LOWER BOUND FRONTIERS IN ARITHMETICAL CIRCUIT COMPLEXITY

by

Maurice Julien Jansen

A dissertation
submitted to the Faculty of the Graduate School
of State University of New York at Buffalo
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

September 2006

Voor mijn Ouders

Acknowledgement

I would like to thank my advisor Kenneth Regan for introducing me to the study of computational complexity. Thank you for your guidance and encouragement. The last six years have been great working together. Thank you also for teaching me how to teach.

I would like to thank Alan Selman and Xin He for being on the Ph.D. committee and Martin Lotz for his work as the outside reader. Thank you for your help in improving this work. Prof. Selman I must also thank for his engaging seminars that I had the pleasure of attending, and for all those fun times being his teaching assistant for the theory of computation course.

Thanks also to fellow theory students Charles Xavier Dennis, Samik Sengupta and Pavan Aduri for creating an enjoyable atmosphere. At the current time they have all vanished from UB, but they were here during most of my stay. I will be remembering those glorious days of taking U-turns in New Jersey!

Finally, I would like to thank my wife, for her support, and, together with my son Rohan, for putting up with all of this.

Even within the limits of the possible, the possibilities are limitless.

- Jules Deelder

Contents

Chapter 1

Introduction

The P vs. NP conundrum, and similar questions posed by theoretical computer science, contain profound mathematical content and carry immediate practical importance. Even without resolution of the main problems, the theory of NP-completeness and more recent extensions regarding hardness of approximation and pseudo-randomness, have been useful companions for practitioners in the field indicating when problems may be too hard to solve. Furthermore, hardness results do not solely have negative implications. For example, the security of most cryptographic systems used in practice is based on unproven hardness assumptions. Also, through the hardness vs. randomness paradigm, hardness of functions has applications in the derandomization of algorithms.

However, proving hardness, i.e., proving lower bounds on the complexity of explicit functions, has turned out to be extremely difficult. For example, currently we still cannot exclude the possibility of solving *SAT* in linear time on a Turing machine. Traditional techniques for reasoning about complexity—such as simulation and diagonalization—do not seem to be adequate because of the so called relativization phenomena [BGS75]. Recently, researchers have taken a new approach by studying Boolean circuit complexity. Circuits promise to lend themselves better to mathematical analysis than Turing machines because they are static finite objects, and their analysis is not subject to diagonalization. Instead of proving $\mathrm{P} \neq \mathrm{NP}$ directly, the focus in this approach is to prove the stronger result that *SAT* does not have polynomial-size Boolean circuits.

Unfortunately, proving lower bounds on general Boolean circuits has turned out to be even more difficult. Currently, there is no explicit Boolean function in $\mathrm{NP} \cup \mathrm{E}$ known to have super-linear circuit size. The current best-known lower bound on the size of an $\{\wedge, \vee, \neg\}$-circuit of an explicit function is $5n - o(n)$ [IM02]. Most progress with circuit complexity has been made by restricting the model. For example, exponential lower bounds are known for constant-depth circuits computing the parity function [FSS81, Hås86, Yao85, Ajt83]. Progress has not been made much beyond this low level, e.g., for all we know non-uniform TC^0 might still contains all of nondeterministic exponential time! We have a good indication of where current techniques are lacking, namely, all circuit lower bounds to this date have been obtained by so-called natural proofs [RR97]. In the presence of pseudo-random generators (PRGs) of certain hardness, for example in TC^0, this type of argument is provably self-defeating. Namely, proving circuit lower bounds for a given class would yield a statistical test for breaking PRGs contained in that class.

Arguably, the most promising approach for obtaining "non-natural" proofs is by the involvement of sophisticated concepts from mathematics that are hard in a certain respect. A promising area for such concepts is algebraic geometry. Algebraic geometry has a long history of development and has many beautiful techniques and deep results. It already has a track record of providing lower bounds, in work by Strassen et al. [BS82][Str73b], Björner, Lovász and Yao [BLY92], and Ben-Or [Ben83].

In order to increase the likelihood of being able to apply algebraic techniques, researchers have considered arithmetical circuits instead of Boolean circuitry. Arithmetical circuits are circuits built from addition and multiplication gates computing a polynomial in the input variables. An analog to the NP-theory exists in this model in the form of Valiant's classes VP and VNP [Val79a, Bür98]. Separation of these classes provides the same intellectual challenge as the P vs. NP question. Over fields of characteristic zero, under assumption of the generalized Riemann hypothesis (GRH), it can be shown that VP = VNP implies that $\mathrm{NC}^3/\mathrm{poly} = \mathrm{PH}/\mathrm{poly}$, and $\#\mathrm{P}/\mathrm{poly} = \mathrm{FP}/\mathrm{poly}$ [Bür00].

However, in this model the best-known lower bounds for explicit functions are obtained by Strassen's degree method [BCS97]. This method relates the size of the arithmetical circuit to a well-studied algebraic invariant, namely the geometric degree, of a certain geometric object obtained from the circuit. Unfortunately, the best possible lower bound we can prove with this technique for an n-variate polynomial of degree d is $\Omega(n \log d)$, i.e., barely non-linear for $d = n^{O(1)}$.

In order to make further progress, researchers have considered more restricted arithmetical circuits [SW99, Shp01]. A natural one is the restriction to constant depth. Contrary to the Boolean case, for fields of characteristic 0 (such as the complex numbers **C**, the real numbers **R**, or the rational numbers **Q**) no super-polynomial lower bounds are known. For finite fields the situation is similar to the Boolean case, and exponential lower bound are known [GR98].

1.1 $\Sigma\Pi\Sigma$-formulas

In characteristic zero, one of the first non-trivial constant-depth models is that of $\Sigma\Pi\Sigma$-formulas, i.e., sums of products of sums of input variables. These networks turn out to be surprisingly powerful. They capture a general form for computing polynomials via Lagrange interpolation. For example, the elementary symmetric polynomial of degree d in n variables has $O(n^2)$ $\Sigma\Pi\Sigma$-formula size, a result first noted by Ben-Or (See Chapter 2). In [SW99] quadratic lower bounds are proved in this model, and optimal lower bounds are obtained for high-degree elementary symmetric polynomials.

Their technique is based on considering the behaviour of the higher order partial derivatives of a given polynomial f, under restriction to arbitrary affine linear subspaces. For a polynomial f in variables $x_1, x_2, \ldots, x_n$, one can define the dth-order partial derivative $\frac{\partial^d f}{\partial X}$ with respect to a multiset of variables X of size d syntactically, with no need for considering a limiting process. Letting $\partial^d(f)$ stand for the set of all such dth-order partial derivatives of f, the dimension of the linear span of the collection of polynomials in $\partial^d(f)$ defines a measure of complexity of the polynomial f. One can generalize this to considering the dimension of the set of dth-order partial derivatives *after restriction* to some affine linear space A, which is denoted by $\dim[\partial^d(f)_{|A}]$. This defines a "progress measure" that is subadditive: for any f, g and A, $\dim[\partial^d(f+g)_{|A}] \leq \dim[\partial^d(f)_{|A}] + \dim[\partial^d(g)_{|A}]$. Let us sketch one of the lower bound arguments of [SW99].

"Reasonable" estimates can be given that bound $\dim[\partial^d(\prod_{i=1}^r L_i)_{|A}]$ for a product of linear forms $L_1, L_2, \ldots, L_r$, provided the degree r is "low". For high degree multiplications this cannot be done. They are dealt with by cancelling them out by means of the restriction to an affine linear space. Polynomials f for which $\dim[\partial^d(f)_{|A}]$ is "high" for *any* affine linear space A can be seen to require large size $\Sigma\Pi\Sigma$-formulas by means of a trade-off argument. Namely, if there are many high degree multiplication gates, the formula must be large to start with, but otherwise, it becomes possible to define a restriction to an affine space A, which is designed to set to zero at least one input of each high degree multiplication gate. Next using the subadditivity property and the "reasonable" bound for low degree multiplication gates, and the fact that $\dim[\partial^d(f)_{|A}]$ is high, one obtains a lower bound on the *multiplicative* size of the formula for f.

In Chapter 3 we continue the study of $\Sigma\Pi\Sigma$-formula. We wil show a refinement of the above described partial derivatives technique, which enables us to account for the number of *addition gates* in the formula, rather than just multiplicative size. Taking circuit size to be the total number of wires in the circuit, we obtain somewhat sharper lower bounds than the Shpilka-Wigderson method would imply for a variety of polynomial families.

Also in Chapter 3, we introduce a companion technique for proving $\Sigma\Pi\Sigma$-formula size lower bounds, which we'll show to be useful in case the partial derivatives technique fails due to an a priori low value of $\dim[\partial^d(f)]$. Our technique exploits a certain *cancellation avoidance* property of polynomials under restriction to affine linear spaces. The crucial notion is that of *resistance* of a polynomial f. Resistance depends on whether f, or more generally whether some higher order partial derivative of f, is non-constant on *all* affine linear spaces of a given dimension k. The smaller this dimension k, the more resistant the polynomial f is, and the larger the $\Sigma\Pi\Sigma$-formulas for computing f are required to be.

All techniques, those of [SW99] and ours, currently known for proving $\Sigma\Pi\Sigma$-formula, are limited to proving at best quadratic lower bounds. A major open problem is to prove super-polynomial lower bounds for explicit functions on $\Sigma\Pi\Sigma$-formula size. Likely candidates to require exponential size in this model are the determinant and permanent polynomials. In light of [Val79a], polynomial-size $\Sigma\Pi\Sigma$-formulas for either one of these implies that all polynomials in VP have polynomial-size depth 3 formulae. Note that recently Mulmuley and Sohoni proposed a representation theoretic approach to prove the permanent requires super-polynomial arithmetical circuit size [MS01].

1.2 Restricting the Role of Constants

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, and recent methods by translation to large finite fields (see [Bür00] after [Koi96]) seem to have limited domain of application. 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 can be employed equally as well as large ones.

To probe the significance of (the magnitude of) field constants, several researchers have obtained (often asymptotically tight) size lower bounds on arithmetical circuits in which a uniform bound is imposed on 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 bilinear 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 bilinear functions. These lower bounds hold even when the bc-restriction is lifted for $O(n^{1-\varepsilon})$-many "help gates." The natural question is, can one obtain similar lower bounds without the bc-restriction at all?

We will continue the study of bilinear circuits with bounded coefficients. In particular our focus will be on the cyclic convolution mapping. It can be computed using the discrete Fourier transform and its inverse by a $O(n \log n)$ size bounded coefficient bilinear circuit, as is a well-known folkore result. Our goal is to generalize the arguments of [Raz02, BL03] to more general models of computation that allow for more unbounded coeffcients.

For this purpose we introduce in Chapter 4 our main bridging concept, resulting in a model whose computational power lies somewhere in between the general unbounded coefficient and bounded coefficient models. This is done by allowing certain linear transformations to be done by the bilinear circuit at the input free of charge. For a bilinear function $f(\vec{x}, \vec{y})$, 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. Such actions on multilinear maps f, like the determinant and permanent polynomials, form the basis of Mulmuley and Sohoni's above mentioned proposal on super-polynomial (arithmetical or Boolean) circuit lower bounds [MS02]. Note that this model not only works past the above-mentioned $O(n^{1-\varepsilon})$ 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*.

In Chapters 5 and 6 we will invesigate the scenerio where the matrices D and E are restricted to be diagonal, focusing on the circular convolution bilinear function. Here one is naturally lead to questions about minors of the $n \times n$ *Fourier matrix* DFT_n. Relations will be established between our aims of proving lower bounds for the diagonal orbit model and discrete analogues of the *Heisenberg uncertainty principle*. Part of our lower bounds will be derived from the Donoho-Stark discrete uncertainty principle [DS89], which gives bounds on the measure of simultaneous concentration of an n-vector x and its discrete Fourier transform $DFT_n x$. As a main result, which will be of independent interest, we will establish a quantative bound on the expected value of the determinant of certain Random Vandermonde matrices with nodes on the unit circle in the complex plane. This result is then used to prove circuit lower bounds. As a by-product we will deduce also an uncertainty type relation for the discrete analog of the *band-limited* functions. Certain limitations of this approach will be probed by considering results known about the so-called *prolate spheroidal wave functions* studied in [Sle78].

1.3 Depth Restrictions

Finally in Chapter 7, we will consider arithmetical circuits of constant bounded depth (not just depth 3 as was done with the $\Sigma\Pi\Sigma$-formulas). First we will establish several structural results that focus on the relation that exists between arithmetical circuits computing a polynomial, and circuits that compute all of it partial derivatives. An analogue will be proved of the Baur-Strassen Derivative Lemma [BS82], in which a circuit for a polynomial p is transformed into a circuit that compute a *linear combination* of all the partial derivatives of p with only constant factor increase in size. This form of the Derivative Lemma has the additional advantage that it truly does not introduce any new constants in the circuit. This is a property the Baur-Strassen lemma notoriously is known not to satisfy. We will extend some of the results of [Lok01] to a particular kind of bounded depth bounded constant ”linear combination” bilinear formula.

Next, we will consider bounded depth bilinear circuits without any kind of assumption on the magnitude of constants. In [RR03] a weakly non-linear lower bound is proved for the matrix multiplication bilinear map. The proof involves a “superconcentrator Lemma” to prove the lower bound. We combine this lemma with the *discrete uncertainty principle for cyclic groups of prime order*, as proved by Tao [Tao91], to obtain a non-linear lower bound for the cyclic convolution bilinear map.

Chapter 2

Preliminaries

2.1 Background Material

All rings are assumed to be commutative and have a multiplicative identity 1. We write $[n]$ as shorthand for $\{1,\ldots,n\}$. We assume familiarity with standard notation for complexity classes such as $\mathrm{P} = \cup_{k\geq 0}\mathrm{DTIME}[n^k]$, $\mathrm{NP} = \cup_{k\geq 0}\mathrm{NTIME}[n^k]$, and so on.

2.1.1 Computational Models

Let $R[x_1,\ldots,x_n]$ denote the polynomial ring in variables $x_1,\ldots,x_n$ over a ring R.

Definition 2.1.1. Let R be a ring and $x_1,x_2,\ldots,x_n$ be a set of variables. An **arithmetical circuit over** R is a 3-tuple (G,γ,κ), where $G = (V,E)$ is a directed acyclic graph and $\gamma : V \to R \cup \{x_1,x_2,\ldots,x_n\} \cup \{+,\times\}$ is the *gate identification function* and $\kappa : E \to R$ is the *wire constants function*, satisfying:

1. if in-degree$(v) = 0$, then $\gamma(v) \in R \cup \{x_1,x_2,\ldots,x_n\}$,

2. if in-degree$(v) > 0$, then $\gamma(v) \in \{+,\times\}$,

The vertices and edges in an arithmetical circuit are called *gates* and *wires*. Gates with in-degree 0 are called **input gates**, or *inputs* for short. All other gates are called *regular gates*. For a regular gate v, if $\gamma(v) = +$, then v is called an *addition gate*, and if $\gamma(v) = \times$, v is called a *multiplication gate*. For a gate v, a wire of the form (u,v) is called an *input wire to* v, and a wire of the form (v,u) is called an *output wire from* v. Note that constants can appear on wires and as inputs.

Definition 2.1.2. Given an arithmetical circuit $\mathcal{C} = (G,\gamma,\kappa)$ we define the **polynomials computed by** $\mathcal{C}$ to be the function $\phi : V[G] \to R[x_1,\ldots,x_n]$ inductively as follows:

1. $\phi(v) = \gamma(v)$, if v is an input gate,

2. $\phi(v) = \sum_{i=1}^{r} \kappa(e_i)\phi(v_i)$, if v is an addition gate with input wires $e_1 = (v_1, v)$, $e_2 = (v_2, v), \ldots, e_r = (v_r, v)$, and

3. $\phi(v) = \prod_{i=1}^{r} \kappa(e_i)\phi(v_i)$, if v is a multiplication gate with input wires $e_1 = (v_1, v)$, $e_2 = (v_2, v), \ldots, e_r = (v_r, v)$.

In the above definition, $\phi(v)$ is called the *polynomial computed by the gate* v. If for a polynomial $p \in R[x_1, x_2, \ldots, x_n]$ there exists a gate $v \in V[G]$ for which $\phi(v) = p$, we say *p is computed by $\mathcal{C}$*.

The *size* of an arithmetical circuit $\mathcal{C} = (G, \gamma, \kappa)$, denoted by $s(\mathcal{C})$, is defined to be the total number of wires in G. The *multiplicative* and *additive* size of $\mathcal{C}$, denoted by $s^*(\mathcal{C})$ and $s^+(\mathcal{C})$, respectively are defined by

$$s^*(C) = |\{(u,v) \in E[G] : \gamma(v) = \times\}|,$$

and

$$s^+(C) = |\{(u,v) \in E[G] : \gamma(v) = +\}|.$$

Definition 2.1.3. An **arithmetical formula** is an arithmetical circuit $\mathcal{F} = (G, \gamma, \kappa)$ for which all regular gates have out-degree at most one. For formulae, their size, multiplicative size and additive size are denoted by $\ell(\mathcal{F}), \ell^*(\mathcal{F})$, and $\ell^+(\mathcal{F})$, respectively.

Note that in the above definition we did not provide subtraction and division gates. The former can be handled in our model using addition gates with -1 on the second input wire. By standard robustness results [BCS97], it is not necessary to include division gates for computing polynomials.

Definition 2.1.4. Let $p_1, p_2, \ldots, p_m$ be a collection of polynomials from $R[x_1, \ldots, x_n]$. The **circuit/formula complexity of** p **over** R, denoted by $s_R(p_1, p_2, \ldots, p_m)$ and $\ell_R(p_1, p_2, \ldots, p_m)$ respectively, is the size of a smallest circuit/formula computing all of $p_1, p_2, \ldots, p_m$. For multiplicative and additive size, these are denoted by $s_R^{\square}(p_1, p_2, \ldots, p_m)$ and $l_R^{\square}(p_1, p_2, \ldots, p_m)$, with $\square \in \{*, +\}$.

In case it is clear from the context which underlying ring R we are working over, we will drop the subscript R in our notation. Sometimes the underlying field matters. For example, over the complex numbers $\mathbf{C}$, $\ell^*_{\mathbf{C}}(x_1^2 + x_2^2) = 1$ witnessed by the formula $(x_1 + ix_2)(x_1 - ix_2)$, but over the real numbers $\mathbf{R}$ one has $\ell^*_{\mathfrak{R}}(x_1^2 + x_2^2) = 2$. Suprisingly however, many results and properties are independent of R, or care only whether R is finite or infinite, and if so whether its charateristic is 0, 2, or an odd prime.

We will now define several computational models of importance, that satisfy additional restrictions.

2.1.2 ΣΠΣ-formulae

As the main object of study in Chapter 3 we have the following model introduced by Shpilka and Wigderson [SW99]:

Definition 2.1.5. A ΣΠΣ-formula is an arithmetical formula $\mathcal{F} = (G, \gamma, \kappa)$ such that on any directed path $(u_1, u_2), (u_2, u_3), \ldots, (u_{m-1}, u_m)$ in G there do not exists indices $1 \leq i < j < k \leq m$ such that $\gamma(u_i) = \gamma(u_k) = \times$ and $\gamma(u_j) = +$.

In other words, a ΣΠΣ-formula can be thought of as having the following structure. First there is a group of addition gates computing linear forms of the input variables, then there is a group of multiplication gates that multiply these linear forms. Finally there is a last group of gates that compute linear combinations of these products.

For a collection of polynomials $p_1, p_2, \ldots, p_m$ we will denote the size of a smallest ΣΠΣ-formula computing $p_1, p_2, \ldots, p_m$ by $\ell_{3,R}(p_1, p_2, \ldots, p_m)$. Similarly as before, we define $\ell^*_{3,R}$ and $\ell^+_{3,R}$ for multiplicative and addive complexity. Note that in Chapter 3 the underlying ring is assumed to be an arbitrary field of characteristic 0, e.g., the complex numbers **C**, and we will drop the R subscript there.

Given a ΣΠΣ-formula computing a single polynomial p with s multiplication gates in some fixed order, we can write

$$p = \sum_{i=1}^{s} M_i,$$

where

$$M_i = \Pi_{j=1}^{d_i} l_{i,j},$$

and

$$l_{i,j} = c_{i,j,1}x_1 + c_{i,j,2}x_2 + \ldots + c_{i,j,n}x_n + c_{i,j,0}.$$

Here d_i is the in-degree of the ith multiplication gate, and $c_{i,j,k}$ is nonzero if and only if there is a wire from x_k to the addition gate computing $l_{i,j}$. Note that $l_{i,j}$ is homogeneous of degree 1, i.e., strictly linear, if $c_{i,j,0} = 0$, and is affine linear otherwise. For an affine linear form l, we will denote its strictly linear part by l^h.

2.1.3 Linear and Bilinear Circuits

Definition 2.1.6. A circuit $\mathcal{L} = (G, \gamma, \kappa)$ is called a **linear circuit** if it has no multiplication gates, i.e., for each gate v, $\gamma(v) = +$, $\gamma(v) \in R$, or $\gamma(v) = x_i$ for some variable x_i. If for no gate v, $\gamma(v) \in R$, the circuit is called **homogeneous**.

For linear circuits, R will be assumed to be a field. In a homogeneous linear circuit each gate computes a homogeneous linear form: for each $g \in V[G]$, $\phi(g) = a_1x_1 + a_2x_2 + \ldots + a_nx_n$ with $a_i \in R$. An ordered list of k gates $(g_1, g_2, \ldots, g_k)$ thus define a linear transformation $R^n \to R^k$ given by mapping $a = (a_1, a_2, \ldots, a_n) \mapsto (\phi(g_1)(a), \phi(g_2)(a), \ldots, \phi(g_k)(a))$. A $k \times n$ matrix A likewise determines a linear transformation $R^n \to R^k$ defined by mapping

$a = (a_1, a_2, \ldots, a_n)^T \mapsto Aa$. We denote by $s_{lin}(A)$ the minimum size of a linear circuit that computes this linear transformation.

For bilinear circuits the set of variables is assumed to be partitioned into two sets $\{x_1, x_2, \ldots, x_n\} \cup \{y_1, y_2, \ldots, y_m\}$. We will study the following homogeneous bilinear circuit model:

Definition 2.1.7. A **homogeneous bilinear circuit** is an arithmetic circuit $\mathcal{B} = (G, \gamma, \kappa)$ satisfying:

1. for each multiplication gate v, the polynomial $\phi(v)$ computed at v is a homogeneous bilinear form in variables $\{x_1, x_2, \ldots, x_n\} \cup \{y_1, y_2, \ldots, y_m\}$, and
2. no input gate v has $\gamma(v) \in R$.

For a set of bilinear polynomials $p_1, p_2, \ldots, p_k \in R[x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m]$, we denote bilinear circuit complexity by $s_{b,R}(p_1, p_2, \ldots, p_k)$. Similarly as before we define notation s_b^*, s_b^+, l_b, l_b^*, and l_b^+ for additive/multiplicative circuit/formula size.

Any homogeneous circuit computing a linear transformation wlog. can be assumed to have no multiplication gates. Any homegeneous circuit computing a set of bilinear forms can therefore wlog. be assumed to a homegeneous bilinear circuit of the structure defined above.

The above models will be considered under restriction of constants on the wires to bounded coefficients. Generally, one could define (families of) bounded-coefficient circuits over **C** or **R** by restricting constants on the wires to have norm $O(1)$. We will adhere to a stricter definition, with the knowledge that typical results easily generalize to $O(1)$ size constants:

Definition 2.1.8. A circuit $\mathcal{C} = (G, \gamma, \kappa)$ over **C** or **R** is called a *bounded-coefficient* circuit if for every $e \in E[G]$, $|\kappa(e)| \leq 1$.

We will use the sub/superscript "bc" to indicate bounded coefficient size of polynomials. For bounded coefficient homogeneous linear circuits, lower bounds can be obtained through the following result by Morgenstern:

Theorem 2.1.1 ([Mor73]) *Let A be an $n \times n$ matrix, then $s_{lin}^{bc}(A) \geq \log_2 |\det(A)|$.*

We define the discrete Fourier transform matrix DFT_n by

$$(DFT_n)_{ij} = \omega^{ij},$$

where $\omega = e^{2\pi i/n}$. Its unitary version we denote by F_n:

$$F_n = \frac{DFT_n}{\sqrt{n}}.$$

The conjugate transpose of a matrix A will be denoted by A^*. A matrix A is called *Hermitian or self-adjoint* if $A^* = A$. A matrix is called *unitary* if $AA^* = A^*A = I$. As indicated above

$F_nF_n^* = F_n^*F_n = I$. A little elementary linear algebra shows:

$$|det(DFT_n)|^2 = det(DFT_n)\overline{det(DFT_n)} = det(DFT_n)det(DFT_n^*) = n^n.$$

So by Morgensterns result:

$$s_{lin}^{bc}(DFT_n) \geq \frac{n}{2}\log_2 n,$$

which is asymptotically tight, given that the circuits for DFT_n as given by Cooley and Tukey [CT65] are of size $O(n \log n)$ and have bounded coefficients.

2.1.4 Perturbation Theory

We require the following basic results from perturbation theory, see e.g. [Bha97]. For vector $x = (x_1, x_2, \ldots, x_n) \in \mathbf{C}^n$, we define its ℓ_2-norm by $||x||_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$. The ℓ_2-norm (or *spectral norm*) of an $m \times n$ matrix A is defined by

$$||A||_2 = \max_{x \neq 0} \frac{||Ax||_2}{||x||_2},$$

and the *Frobenius norm* is defined by

$$||A||_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^2}.$$

An *eigenvalue* of a complex square marix A is a complex number λ for which there exist a vector x such that $Ax = \lambda x$. For Hermitian matrices all eigenvalues are real numbers. We denote the ith largest eigenvalue of an $n \times n$ Hermitian matrix A by $\lambda_i(A)$, i.e. we have $-\infty < \lambda_n(A) \leq \lambda_{n-1}(a) \leq \ldots \leq \lambda_1(A) < \infty$.

Theorem 2.1.2 (Weyl's Perturbation Theorem) *Let A and E be Hermitian matrices. Then*

$$\max_j |\lambda_j(A) - \lambda_j(A+E)| \leq ||E||_2.$$

We also need the following theorem.

Theorem 2.1.3 (Hadamard Inequality) *For an $n \times n$ complex matrix A with columns $a_1, a_2, \ldots, a_n$,*

$$|det(A)| \leq \prod_{i=1}^n ||a_i||_2.$$

Intuitively speaking, for an $n \times n$ matrix A, $|\det(A)|$ is the volume of the parallelipiped spanned by its columns (or rows). This volume is maximized by making the columns orthogonal, and it can then be computed by just taking the n-product of the lengths of these vectors. This is essentially the content of the above theorem.

2.1.5 Cyclic Convolution

Definition 2.1.9. The **cyclic convolution** $x \circ y$ of two n-vectors $x = (x_0, x_1, \ldots, x_{n-1})^T$ and $y = (y_0, y_1, \ldots, y_{n-1})^T$ is the n-vector $(z_0, \ldots, z_{n-1})^T$ 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 $\mathrm{Circ}(a)$, which we define by

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

That is, we have that

$$x \circ y = \mathrm{Circ}(x)y = \mathrm{Circ}(y)x.$$

Convolution can be computed using the Fourier transform, according to the following folklore result:

Theorem 2.1.4 (The Convolution Theorem) *For any $a \in \mathbf{C}^n$,*

$$\mathrm{Circ}(a) = F_n diag(DFT_n a) F_n^*.$$

In the above, for a vector $x = (x_1, x_2, \ldots, x_n)^T$,

$$\mathrm{diag}(x) = \begin{pmatrix} x_1 & 0 & \cdots & 0 & 0 \\ 0 & x_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \\ 0 & 0 & \cdots & x_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & x_n \end{pmatrix}.$$

Through the convolution theorem and using the $O(n \log n)$ circuits for the Fourier transform, we thus obtain

$$s_{bc}(x \circ y) = O(n \log n).$$

We also find it convenient to consider the "half convolution" defined by $\mathrm{HCirc}(x)y$, where $\mathrm{HCirc}(a)$ 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 $\mathrm{HCirc}(x)y$ to the inverted vector $\mathrm{HCirc}(x_{n-1}, x_{n-2}, \ldots, x_1)(y_1, y_2, \ldots, y_{n-1})$, which can be done by bilinear (bc) circuits with linearly many extra $+$ gates. Thus lower bounds on $x \circ y$ extend immediately to $\mathrm{HCirc}(x)y$. The convenience is that $\mathrm{HCirc}(x)y$ is definable by recursion from $\mathrm{HCirc}(x_1, \ldots, x_{n-2})(y_1, \ldots, y_{n-2})$, 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 $\mathrm{HCirc}(a)$, observe that the piece in between is the lower triangle of $\mathrm{HCirc}(a_1, \ldots, a_{n-2})$ 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 4.0.1.

2.1.6 Families of Polynomials

Let R be a ring. In general if $l(n)$ is a strict monotone increasing function on natural numbers and

$$P = \{p_n \in R[x_1, x_2, \ldots, x_{l(n)}]\}_{n>0}$$

is a family of polynomials one can define the (non-uniform) complexity as the function defined by $s(n) = s(p_{l(n)})$. For uniform complexity one would requires in addition the existence of some Turing machine that can output descriptions of minimum circuits for each n, but in this document we will only consider non-uniform complexity.

Let S_n be the symmetric group. The determinant polynomial Δ_n and permanent polynomial Π_n on n^2 variables are defined by

$$\Delta_n = \sum_{\sigma \in S_n} \mathtt{sgn}(\sigma) \prod_{i=1}^{n} x_{i,\sigma(i)}, \qquad \Pi_n = \sum_{\sigma \in S_n} \prod_{i=1}^{n} x_{i,\sigma(i)},$$

where sgn(σ) is the sign of the permutation σ. Note that Π_n is the same as Δ_n except without the sign alternations, and these are the same polynomial when the underlying field has characteristic 2. Valiant [Val79a] proposed a theory analogous to the theory of NP-completeness in which the determinant and permanent play the roles of feasible and infeasible complete problem. The determinant has polynomial size arithmetical and Boolean circuits. The permanent is strongly suspected not to have polynomial size circuits of either kind [Val79b, Bür98]. Raz [Raz04a] recently showed that any multilinear formula computing the permanent or determinant must have size $n^{\Omega(\log n)}$. Both Δ_n and Π_n are expected to require exponential size in the ΣΠΣ-formula model. However, the best-known lower bound for both Δ_n and Π_n is $\Omega(n^4/\log n)$, i.e., $\Omega(N^2/\log N)$ in the number $N = n^2$ of variables [SW99].

Next we define the elementary symmetric polynomial of degree d:

$$S_n^d = \sum_{\substack{T\subset[n]\\|T|=d}} \prod_{i\in T} x_i.$$

Ben-Or observed the surprising fact that S_n^d has $O(n^2)$ size ΣΠΣ-formulas, where the constant in the big-O does not depend on d. This is done as follows. Define the polynomial $g(t) = \Pi_{t=1}^n (t+x_i)$. Observe that $g(t) = \sum_{d=0}^n S_n^d(X)t^{n-d}$. We can compute $g(t_0),\dots,g(t_n)$ for any given constants $t_0 \dots t_n$ in parallel with $n+1$ multiplication gates of degree n. Now, from the Lagrange interpolation formula, it follows that the coefficient of t^{n-d}, which equals S_n^d, is a linear combination of $g(t_0)\dots g(t_n)$. Hence we obtain a ΣΠΣ-formula for S_n^d using a total of at most $3n^2+4n+1$ wires. In [SW99] the following lower bound was obtained for S_n^d:

$$\ell_3^*(S_n^{2d}) \geq \max\left(\Omega(\frac{n^{\frac{2d}{d+2}}}{d}), \Omega(nd)\right), \forall d \leq 4n/9.$$

In light of the Ben-Or upper bound, we see that this is tight for $d = \Omega(n)$.

2.1.7 Algebraic Geometry

For results and definitions stated in this section see [CLO92, BCS97].

Definition 2.1.10. Let R be a ring. A subset I of R is an **ideal** if,

1. for any $a \in I$, for all $r \in R$, $ra \in I$, and
2. for all $a,b \in I$, $a+b \in I$.

For example, if $a_1,\dots,a_s$ are elements of R, then the set of all elements of the form $r_1a_1+\dots+r_sa_s$, with all $r_i \in R$, is an ideal. It is called the *ideal generated by* $a_1,\dots,a_s$, and denoted by $a_1R+\dots+a_sR$ or just $(a_1,\dots,a_s)$. If for an ideal there exist finitely many elements $a_1,\dots,a_s$, such that $I = (a_1,\dots,a_s)$, then I is called *finitely generated*. It is a fact that the polynomial ring $F[x_1,\dots,x_n]$ is *Noetherian*, implying that all its ideals are finitely generated.

Let I,J be ideals. Observe that $I \cap J$ is an ideal, and that the set of all elements $a+b$ with $a \in I, b \in J$, is an ideal. We denote it by $I+J$. More generally, for a family of ideals $\{I_s\}_{s\in S}$, define $\sum_{s\in S} I_s$, to be the set of all sums $\sum_{s\in S} a_s$, with $a_s \in I_s$, $a_s \neq 0$, for only finitely many s. Let $I \cdot J$ be the set of all finite sums $\sum_i a_ib_i$, with $a_i \in I, b_i \in J$; then $I \cdot J$ is an ideal.

Now let $R = F[x_1,\dots,x_n]$. The set F^n, of all n-tuples $(a_1,\dots,a_n)$ with $a_i \in F$ is called *n-dimensional affine space over* F. The elements of F^n are called *points*.

Definition 2.1.11. Let I be an ideal in R. The *affine variety defined by* I, denoted by $V(I)$, is the subset of tuples $(a_1,\dots,a_n) \in F^n$, such that $f(a_1,\dots,a_n) = 0$, for every polynomial $f \in I$.

We have the following elementary proposition:

Proposition 2.1.5 *For ideals $I, J, \{I_s\}_{s\in S}$ in R, polynomials $f_1, \dots, f_s \in R$,*

1. $V(\sum_{s\in S} I_s) = \cap_{s\in S} V(I_s)$.
2. $V(I \cdot J) = V(I) \cup V(J)$.
3. $V(R = (1)) = \emptyset$.
4. $V((0)) = F^n$.

The above Proposition shows that we can define a topology on n-dimensional affine space, by taking as closed sets all varieties in F^n. This topology is called the *Zariski topology*.

Proposition 2.1.6 *Let V be a subset of F^n. Then the set of all polynomials $f \in F[x_1, \dots, x_n]$ such that $f(a_1, \dots, a_n) = 0$, for every point $(a_1, \dots, a_n) \in V$, is an ideal. This ideal is denoted by $I(V)$.*

We would like to define the geometric degree of an affine variety. In order to do so we must introduce the concept of projective space. In the following, let $R = F[x_0, \dots, x_n]$.

Definition 2.1.12. Let P^n be the set of all $(n+1)$-tuples $(a_0, \dots, a_n) \in F^{n+1}/(0, \dots, 0)$, where we identify points $(a_0, \dots, a_n)$ and $(b_0, \dots, b_n)$, if there exist a nonzero $\lambda \in F$, such that $a_i = \lambda b_i$, for all $i \in \{0, \dots, n\}$. P^n is called *n-dimensional projective space*. The equivalence class of a point $(a_0, \dots, a_n)$ is denoted by $[a_0 : \dots : a_n]$.

A polynomial is called *homogeneous*, if all its monomials are of the same degree. An ideal $I \in R$ is called *homogeneous* if it can be generated by homogeneous polynomials. For a polynomial $f \in F[x_1, \dots, x_n]$, it homogenization f^h is defined by $x_0^{\deg(f)} f(\frac{x_1}{x_0}, \dots, \frac{x_1}{x_0})$.

Definition 2.1.13. Let I be a homogeneous ideal in R. Let $V(I)$ be the set of point $(a_0 : \dots : a_n) \in P^n$ such that $f(a_0, \dots, a_n) = 0$, for all homogeneous $f \in I$. $V(I)$ is called the *projective variety defined by I*. Conversely, if V is a subset of P^n, then the ideal generated by all homogeneous polynomials $f \in F[x_0, \dots, x_n]$ that vanish on V, is called the *ideal of the variety V*, and denoted by $I(V)$.

As in affine space, the set of all varieties in P^n forms a topology. We can embed n-dimensional affine space into P^n via the map $\phi : F^n \to P^n$, defined by mapping $(a_1, \dots, a_n)$ to $[1 : a_1 : \dots : a_n]$.

For a homogeneous ideal I in R, let $I^{(t)}$ be the set of all homogeneous polynomials of I of degree t, and let $R^{(t)}$ be the set of all homogeneous polynomials of degree t. $I^{(t)}$ is a vector subspace of $R^{(t)}$. Define $H_I(t) =$ codimension of $I^{(t)}$ in $R^{(t)}$. The function $H_I(t)$ is called the *Hilbert function* of the ideal I. We have the following classical result.

Theorem 2.1.7 (Hilbert-Serre, see [BCS97], p. 178) *Let I be a homogeneous ideal of $R = F[x_0, \dots, x_n]$, and assume that $V(I)$ is nonempty and of dimension d. Then there exist unique integers $h_0, h_1, \dots, h_d$, such that the polynomial*

$$h(T) = \sum_{j=0}^{d} h_j \binom{T}{d-j}$$

satisfies $h(t) = H_I(t)$, for all sufficiently large $t \geq 0$. The uniquely determined polynomial h, is called the Hilbert polynomial of the ideal I.

Definition 2.1.14. We define the **geometric degree** $\mathrm{GDEG}(I)$ of the homogeneous ideal I, to be the uniquely determined integer h_0 of Theorem 2.1.7. The geometric degree of a projective variety V is defined as the geometric degree of $I(V)$.

The above is the classical definition of geometric degree of a projective variety found in algebraic geometry.

Definition 2.1.15. A subset V of a topological space X is **irreducible**, if it is nonempty, and whenever we can write $V = U \cup W$, for sets U and W that are closed in V, then one of U and W equals V.

Affine and projective n-space are *Noetherian* topological spaces, which implies that every variety V has a unique decomposition $V = V_1 \cup \ldots \cup V_s$ into irreducible varieties, up to order of terms. The V_i's are called the *components* of V. For an ideal I, define the *radical of I* to be the ideal $\{f \mid \exists n > 0, f^n \in I\}$. The radical of I is denoted by $\sqrt{I}$. A cornerstone of algebraic geometry is the following theorem:

Theorem 2.1.8 (Hilberts's (Strong) Nullstellensatz) *Over algebraically closed fields, for any ideal I, it holds that*

$$I(V(I)) = \sqrt{I}.$$

This gives a 1-1 correspondence between radical ideals and varieties called the algebra-geometry dictionary. An ideal I is called *primary*, if for every $a \notin I$, for every b with $ab \in I$, it holds that $b^n \in I$, for some $n > 0$. An ideal I is called *prime*, if for every $a, b \in R$, if $ab \in I$, then a or b is in I. In the algebra-geometry dictionary prime ideals correspond 1-1 with irreducible varieties. Every ideal I can be written as an intersection $I = I_1 \cap \ldots \cap I_s$ of primary ideals, called the *primary decomposition* of I, such that $V(I_1) \cup \ldots \cup V(I_s)$ is a decomposition of $V(I)$ into irreducible components, and with radicals $\sqrt{I}_j$ being unique prime ideals.

Definition 2.1.16. For a non-empty affine variety V, let $V_1, \ldots, V_s$ be the irreducible components of the closure of $\phi(V)$ in the Zariski topology. We define[1] the *affine geometric degree* $\mathrm{gdeg}(V)$ of V by

$$\sum_{i=1}^{s} \mathrm{GDEG}(V_i),$$

This can be computed from any ideal I such that $V = V(I)$ by calculating a primary decomposition of I^h as $I_1 \cap \ldots \cap I_s$ and then summing $\mathrm{GDEG}(\sqrt{I}_j)$ over the factors. By convention we let $\mathrm{gdeg}(\emptyset) = -1$.

[1]Caution to the reader: this differs from [BCS97], def. 8.22., by decomposing $\overline{\phi(V)}$ rather than V. This makes the affine case subordinate to the projective case, and Theorem 2.1.9 merely specializes the statement in [BCS97].

Two facts about affine geometric degree:

1. $\mathrm{gdeg}(F^n) = 1$.
2. If f is a polynomial of degree $d \geq 1$, then $\mathrm{gdeg}(V(f)) \leq \deg(f)$.

The main fact we will use about degree is the following from of Bézout's Theorem stated as an inequality:

Theorem 2.1.9 (cf. [BCS97], p. 181) *For affine varieties X and Y, we have that*

$$gdeg(X \cap Y) \leq gdeg(X) \cdot gdeg(Y).$$

Chapter 3

Lower Bounds on $\Sigma\Pi\Sigma$-formulae

In contrast to the case of Boolean circuit complexity, in *arithmetical* circuit complexity we do not currently have exponential lower bounds (for "natural" mathematical functions) against constant-depth circuits, or even constant-depth formulas, in case the underlying field has characteristic zero. Shpilka and Wigderson [SW99] noted that such lower bounds are unknown even for formulas that are sums-of-products-of-sums, the $\Sigma\Pi\Sigma$ formulas defined in Chapter 2.

These formulas have notable *upper-bound power* because they can carry out forms of Lagrange interpolation, including that needed to compute the symmetric polynomials S_n^d (defined to be the sum of all degree-d monomials in n variables) in quadratic size. This heightens the contrast because the Boolean majority function, which is analogous to $S_n^{\lceil n/2 \rceil}$, requires exponential size in constant-depth Boolean circuits [Hås88]. Thus $\Sigma\Pi\Sigma$ formulas present a substantial challenge for lower bounds, as well as being a nice small-scale model to study.

The *multiplicative size* ℓ^* of an arithmetical formula or circuit with gates of bounded or unbounded fan-in can be taken as the total fan-in to multiplication gates. Lower bounds on ℓ^* imply lower bounds on the total circuit/formula size ℓ, taken as the number of wires in the circuit/formula. The best known lower bound for general arithmetical circuits has remained for thirty years the $\Omega(n \log n)$ lower bound on ℓ^* by the "Degree Method" of Strassen [Str73a] (see also [BS82, BCS97]), which however applies to some simple functions such as $f(x_1, \ldots, x_n) = x_1^n + \ldots + x_n^n$. Shpilka and Wigderson [SW99] proved lower bounds on ℓ^* of $\Omega(n^2)$ for S_n^d when $d = \Theta(n)$, $n^{2-\varepsilon(d)}$ for S_n^d with small values of d, and $\Omega(N^2/\operatorname{polylog}(N))$ lower bounds for the determinant, with $N = n^2$. Of course, many natural arithmetical functions including the permanent [Val79b] are conjectured to require exponential size (for ℓ^*) for general circuits, let alone $\Sigma\Pi\Sigma$ ones. Straight counting of equations for monomial coefficients show that "generically" functions need exponential size. However, Strassen's technique has the limitation that $\Omega(n \log n)$ is the best lower bound for a polynomial of total degree $n^{O(1)}$ in n variables that it can prove, and the main methods of Shpilka and Wigderson [SW99] have a similar limitation of $\Omega(n^2)$ for $\Sigma\Pi\Sigma$ formulas. Shpilka [Shp01] gets past this only in some further-restricted cases, and also considers a depth-2 model consisting of an arbitrary symmetric function of sums. This barrier provides another reason to study the $\Sigma\Pi\Sigma$ model, in order to understand the obstacles and what might be needed to surpass them.

In this chapter we prove a sharp n^2 lower bound on ℓ^* for $\Sigma\Pi\Sigma$ formulas for the function

$f(x_1,\ldots,x_n) = x_1^n + \ldots + x_n^n$ computed over the real or rational numbers, and a lower bound of $n^2/2$ over any field of characteristic zero. Note the absence of "O,Ω" notation. These lower bounds are obtained via a new notion we introduce, namely the *resistance* of a polynomial. A technique is introduced for proving up to quadratic ΣΠΣ-formula size lower bounds for polynomials with high resistance.

Next we prove lower bounds on the total complexity ℓ for some of Shpilka and Wigderson's functions that are significantly higher (but still sub-quadratic) than their bounds on ℓ^* when the degree d of the function is small. This is done intuitively by exploiting a closed-form application of the Baur-Strassen "Derivative Lemma" to ΣΠΣ formulas, showing that f and all of its n first partial derivatives can be computed with only a constant-factor increase in ℓ and ℓ^* over ΣΠΣ formulas for f.

3.1 Preliminaries

Let us recall the computational model. A ΣΠΣ-formula is an arithmetic formula consisting of four consecutive layers: a layer of input gates, followed by a layer of addition gates, followed by a layer of multiplication gates, followed by the output. Wires can be assumed to be present only between consecutive layers. For a polynomial p, $\ell_3(p)$ will denote the size of a smallest ΣΠΣ-formula computing p. Given a ΣΠΣ-formula for a polynomial p, we can write

$$p = \sum_{i=1}^{s} M_i,$$

where

$$M_i = \Pi_{j=1}^{d_i} l_{i,j},$$

and

$$l_{i,j} = c_{i,j,1}x_1 + c_{i,j,2}x_2 + \ldots + c_{i,j,n}x_n + c_{i,j,0}.$$

Here d_i is the in-degree of the ith multiplication gate (fix any order on the multipication gates), and $c_{i,j,k}$ is nonzero iff there is a wire from x_k to the addition gate computing $l_{i,j}$. Note that $l_{i,j}$ is homogeneous of degree 1, i.e. *strictly linear*, if $c_{i,j,0} = 0$, and is *affine linear* otherwise.

3.1.1 Affine Linear Subspaces and Derivatives

An *affine linear* subspace A of F^n is a set of the form $A = V + w = \{\, v + w : v \in V \,\}$, where V is a linear subspace of F^n, and w is a vector in F^n. The dimension of A is defined to be the vector space dimension of V.

Let $X = \{x_1,\ldots,x_n\}$ be a set of variables. For any affine subspace A, we can always find a set of variables $B \subset X$, and affine linear forms l_x in the variables $X \setminus B$, for each $x \in B$, such that A is the set of solutions of $\{x = l_x : x \in B\}$. This representation is not unique. The set B is called a *base* of A. The size $|B|$ always equals the co-dimension of A.

To indicate how one obtains a base, say $\dim V = r$ and let R be an $n \times r$ matrix whose columns form a basis of V. Then

$$A = \{R\beta + w : \beta \in F^r\}.$$

Since row-rank(R) = col-rank(R) = r, there must be r independent rows. Let S be any $r \times r$ submatrix of R with independent rows. B is taken to be the set of variables corresponding to rows not in S. Any specified vector of values a can be obtain for variables in X/B: set $\beta = S^{-1}(a - w_S)$, where w_S is the r-vector obtain by limiting w to the rows of S. Then the variables in B are determined. Thus the affine linear forms l_b are given by $R_B\beta = R_B S^{-1}(X/B - w)$, where we denote by R_B the rows of R that are in B, and X/B the r-vector of variables not in B.

In the following, whenever we consider an affine linear subspace A, we assume we have fixed some base B of A. Any of our numerical "progress measures" used to prove lower bounds will not depend on the choice of a base. The following notion *does* depend on the choice of a base:

Definition 3.1.1 ([SW99]). Let A be an affine linear subspace of F^n, and let $f \in F[x_1, \ldots, x_n]$. Then the *restriction of f to A* is the polynomial obtained from f by substituting l_b for the variable x_b for each $b \in B$, is denoted by $f_{|A}$. If W is a set of polynomials, define $W_{|A} = \{f_{|A} \mid f \in W\}$.

Then we define:

Definition 3.1.2. For polynomial $f \in F[x_1, \ldots, x_n]$, define the first order gradient mapping $\nabla f : F^n \to F^n$ by

$$\nabla f(a_1, \ldots, a_n)_k = \frac{\partial f}{\partial x_k}(a_1, \ldots, a_n).$$

For linear polynomial $l = c_1x_1 + \ldots + c_nx_n + c_0$, we denote $l^h = c_1x_1 + \ldots + c_nx_n$. For a set S of linear polynomials, $S^h = \{l^h : h \in S\}$. We have the following proposition:

Proposition 3.1.1 *Let S be a set of s polynomials of degree 1 from $F[x_1, \ldots, x_n]$, such that S^h is an independent set. Then the set of common zeroes of S is affine linear of dimension $n - s$.*

Proof. Let V be the set of common zeroes of S^h. V is a linear space of dimension $n - s$. Since S^h is an independent set, one can conclude there exists v, such that for all $l \in S$, $l(v) = 0$. All of $v + V$ vanishes on S: for $v' \in V$, $l(v + v') = l^h(v + v') + c = l^h(v) + l^h(v') + c = l(v) + l^h(v') = 0$. Conversely, if for w, for all $l \in S$, $l(w) = 0$, then writing $w = w' + v$. $0 = l(w) = l(w' + v) = l^h(w' + v) + c = l^h(w') + l^h(v) + c = l^h(w') + l(v) = l^h(w')$, so $w' \in V$, so $w \in v + V$. □

3.2 Resistance of polynomials

We introduce the following notion.

Definition 3.2.1. A polynomial f in variables $x_1, x_2, \ldots, x_n$ is (d, r, k)*-resistant* if for any polynomial $g(x_1, x_2, \ldots, x_n)$ of degree at most r, for any affine linear subspace A of codimension k, there exists a dth order partial derivative of $f - g$ that is non-constant on A.

For a multiset X of size d with elements taken from $\{x_1, x_2, \ldots, x_n\}$, we will use the notation $\frac{\partial^d f}{\partial X}$ to indicate the dth-order derivative with respect to the variables in X. An elementary fact is that the order of taking derivatives does not matter.

For polynomials with terms of different degrees, the middle parameter r in the definition might be useful. However, typically in the applications r is set to be $\deg(f)-1$. Convention will be that when we say a polynomial f is (d,k)-*resistant*, we mean f is $(d, \deg(f)-1, k)$-resistant.

Definition 3.2.2. For a polynomial $f(x_1, x_2, \ldots, x_n)$ we define its *resistance factor* $\mu(f)$ by

$$\mu(f) = max\{\frac{k+1}{d+1} : f \text{ is } (d,k)\text{-resistant}\},$$

and we take $\mu(f) = 0$ for constant f.

We have the following theorem:

Theorem 3.2.1 $\ell_3^*(f) \geq \deg(f)\mu(f)$.

The above theorem will follow from the following general result:

Theorem 3.2.2 *Suppose $f(x_1, x_2, \ldots, x_n)$ is (d,r,k)-resistant, then*

$$\ell_3^*(f) \geq (r+1)\frac{k+1}{d+1}.$$

Proof. Consider a ΣΠΣ-formula that computes f. Remove all multiplication gates that have degree at most r. Doing so we obtain a ΣΠΣ formula $\mathcal{F}$ computing $f-g$, where g is some polynomial of degree at most r. Say $\mathcal{F}$ has s multiplication gates. Write:

$$f - g = \sum_{i=1}^{s} M_i,$$

where

$$M_i = \Pi_{j=1}^{d_i} l_{i,j},$$

and

$$l_{i,j} = c_{i,j,1}x_1 + c_{i,j,2}x_2 + \ldots + c_{i,j,n}x_n + c_{i,j,0}.$$

The degree of each multiplication gate in $\mathcal{F}$ is at least $r+1$, i.e. $d_i \geq r+1$, for each $1 \leq i \leq s$. Now select a set S of input linear forms using the following algorithm:

$S = \emptyset$
for $i = 1$ to s **do**
  repeat $d+1$ times:
    if($\exists j \in \{1, 2, \ldots, d_i\}$) such that $S^h \cup \{l_{i,j}^h\}$ is a set of independent vectors **then**
      $S = S \cup \{l_{i,j}\}$

Let A be the set of common zeroes of the linear forms in S. Since S^h is an independent set, by Lemma 3.1.1, A is affine linear of co-dimension $|S| \leq (d+1)s$.

Claim 3.2.3 *If at a multiplication gate M_i we picked strictly less than $d+1$ linear forms, then any linear form that was not picked is constant on A.*

Proof. Each linear form l that was not picked had l^h already in the span of S^h, for the set S built up so far. Hence we can write $l = c + l^h = c + \sum_{g \in S} c_g g^h$, for certain scalars c_g. Since each g^h is constant on A, we conclude l is constant on A. □

We conclude that for each multiplication gate at least one of the following holds:

1. $(d+1)$ input linear forms vanish on A, or
2. less than $(d+1)$ linear form vanishes on A, and all others are constant on A.

For each multiset X of size d with elements from $\{x_1, x_2, \ldots, x_n)$, the dth order partial derivative

$$\frac{\partial^d (f-g)}{\partial X} \tag{3.1}$$

is in the linear span of the set

$$\{\prod_{\substack{j=1 \\ j \notin J}}^{d_i} l_{ij} : \ 1 \leq i \leq s, J \subseteq \{1, 2, \ldots, d_i\}, |J| = d \ \}$$

Consider $1 \leq i \leq s$ and $J \subseteq \{1, 2, \ldots, d_i\}$ with $|J| = d$. If item 1 hold for multiplication gate M_i, then

$$\prod_{\substack{j=1 \\ j \notin J}}^{d_i} l_{ij} \tag{3.2}$$

vanishes on A, since there must be one l_{ij} that vanishes on A that was not selected, given that $|J| = d$. If item 2 holds for M_i, then (3.2) is constant on A.

Hence, we conclude that (3.1) is constant on A. Since f is (d,r,k)-resistant, we must have that the codimension of A is at least $k+1$. Hence $(d+1)s \geq k+1$. Since each gate in $\mathcal{F}$ is of degree at least $r+1$, we get that

$$\ell_3^*(\mathcal{F}) \geq (r+1)\frac{k+1}{d+1}.$$

Since $\mathcal{F}$ was obtained by removing zero or more multiplication gates from a $\Sigma\Pi\Sigma$-formula computing f, we have proven the statement of the theorem. □

To prove lower bounds on resistance, we supply the following lemma that uses the syntactic notion of affine restriction. In certain cases this will be convenient.

Lemma 3.2.4 *Over fields of characteristic zero, for any $d \leq r$, $k > 0$, and any polynomial $f(x_1, x_2, \ldots, x_n)$, if for every affine linear subspace A of codimension k, there exists some dth order partial derivative of f such that*

$$\deg(\left(\frac{\partial^d f}{\partial X}\right)_{|A}) \geq r - d + 1$$

then f is (d, r, k)-resistant.

Proof. Assume for every affine linear subspace A of codimension k, there exists some dth order partial derivative derivative of f such that

$$\deg(\left(\frac{\partial^d f}{\partial X}\right)_{|A}) \geq r - d + 1.$$

Let g be an arbitrary polynomial of degree r. Then

$$\begin{aligned} \left(\frac{\partial^d f - g}{\partial X}\right)_{|A} &= \left(\frac{\partial^d f}{\partial X} - \frac{\partial^d g}{\partial X}\right)_{|A} \\ &= \left(\frac{\partial^d f}{\partial X}\right)_{|A} - \left(\frac{\partial^d g}{\partial X}\right)_{|A}. \end{aligned}$$

The term $\left(\frac{\partial^d f}{\partial X}\right)_{|A}$ has degree at least $r - d + 1$, whereas the term $\left(\frac{\partial^d g}{\partial X}\right)_{|A}$ can have degree at most $r - d$. Hence $\deg(\left(\frac{\partial^d f - g}{\partial X}\right)_{|A}) \geq r - d + 1 \geq 1$. Since over fields of characterstic zero, syntactically different polynomials define different mappings, we conclude $\frac{\partial^d f - g}{\partial X}$ must be non-constant on A. □

Let us make the following important remark: taking partials does not commute with affine restrictions. For example, it is possible for all $\frac{\partial^d f}{\partial X}$ to vanish on A, but to have some $\frac{\partial^{d+1} f}{\partial X}$ to be non-constant on A. This is slightly counter-intuitive.

3.2.1 Applications

We will now prove lower bounds on the ΣΠΣ-formula size of several explicit polynomials.

Sum of Nth Powers Polynomial

Consider $f = \sum_{i=1}^n x_i^n$. For this polynomial we have ΣΠ-circuits of size $O(n \log n)$: for each variable x_i separately use $\approx \log n$ repeated multiplications to compute x_i^n and add up the results. This can be shown to be optimal using Strassen's degree method. By that method we know any circuit for f has size $\Omega(n \log n)$. The following section investigates lower bounds on ΣΠΣ-formula size for f. The obvious ΣΠΣ-formula has additive size n^2 wires in the top linear layer, and has n multiplication gates of degree n. We prove that this is essentially optimal.

Lemma 3.2.5 *Over fields of characteristic zero, the polynomial $f = \sum_{i=1}^{n} x_i^n$ has resistance factor $\mu(f) \geq n/2$.*

Proof. We will show that f is $(1, n-1)$-resistant. Let g be an arbitrary polynomial of degree $deg(f) - 1 = n - 1$. Letting $g_1, \ldots, g_n$ denote the first order partial derivatives of g, we get that the ith partial derivative of $f - g$ equal

$$nx_i^{n-1} - g_i(x_1, \ldots, x_n).$$

Note that the g_i's are of total degree at most $n-2$.

We claim there is no affine linear subspace of dimension greater than zero on which $\hat{f}$ is constant. To show this, it suffices to show that $\hat{f}$ is not constant on any affine line in F^n. Consider an arbitrary affine line, parameterized by a variable t:

$$x_i = c_i + d_i t,$$

where c_i and d_i are constants for all $i \in [n]$, and with at least one d_i nonzero. Then $\frac{\partial(f-g)}{\partial x_i}$ restricted to the line is given by

$$n(c_i + d_i t)^{n-1} - h_i(t),$$

for some univariate polynomials $h_i(t)$ of degree $\leq n-2$. Since there must exist *some* i such that d_i is nonzero, we know some partial derivative restricted to the affine line is parameterized by a univariate polynomial of degree $n-1$, and thus, given that the field is of characteristic zero, is not constant for all t. $\square$

Corollary 3.2.6 *Over fields of characteristic zero, any $\Sigma\Pi\Sigma$-formula for $f = \sum_{i=1}^{n} x_i^n$ has multiplicative size at least $n^2/2$.*

Proof. By Theorem 3.2.1, $\ell_3^*(f) \geq \deg(f)\mu(f)$. Applying Lemma 3.2.5, we get that $\ell_3^*(f) \geq n^2/2$. $\square$

In case the underlying field is the real numbers $\mathbf{R}$ and n is even, we can improve the above result to prove an absolutely tight n^2 lower bound. We start with the following lemma:

Lemma 3.2.7 *Let $f = \sum_{i=1}^{n} x_i^n$. Over the real numbers, if n is even, we have that for any affine linear subspace A of dimension $k \geq 1$, $deg(f_{|A}) = n$.*

Proof. Since f is symmetric we can assume without loss of generality that the following is a base representation of A:

$$\begin{aligned}
x_{k+1} &= l_1(x_1, \ldots, x_k) \\
x_{k+2} &= l_2(x_1, \ldots, x_k) \\
&\vdots \\
x_n &= l_{n-k}(x_1, \ldots, x_k).
\end{aligned}$$

Then

$$f_{|_A} = x_1^n + \dots x_k^n + l_1^n + \dots + l_{n-k}^n.$$

We conclude that $f_{|_A}$ must include the term x_1^n, since each l_j^n has a non-negative coefficient for the term x_1^n, since n is even. □

Theorem 3.2.8 *Over the real numbers, for even n, any* $\Sigma\Pi\Sigma$*-formula for* $f = \sum_{i=1}^n x_i^n$ *has multiplicative size at least* n^2.

Proof. Using Lemma's 3.2.4 and 3.2.7 we conclude that over the real numbers f is $(0, n-1)$-resistant. Hence, by Theorem 3.2.2 we get that $\ell_3^*(f) \geq \deg(f)\frac{n}{1} = n^2$. □

Let us note that $f = \sum_{i=1}^n x_i^n$ is an example of a polynomial that, even for large d, has relatively few, namely only n, partial derivatives. This makes application of the partial derivatives technique of [SW99], which we will describe and extend in the next section, problematic. Conversely, for polynomials that have many partial derivatives, in a sense to be made more precise, the technique of [SW99] can be more straightforward in its application than the resistance technique. The problem of analyzing precisely what is the minimal dimension of an affine linear space on which $f - g$ is non-constant can be quite hard for a given polynomial f and arbitrary g with degree less than $\deg(f)$.

Blocks of Powers

Suppose $n = m^2$ for some m. Consider the "m blocks of m powers" polynomial

$$f = \sum_{i=1}^{m} \prod_{j=(i-1)m+1}^{im} x_j^m.$$

The straightforward $\Sigma\Pi\Sigma$-formula for f, that computes each term/block using a multiplication gate of degree n, is of multiplicative size $n^{3/2}$. We will show this is tight.

Proposition 3.2.9 *The blocks of powers polynomial f defined above is* $(0, m-1)$*-resistant.*

Proof. Consider an affine linear space of codimension $m-1$. For any base B of A, restriction to A consists of substitution of the $m-1$ variables in B by linear forms in the remaining variables X/B. This means there is at least one term/block $B_i := \prod_{j=(i-1)m+1}^{im} x_j^m$ of f whose variables are disjoint from B. This block B_i remains the same under restriction to A. Also, for every other term/block there is at least one variable that is not assigned to. As a consequence, B_i cannot be cancelled against terms resulting from restriction to A of other blocks. Hence $\deg(f_{|_A}) = \deg(f)$. Hence by Lemma 3.2.4 we have that f is $(0, m-1)$-resistant. □

Corollary 3.2.10 *For the blocks of powers polynomial f defined above,* $\ell_3^*(f) \geq nm = n^{3/2}$.

Proof. Follows immediately from Theorem 3.2.2 and Proposition 3.2.9. □

Alternatively, one can observe that by substitution of a variable y_i for each variable appearing in the ith block one obtains from a $\Sigma\Pi\Sigma$-formula $\mathcal{F}$ for f a formula for $f' = \sum_{i=1}^{m} y_i^n$ of the same size as $\mathcal{F}$. Corollary 3.2.6 generalizes to show that $\ell_3^*(f') \geq \frac{1}{2}n^{3/2}$, which implies $\ell_3^*(f) \geq \frac{1}{2}n^{3/2}$.

Polynomials depending on distance to the origin

Over the real numbers, $x_1^2 + x_2^2 + \ldots + x_n^2$ is the Euclidean distance of the point $(x_1, x_2, \ldots, x_n)$ to the origin. Polynomials defined in terms of this distance can easily be seen to be of high resistance.

For example, consider $f = (x_1^2 + x_2^2 + \ldots + x_n^2)^m$. On any affine line L in $\mathbf{R}^n$, $\deg(f_{|L}) = 2m$. Therefore, by Lemma 3.2.4, over the reals, f is $(0, n-1)$-resistant. Hence by Theorem 3.2.2 we get that

Proposition 3.2.11 *Over the real numbers, $\ell_3^*((x_1^2 + x_2^2 + \ldots + x_n^2)^m) \geq 2mn$.*

Observe that by reduction this means that the "mth-power of an inner product polynomial", defined by $g = (x_1y_1 + x_2y_2 + \ldots + x_ny_n)^m$, must also have $\Sigma\Pi\Sigma$-size at least $2mn$ over the reals numbers.

Symmetric Polynomials

The special case of $(0,k)$-resistance implicitly appears in [Shp01], or at least in so far that the sufficient condition of Lemma 3.2.4 is used for the special case $d = 0$ in which no derivatives are taken. For the elementary symmetric polynomial S_n^r of degree $r \geq 2$ in n variables Theorem 4.3 of [Shp01] implies, using Lemma 3.2.4, that S_n^r is $(0, n - \frac{n+r}{2})$-resistant. Shpilka proves for $r \geq 2$, $\ell_3(S_n^r) = \Omega(r(n-r))$, which can be verified using Theorem 3.2.2: $\ell_3(S_n^r) \geq (r+1)(n - \frac{n+r}{2}) = \Omega(r(n-r))$. For $r = \Omega(n)$ this yields a tight $\Omega(n^2)$ bound as observed in [Shp01].

3.3 Bounds for +,*-Complexity

The partial derivatives technique of [SW99] ignores the wires of the formula present in the first layer. In the following we show how to account for them. As a result we get a sharpening of several lower bounds, though not on ℓ_3^* but on total formula size. The main idea is to utilize a closed form of the Baur-Strassen Derivative Lemma as one can derive it for $\Sigma\Pi\Sigma$-formulae. Let us describe this closed form here.

Consider a $\Sigma\Pi\Sigma$-formula $\mathcal{F}$ computing a polynomial p. Then one can write

$$p = \sum_{i=1}^{s} M_i,$$

where

$$M_i = \Pi_{j=1}^{d_i} l_{i,j},$$

and

$$l_{i,j} = c_{i,j,1}x_1 + c_{i,j,2}x_2 + \ldots + c_{i,j,n}x_n + c_{i,j,0}.$$

Here d_i is the in-degree of the ith multiplication gate, and $c_{i,j,k}$ is nonzero if and only if there is a wire from x_k to the addition gate computing $l_{i,j}$. Hence, using the addition and product rule for partial derivatives, we get for any k,

$$\begin{aligned} \frac{\partial p}{\partial x_k} &= \sum_{i=1}^{s}\sum_{p=1}^{d_i} \frac{\partial l_{i,p}}{\partial x_k} \prod_{\substack{j=1\\ j\neq p}}^{d_i} l_{i,j} \\ &= \sum_{i=1}^{s}\sum_{p=1}^{d_i} c_{i,p,k} \prod_{\substack{j=1\\ j\neq p}}^{d_i} l_{i,j}. \end{aligned} \tag{3.3}$$

We need a circuit-gadget $G(z_1, z_2, \ldots, z_d)$ that computes all d products of $d-1$ distinct input variables. Such a gadget can be constructed with size $O(d)$ many wires:

Proposition 3.3.1 *For each $d > 1$, there exists a circuit $G_d(z_1, z_2, \ldots, z_d)$ that consists of at most $3d-6$ multiplication gates and at most $6d-12$ wires that computes all $d-1$ products of $d-1$ distinct input variables.*

Proof. Let us construct G inductively. $G_2(z_1, z_2)$ is taken to consist of just the input variables z_1 and z_2. Suppose we have constructed G_d. Let g_i be the gate in G_d that computes $z_1 z_2 \ldots z_{i-1} z_{i+1} \ldots z_d$. Add a new input gate for variable z_{d+1}. Add a g gate that multiplies z_d and z_{d+1}. Peform the substitution $z_d = z_d \cdot z_{d+1}$ by replacing each wire going from z_d to a gate by a wire that goes from g to that gate. For $1 \leq j < d$, g_i now computes $z_1 z_2 \ldots z_{i-1} z_{i+1} \ldots z_d z_{d+1}$. The gate g_d computes $z_1 z_2 \ldots z_{d-1}$. Hence add a multiplication gate with input g_d and z_d and one with input g_d and z_{d+1} to compute the products "excluding z_d" and "excluding z_{d+1}". We added three multiplication gates and 6 wires in the induction, which proves the proposition. □

From the expression given for $\frac{\partial p}{\partial x_k}$ in (3.3), one can thus obtain a circuit that computes $(\frac{\partial p}{\partial x_1}, \frac{\partial p}{\partial x_2}, \ldots, \frac{\partial p}{\partial x_n})$ from $\mathcal{F}$ by first replacing each multiplication gate M_i, which has arity d_i, by a gadget G_{d_i} taking inputs $l_{i,1}, l_{i,2}, \ldots, l_{i,d_i}$. Then add an addition gate for each k of arity

$$\sum_{i=1}^{s} \sum_{\substack{p=1\\ c_{i,p,k}\neq 0}}^{d_i} 1$$

that computes $\frac{\partial p}{\partial x_k}$ according to (3.3). This layer is the mirror image of the layer computing the linear forms $l_{i,j}$: there is a wire going from variable x_k with constant c if and only if there is a wire with constant c going from the output of the ith gadget that excludes $l_{i,j}$ to the gate for $\frac{\partial p}{\partial x_k}$. (Seen as linear transformations these layers are each others transpose). We can conclude the resulting circuit for the partials has twice the number of wires fanning into addition gates, and by Proposition 3.3.1 has at most 6 times the number of wires fanning into multiplication gates.

When we utilize the above structural results, it turns out that the partial derivatives/affine restrictions technique factors through, allowing us to refine the [SW99] result for *-complexity:

Theorem 3.3.2 ([SW99]) *Let $f \in F[x_1,\ldots,x_n]$. Suppose for integers d, D, κ it holds that for every affine subspace A of co-dimension κ, $\dim(\partial_d(f)_{|A}) > D$. Then*

$$\ell_3^*(f) \geq \min(\frac{\kappa^2}{d}, \frac{D}{\binom{\kappa+d}{d}});$$

—to our result for +,*-complexity:

Theorem 3.3.3 *Let $f \in F[x_1,\ldots,x_n]$. Suppose for integers d, D, κ it holds that for every affine subspace A of co-dimension κ, $\sum_{i=1}^n \dim[\partial_d(\frac{\partial f}{\partial x_i})_{|A}] > D$. Then*

$$\ell_3(f) \geq \min(\frac{\kappa^2}{d+2}, \frac{D}{\binom{\kappa+d}{d}}).$$

Comparing the two theorems, we see that the result by Shpilka and Wigderson provides a lower bound on multiplicative complexity, while our result gives a lower bound on the *total* number of wires. We do get an extra "factor n" of additions with the $\sum_{i=1}^n \dim[\partial_d(\frac{\partial f}{\partial x_i})_{|A}] > D$ condition compared to just $\dim(\partial_d(f)_{|A}) > D$. Potentially this can lead to improved lower bounds on the *total* size of the formula, better than one would be able to infer from the lower bound on *multiplicative* complexity of Theorem 3.3.2 alone. We shall see that we can indeed get such kinds of improvements in the applications section below.

We employ the following suite of concepts and lemmas from [SW99] directly. We include proofs for completeness in case they are fairly short.

Definition 3.3.1 ([SW99]). For $f \in F[x_1,\ldots,x_n]$, let $\partial_d(f)$ be the set of all dth order *formal* partial derivatives of f w.r. to variables from $\{x_1,\ldots,x_n\}$.

For a multiset X of d variables, for any polynomial f, denote the d-th derivative of f by variables X by $\frac{\partial^d f}{\partial X}$. Then

$$\partial_d(f) = \{\frac{\partial^d f}{\partial X} : X \text{is a multiset of } d \text{ variables} \in \{x_1,x_2,\ldots,x_n\}\}.$$

For a set of polynomials $A = \{f_1,\ldots,f_t\}$, let $\text{span}(A) = \{\sum_{i=1}^t c_i f_i \mid c_i \in F\}$, i.e., span$(A)$ is the linear span of A. We write $\dim[A]$ as shorthand for $\dim[\text{span}(A)]$. We have the following elementary sub-additivity property for the measure $\dim[\partial_d(f)]$.

Proposition 3.3.4 ([SW99]) *For $f_1, f_2 \in F[x_1, \ldots, x_n]$ and constants $c_1, c_2 \in F$,*

$$\dim[\partial_d(c_1 f_1 + c_2 f_2)] \leq \dim[\partial_d(f_1)] + \dim[\partial_d(f_2)].$$

Proof. By the addition rule for (formal) partial derivatives:

$$\frac{\partial^d c_1 f_1 + c_2 f_2}{\partial X} = c_1 \frac{\partial^d f_1}{\partial X} + c_2 \frac{\partial^d f_2}{\partial X}$$

Hence each basis vector in $\partial_d(c_1 f_1 + c_2 f_2)$ is in the span of $\partial_d(f_1) \cup \partial_d(f_2)$. Since for vector spaces A and B, $\dim(\text{span}(A \cup B)) \leq \dim(A) + \dim(B)$, we get the statement. $\square$

One also needs to bound the growth of $\dim[\partial_d(f)]$ in case of multiplication. For multiplication of affine linear forms, we have the following two bounds.

Proposition 3.3.5 ([SW99]) *Let $M = \Pi_{i=1}^m l_i$, where each l_i is affine linear. Then*

$$\dim[\partial_d(M)] \leq \binom{m}{d}.$$

Proof. $\text{span}(\partial_d(M)) \subset \text{span}\ \{\Pi_{i \in S} l_i \mid S \subset [m], |S| = m - d\}$. $\square$

For a product $M = \Pi_{i=1}^t l_i$ of affine linear forms, we define M^h to be the set $\{l_1^h, \ldots, l_t^h\}$ of strictly linear parts of its input linear forms.

Proposition 3.3.6 ([SW99]) *Let M be a product gate with $\dim[M^h] = m$, then for any d,*

$$\dim[\partial_d(M)] \leq \binom{m+d}{d}.$$

Proof. Let $l_1, l_2, \ldots, l_m$ be a set of input linear forms for which $\{l_1^h, \ldots, l_m^h\}$ are independent. Then any other input linear form r_j of M is an affine linear combination $r_j = a_{0,j} + a_{1,j} l_1 + a_{2,j} l_2 + \ldots + a_{m,j} l_m$. We have

$$M = \prod_{i=1}^m l_i \cdot \prod_{j=1}^k (a_{0,j} + a_{1,j} l_1 + a_{2,j} l_2 + \ldots + a_{m,j} l_m) = p(l_1, l_2, \ldots, l_m)$$

for some polynomial $p(y_1, y_2, \ldots, y_m)$. Hence by the chain rule, and the fact that any $\frac{\partial l_i}{\partial x_j}$ is a constant, we can see that the set of all dth-order derivatives of M is contained in the linear span of

$$\{ \left(\frac{\partial^d p}{\partial^{d_1} y_1 \partial^{d_2} y_2 \ldots \partial^{d_m} y_m} \right)_{|A} : \text{for any } d_i \geq 0 \text{ with } d_1 + d_2 + \ldots d_m = d \},$$

where "$|_A$" is the substitution $y_1 = l_1, y_2 = l_2, \ldots, y_m = l_m$. Since there are $\binom{m+d}{d}$ ways of writing d as a sum of m non-negative integers, we get the result. $\square$

Note that for polynomials $f_1, \ldots, f_s$, $\text{span}(f_1, \ldots, f_s)_{|A} = \text{span}(f_{1|A}, \ldots, f_{s|A})$, and that $\dim[W_{|A}] \leq \dim[W]$. Now we modify Proposition 3.3.4 a little to get a result implicitly used by Shpilka and Wigderson in their arguments.

Proposition 3.3.7 (cf. [SW99]) *For $f_1, f_2 \in F[x_1, \ldots, x_n]$ and constants $c_1, c_2 \in F$, and affine linear subspace A, we have that* $\dim[\partial_d(c_1 f_1 + c_2 f_2)_{|A}] \leq \dim[\partial_d(f_1)_{|A}] + \dim[\partial_d(f_2)_{|A}]$.

Proof. By the addition rule for (formal) partial derivative and by the fact that substitution is a homeomorphism one gets that

$$\frac{\partial^d c_1 f_1 + c_2 f_2}{\partial X}_{|A} = c_1 \frac{\partial^d f_1}{\partial X}_{|A} + c_2 \frac{\partial^d f_2}{\partial X}_{|A}.$$

Hence each basis vector in $\partial_d(c_1 f_1 + c_2 f_2)_{|A}$ is in the span of $\partial_d(f_1)_{|A} \cup \partial_d(f_2)_{|A}$. Since for vector spaces A and B, $\dim(\mathrm{span}(A \cup B)) \leq \dim(A) + \dim(B)$, we get the statement. □

Finally, we require:

Lemma 3.3.8 ([SW99]) *For every n, κ, d, and every affine subspace A of co-dimension κ, we have that*

$$\dim[\partial_d(S_n^{2d})_{|A}] \geq \binom{n-\kappa}{d}.$$

Proof. The polynomial S_n^{2d} is multilinear, so only dth-order derivatives with respect to d distinct variables $D = \{x_{i_1}, x_{i_2}, \ldots, x_{i_d}\}$ will be potentially non-zero. Let X be the set of all n variables x_i. Observe that

$$\frac{\partial^d S_n^{2d}(X)}{\partial D} = S_{n-d}^d(X/D).$$

From [Got66] as used by [SW99] one has that

$$\mathrm{span}(\{S_{n-d}^d(X/D) : \text{for all subsets } D \subset X \text{ of size } d\})$$

has as basis the set of all multilinear monomials in variables X of degree d. There are $\binom{n-\kappa}{d}$ such monomials that are unchanged under the restriction $|_A$, which gives the result.

□

Now we can prove our sideways improvement of Shpilka and Wigderson's main Theorem 3.1 [SW99].

Proof of Theorem 3.3.3. Consider a minimum-size $\Sigma\Pi\Sigma$-formula for f with multiplication gates $M_1, \ldots, M_s$. We have that

$$f = \sum_{i=1}^{s} M_i,$$

where for $1 \leq i \leq s$,

$$M_i = \Pi_{j=1}^{d_i} l_{i,j}$$

with

$$l_{i,j} = c_{i,j,1}x_1 + c_{i,j,2}x_2 + \ldots + c_{i,j,n}x_n + c_{i,j,0},$$

for certain constants $c_{i,j,k} \in F$. Computing the partial derivative of f w.r. to variable x_k we get

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^{s} \sum_{j=1}^{d_i} c_{i,j,k} \frac{M_i}{l_{i,j}}. \tag{3.4}$$

Let
$$S = \{i : \dim[M_i^h] \geq \kappa\}.$$
If $|S| \geq \frac{\kappa}{d+2}$, then $\ell_3(f) \geq \frac{\kappa^2}{d+2}$. Suppose $|S| < \frac{\kappa}{d+2}$. If $S = \emptyset$, then let A be an arbitrary affine subspace of co-dimension κ. Otherwise, construct an affine space A as follows. Since $|S|(d+2) < \kappa$, and since for each $j \in S$, $\dim[M_i^h] \geq \kappa$, 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 $|S|(d+2) < \kappa$ independent homogeneous linear forms. Define
$$A = \{x : \; l_{i,j}(x) = 0, \text{ for any } i \in S, \; j \in [d+2]\}.$$
By Lemma 3.1.1, we have that the co-dimension of A is at most κ. W.l.o.g. assume the co-dimension of A equals κ. For each $i \in S$, d+2 linear forms of M_i vanish on A. This implies that
$$\dim[\partial_d(\frac{M_i}{l_{i,j}})_{|A}] = 0.$$
for any $i \in S$. For any $i \notin S$, by Proposition 3.3.6,
$$\dim[\partial_d(\frac{M_i}{l_{i,j}})_{|A}] < \binom{\kappa+d}{d}.$$
Let $D_k = \dim[\partial_d(\frac{\partial f}{\partial x_k})_{|A}]$. By Proposition 3.3.7 and equation (3.4),
$$D_k \leq \sum_{i \notin S} \sum_{\substack{j \\ c_{i,j,k} \neq 0}} \dim[\partial_d(\frac{M_i}{l_{i,j}})_{|A}].$$
Hence there must be at least $\frac{D_k}{\binom{\kappa+d}{d}}$ terms on the r.h.s., i.e. there are at least that many wires from x_k to gates in the first layer. Hence in total the number of wires to the first layer is at least $\sum_{i=1}^n \frac{D_i}{\binom{\kappa+d}{d}} > \frac{D}{\binom{\kappa+d}{d}}$. □

We can apply a similar idea to adapt the other main theorem from [SW99]:

Theorem 3.3.9 ([SW99]) *Let $f \in F[x_1, \ldots, x_n]$. Suppose for integers d, D, κ it holds that for every affine subspace A of co-dimension κ, $\dim(\partial_d(f_{|A})) > D$. Then for every $m \geq 2$,*
$$\ell_3^*(f) \geq \min(\kappa m, \frac{D}{\binom{m}{d}}).$$

We get:

Theorem 3.3.10 *Let $f \in F[x_1, \ldots, x_n]$. Suppose for integers d, D, κ with $d \geq 1$, it holds that for every affine subspace A of co-dimension κ, $\sum_{i=1}^n \dim[\partial_d(\frac{\partial f}{\partial x_i}{}_{|A})] > D$. Then for every $m \geq 2$,*
$$\ell_3(f) \geq \min(\frac{1}{2}\kappa m, \frac{D}{\binom{m-1}{d}}).$$

Proof. Consider a minimum size $\Sigma\Pi\Sigma$-formula for f with multiplication gates $M_1, \ldots, M_s$. We have that

$$f = \sum_{i=1}^{s} M_i,$$

where for $1 \leq i \leq s$,

$$M_i = \Pi_{j=1}^{d_i} l_{i,j},$$

with

$$l_{i,j} = c_{i,j,1}x_1 + c_{i,j,2}x_2 + \ldots + c_{i,j,n}x_n + c_{i,j,0}.$$

If there are $\frac{\kappa}{2}$ multiplication gates M_i of degree greater than m then already $\ell_3(f) > \frac{1}{2}\kappa m$. So suppose the number t of multiplication gates of degree greater than m is less than $\frac{\kappa}{2}$. Wlog. assume these gates are given by

$$M_1, M_2, \ldots, M_t.$$

For $i = 1, 2, \ldots, t$, pick two input linear forms $l_{i,1}, l_{i,2}$ of M_i, such that for the total collection $l_{1,1}, l_{1,2}, \ldots, l_{i,1}, l_{i,2}$ we have that the strictly linear parts $l^h_{1,1}, l^h_{1,2}, \ldots, l^h_{i,1}, l^h_{i,2}$ are independent. It might be that at some $i \leq t$, we cannot find any $l_{i,1}$ or $l_{i,2}$ with $l^h_{i,1}$ or $l^h_{i,2}$ independent from the previously collected linear forms. In this case, we just pick $l_{i,1}$ if that one is still independent, and skip to the next index i. If we can't even find $l_{i,1}$ for which $l_{i,1}$ is independent, we pick no linear form and proceed to the next i.

Let A be the zero set of all the collected input linear forms. Then A has co-dimension at most κ, by Lemma 3.1.1. Wlog. we may assume that the co-dimension of A equals κ. Observe that

$$\frac{\partial f}{\partial x_k}_{|A} = \sum_{i=1}^{s} \sum_{j=1}^{d_i} c_{i,j,k} (\frac{M_i}{l_{i,j}})_{|A}. \tag{3.5}$$

Now for a multiplication gate M_i of degree $\geq m$, there are three cases: either we picked two input linear forms of M_i, or we picked just one, or none at all. In the first case,

$$(\frac{M_i}{l_{i,j}})_{|A} = 0$$

in the r.h.s. of (3.5), for all i, j. In the second and third case, we know that for every input l of M_i that was not picked, l^h is a linear combination of l^h_i's for l_i's that were picked. Hence

$$l^h_{|A} = \sum_{i=1}^{r} c_i (l^h_{i\,|A}) = \text{constant}.$$

As a consequence, $(\frac{M_i}{l_{i,j}})_{|A} =$ constant in the r.h.s. of (3.5), for all i, j. Since $d \geq 1$, in either three cases, we obtain that $\partial_d(\frac{M_i}{l_{i,j}}_{|A}) = 0$. For multiplication gates M_i of degree at most m, Proposition 3.3.5 gives us that $\dim[\partial_d((\frac{M_i}{l_{i,j}})_{|A})] \leq \binom{m-1}{d}$. Let $D_k = \dim[\partial_d(\frac{\partial f}{\partial x_k}_{|A})]$. By Proposition 3.3.4, we see there are at least $D_k / \binom{m-1}{d}$ terms in (3.5). This implies that there are at least that many wires fanning out of x_k. Adding up for all variables, we conclude that $\ell_3(f) \geq D / \binom{m-1}{d}$. □

3.3.1 Some Applications

In [SW99] it was proved that for $d \leq \log n$, $\ell_3^*(S_n^{2d}) = \Omega(\frac{n^{\frac{2d}{d+2}}}{d})$. Note for $d = 2$, this lower bound is only $\Omega(n)$. We can apply Theorem 3.3.3 to prove the following stronger lower bound on the total formula size of S_n^{2d}. In particular for $d = 2$, we get an $\Omega(n^{\frac{4}{3}})$ bound.

Theorem 3.3.11 *For* $1 \leq d \leq \log n$, $\ell_3(S_n^{2d}) = \Omega(\frac{n^{\frac{2d}{d+1}}}{d})$.

Proof. For any affine subspace A of co-dimension κ and $d \geq 2$ we have that

$$\sum_{i=1}^{n} \dim[\partial_{d-1}(\frac{\partial S_n^{2d}}{\partial x_i})_{|A}] \geq \dim[\partial_d(S_n^{2d})_{|A}] \geq \binom{n-\kappa}{d}.$$

The latter inequality follows from Lemma 3.3.8. Applying Theorem 3.3.3 we get that

$$\ell_3(S_n^{2d}) \geq \min(\frac{\kappa^2}{d+1}, \frac{\binom{n-\kappa}{d}}{\binom{\kappa+d-1}{d-1}}) = \min(\frac{\kappa^2}{d+1}, \frac{\binom{n-\kappa}{d}}{\binom{\kappa+d}{d}}\frac{\kappa+d}{d}). \tag{3.6}$$

Set $\kappa = \frac{1}{9}n^{\frac{d}{d+1}}$. Then we have that

$$\begin{aligned}
\frac{\binom{n-\kappa}{d}}{\binom{\kappa+d}{d}}\frac{\kappa+d}{d} &\geq (\frac{n-\kappa}{\kappa+d})^d\frac{\kappa+d}{d} \\
&\geq (\frac{8/9n}{2/9n^{\frac{d}{d+1}}})^d\frac{\kappa+d}{d} \\
&= 4^d n^{\frac{d}{d+1}}\frac{\kappa+d}{d} \\
&\geq \frac{4^d}{9d}n^{\frac{2d}{d+1}} \\
&\geq n^{\frac{2d}{d+1}}.
\end{aligned}$$

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

Corollary 3.3.12 $\ell_3(S_n^4) = \Omega(n^{4/3})$.

Another function considered in [SW99] is the product of inner-product function. For two inner-products, i.e. $4n$ variables, it is defined by

$$PIP_n^2 = (\sum_{j=1}^{n} a_j b_j)(\sum_{i=1}^{n} c_i d_i).$$

Note the lower bound in [SW99] on PIP_n^d for the special case $d = 2$ is $\Omega(n)$. We can prove a non-linear lower bound for this function as follows.

Set $d = 1, \kappa = n^{2/3}$. Observe that $\frac{\partial^2 PIP_n^2}{\partial a_i c_j} = b_i d_j$. Let A be any affine subspace of co-dimension 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), $\frac{\partial^2 PIP_n^2}{\partial a_i c_j}_{|A} = \frac{\partial^2 PIP_n^2}{\partial a_i c_j}$. These are independent terms, hence $\dim[\partial_2 (PIP_n^2)_{|A}] \geq (n-\kappa)^2$. Observe the fact that for any $f(x_1, \ldots, x_n)$ and any affine subspace A we have that

$$\sum_{i=1}^{n} \dim[\partial_d (\frac{\partial f}{\partial x_i})_{|A}] \geq \dim[\partial_{d+1}(f)_{|A}].$$

Applying Theorem 3.3.3 we get that $\ell_3 PIP_n^2 \geq \min(\frac{n^{4/3}}{3}, \frac{(n-n^{2/3})^2}{n^{2/3}+1}) = \Omega(n^{4/3})$. We have proved:

Theorem 3.3.13 $\ell_3(PIP_n^2) = \Omega(n^{4/3})$.

More generally, we can apply Theorem 3.3.10 to obtain improved exponent for lower bounds on PIP_n^d. We define over $2d$ variable sets of size n (superscript indicate different variables, each variable has degree one):

$$PIP_n^d = \prod_{i=1}^{d} \sum_{j=1}^{n} x_j^i y_j^i.$$

Theorem 3.3.14 *For any constant $d > 0$, $\ell_3(PIP_n^d) = \Omega(n^{\frac{2d}{d+1}})$.*

Proof. Let $f = PIP_n^d$. Essentially we have that

$$\frac{\partial f}{\partial x_j^i} = y_j^i PIP_n^{d-1},$$

where the PIP_n^{d-1} must be chosen on the appropriate variable set. Let A be an arbitrary affine linear subspace of codimension κ. Then

$$\begin{aligned}\sum_{i=1}^{d} \sum_{j=1}^{n} \dim[\partial_{d-1}(\frac{\partial f}{\partial x_j^i}|A)] &= \sum_{i=1}^{d} \sum_{j=1}^{n} \dim[\partial_{d-1}(y_j^i PIP_n^{d-1}|A)] \\ &\geq (dn-\kappa) \dim[\partial_{d-1}(PIP_n^{d-1}|A)]\end{aligned}$$

The last inequality follows because at least $dn-\kappa$ of the y-variables are not assigned to with the restriction to A. From Lemma 4.9 in [SW99] one gets

$$\dim[\partial_{d-1}(PIP_n^{d-1}|A) \geq n^{d-1} - 2^{2d-1}\kappa n^{d-2}.$$

Using Theorem 3.3.10 we get

$$\ell_3(f) \geq \min(\frac{\kappa^2}{2}, \frac{(dn-\kappa)(n^{d-1} - 2^{2d-1}\kappa n^{d-2})}{\binom{\kappa-1}{d-1}}).$$

Taking $\kappa = n^{\frac{d}{d+1}}$, one gets for constant d that

$$\ell_3(PIP_n^d) = \Omega(n^{\frac{2d}{d+1}}).$$

□

For comparison, in [SW99] one gets $\ell_3^*(PIP_n^d) = \Omega(n^{\frac{2d}{d+2}})$.

3.4 Conclusion—Possible Further Tools

We have taken some further steps after [SW99], obtaining an absolutely tight (rather than asymptotically so) n^2 multiplicative size bound for a natural function, and obtaining somewhat higher bounds on $+,*$-size for low-degree symmetric and product-of-inner-product polynomials. However, these may if anything enhance the feeling from [SW99] that most of the concepts being employed may go no further than quadratic for lower bounds. One cannot after all say that a function $f(x_1,\ldots,x_n)$ is nonvanishing on an affine-linear space of co-dimension more than n. The quest then is for a mathematical invariant that scales beyond linear with the number of degree-d-or-higher multiplication gates in the formula.

One tool that has so far disappointed comes from various forms of the *degree* notion used by Strassen [Str73a]. The gradient of the sum-of-nth-powers function, namely the regular mapping $(x_1^{n-1},\ldots,x_n^{n-1})$, has *algebraic degree* $d_a = (n-1)^n$ at each of its points in the range, and likewise the "mapping ideal" $\langle y_1 - x_1^{n-1},\ldots,y_n - x_n^{n-1}\rangle$ has *geometric degree* $(n-1)^n$ (see ch. 8 of [BCS97]), which is the highest possible for a degree-$(n-1)$ regular mapping. The attraction here is that the gradient of a multiplication gate, i.e. of a product $z_1,\ldots,z_m$, has algebraic degree only $m-1$, although its mapping ideal has exponential geometric degree $1+(m-2)2^{m-1}$. A ΣΠΣ formula with s multiplication gates of degrees d_i and total fan-in $N = \sum_{i=1}^{s} d_i$ can be decomposed as a composition of a linear map from F^n to F^N, then a vector of multiplications in variables $z_1,\ldots,z_N$, and then a singular linear transformation back to F^n. A similar decomposition holds for formulas computing the gradient (and higher derivatives) of the function. If the two linear maps did not affect the algebraic degree of the composition, then by the product rule for degree one would get the inequality

$$d_a \leq \prod_{i=1}^{s}(d_i - 1).$$

Upon finding a way to dispense with multiplication gates of degree less than n (or degree $o(n)$), similar to what we did in the proof of Theorem 3.3.3, this inequality would yield quadratic lower bounds on $\ell_3^*(f)$ for a great variety of functions f. Unfortunately the linear maps *do* affect the algebraic degree, and the inequality is false. In fact, our computer runs have found that random ΣΠΣ formulas consisting of one *-gate of fan-in n and some small number of binary multiplication gates already achieve the maximum possible algebraic degree. It is possible that deeper uses of algebraic/geometric degree may yield invariants that scale to exponential size, but the simple notion's failure to pass even the quadratic threshold is not promising.

Suspiciously absent in current lower bound techniques for $\Sigma\Pi\Sigma$-formulas are random restriction type arguments, whereas all the results for *Boolean* constant depth circuits of [Ajt83, FSS81, Yao85, Hås89] proceed using random restrictions. Note that Raz manages to use random restritions in conjunction with a partial derivatives based technique in his work on *multilinear* arithmetical formulas [Raz04a, Raz04b]. In any event, the search for stronger mathematical techniques to prove exponential lower bounds in the self-contained $\Sigma\Pi\Sigma$ formula case continues.

Chapter 4

Orbit of Bilinear Forms

The seminal motivation of this and the next two chapters is to remove a major restriction from notable recent lower bounds by Raz [Raz02] and Bürgisser-Lotz [BL02]. The work will be done exclusively over the field $\mathbf{C}$ of complex numbers. We are interested in borrowing the following set of concepts from representation theory, see for example [NS82]. Note also the work by Mulmuley and Sohoni [MS02], who have outlined an approach via geometric invariant theory to showing P $\neq$ NP and other questions, involving some of the same basic concepts.

Definition 4.0.1. Let G be a group and X be a complex linear space $\neq \{0\}$ and denote by $\mathrm{Lin}_{\mathbf{C}}(X)$ the set of all linear operators $X \to X$. A *group representation* is a mapping $T : G \to \mathrm{Lin}_{\mathbf{C}}(X)$ such that

1. $T(e) = id_X$, where id_X is the identity operator on X, and e is the identity of the group G.
2. for all $g_1, g_2 \in$G$, T(g_1 g_2) = T(g_1) \circ T(g_2)$.

We are interested in the special case where X is taken to be the vector space $\mathbf{C}[x_0, x_1, \ldots, x_{n-1}]_m$ of homogeneous polynomials of degree m in variables $x_0, x_1, \ldots, x_{n-1}$ over $\mathbf{C}$, and considering G to be a group of $n \times n$ invertible matrices under multiplication. Then for invertible matrix $E \in G$, we can define linear transformation $T(E)$ by mapping $f \in \mathbf{C}[x_0, x_1, \ldots, x_{n-1}]_m$ according to:

$$T(E)(f) = f(E^{-1}x).$$

In other words, for vector of variables $x = (x_0, x_1, \ldots, x_{n-1})^T$, mapping $T(E)$ is defined by performing the substitution

$$x_i := (E^{-1}x)_i. \quad \text{for each } i = 0, 1, \ldots, n-1,$$

on the polynomial f. This defines a linear transformation on X:

$$(\mu f + g)(E^{-1}x) = \mu f(E^{-1}x) + g(E^{-1}x),$$

for any constant μ, any homogeneous polynomials f and g of same degree and invertible matrix E. It also is a representation, for the identity matrix I, $T(I)$ is the identity map and for any two invertible matrices E and D,

$$\begin{aligned} T(DE)(f) &= f((DE)^{-1}x) \\ &= f(E^{-1}D^{-1}x) \\ &= T(D)f(E^{-1}x) \\ &= T(D)\circ T(E)f. \end{aligned}$$

For a homogeneous polynomial f and group representation of G as above the set

$$\{f(E^{-1}x) : E \in G\}$$

is called the *G-orbit* of f. More generally, for multi-output polynomial mappings given by a tuple of polynomials $\mathcal{F} = (f_1, f_2, \ldots, f_m)$, we define the G-orbit of $\mathcal{F}$ to be the set

$$\{(f_1(E^{-1}x), f_2(E^{-1}x), \ldots, f_m(E^{-1}x)) : E \in G\}.$$

We are interested in proving sweeping lower bounds on the arithmetical complexity of all polynomials $f(E^{-1}x)$ that appear in the G-orbit of some explicitly defined polynomial f (or more generally for a multi-output polynomial mapping $\mathcal{F}$), for certain matrix groups G. In particular, we will focus on bilinear multi-output mappings over disjoint variable sets $\{x_0, x_1, \ldots, x_{n-1}\}$ and $\{y_0, y_1, \ldots, y_{n-1}\}$. In that case it is more natural to let two matrices E and D act on the variables separately. We define:

Definition 4.0.2. Let E and D be $n \times n$ non-singular complex matrices, and let $x = (x_0, x_1, \ldots, x_{n-1})^T$ and $y = (y_0, y_1, \ldots, y_{n-1})^T$ be vectors of variables. 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. In Section 4.3 we also consider having an $n \times n$ matrix at the output gates.

If for a bilinear mapping $b(x, y)$ one proves that any orbit circuit that computes it requires size s, this means that for any invertible E and D, the polynomial $b(E^{-1}x, D^{-1}y)$ must have *regular* circuit size at least s. Namely, from the ordinary circuit:

$$\Gamma(x, y) = b(E^{-1}x, D^{-1}y)$$

we obtain an orbit circuit of size s by substitution:

$$\Gamma(Ex, Dy) = b(E^{-1}Ex, D^{-1}Dy) = b(x, y)$$

that computes b. In this sense, any of our results that follow prove *generic* lower bounds on entire families of polynomials. Even when we are forced to make further restrictions on the groups E and D are taken from, or even drop the entire group concept, and just consider sets of matrices, this should be kept in regard. The computational model may seem increasingly exotic this way, but from the point of view of proving *generic* lower bounds no such objection holds.

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). Actually, the above shows that taking G to be the group of all invertible diagonal matrices yields a model equivalent in power as the unbounded-coefficients case. In fact, we could take the matrices at the input to be constant multiples λI of the identity matrix, and multiply by the appropriate constants less than one to correct for this at the cost of adding n unary addition gates.

Note that in Chapter 6 we will establish some orbits model lower bounds relative to diagonal matrices for circular convolution.

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 $n^{1-\varepsilon}$ 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 4.0.1 *Bilinear circuits C of size s computing* $\mathrm{HCirc}(x)y$ *can be converted into "relaxed" SL_n-orbit circuits Γ of size $s+O(n)$ computing* $\mathrm{HCirc}(x)y$.

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 $\mathrm{HCirc}(x_1,\ldots,x_{n-2})(y_1,\ldots,y_{n-2})$. Finally use the free access to the untransformed inputs x_0 and y_0 to re-create $\mathrm{HCirc}(x)y$ as above, adding $2n$-many $*$ gates and $2n-1$ $+$ gates at the outputs. On products x_0y_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_iy_0. This yields the desired "relaxed" orbit bc-circuit Γ. □

The significance of the orbit model is threefold. Firstly, it is natural and bridges between the bounded coefficient and general cases. Secondly, it defeats the proof methods of Raz and Bürgisser-Lotz. Thirdly, the orbit model leads to cutting edge problems in Fourier theory, as we show.

The proofs in [Raz02, BL02] 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.

The backbone of our lower bound technique will be the same as in [Raz02, BL02]: to simplify the bilinear circuit into a linear circuit using the probabilistic method. The idea is to fix scalar values $a = (a_0, a_1, \ldots, a_{n-1})$ for $x = (x_0, x_1, \ldots, x_{n-1})$ such that the "x side" of the bilinear bc-circuit Γ, which computes linear forms say $\ell_1(x), \ell_2(x), \ldots, \ell_k(x)$, keeps the values $|\ell_1(a)|, |\ell_2(a)|, \ldots, |\ell_k(a)|$ "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.

Recall we defined the *cyclic convolution* $x \circ y$ of two n-vectors x, y as 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 $0 \leq k < n$. In terms of circulant matrices:

$$x \circ y = \mathrm{Circ}(x) y.$$

Our main focus in this and the next two chapters will be to establish orbit model lower bounds for this bilinear form. We conjecture:

Conjecture 1. For any two $n \times n$ matrices E and D with determinant equal to one, any bounded coefficient bilinear circuit Γ with $\Gamma(Ex, Dy) = x \circ y$ requires $\Omega(n \log n)$ gates.

We also believe that the statement of the conjecture holds for arbitrary matrices E and D, but obtaining unbounded constant lower bounds seems hard with known techniques, whereas the above conjecture seems to lie within our present reach. The conjecture is equivalent to asserting that any bilinear map in the $SL_n(\mathbf{C})$ orbit of $x \circ y$ requires bounded coefficient circuit size $\Omega(n \log n)$. One must be careful here, for example one cannot prove an $SL_n(\mathbf{C})$-orbit lower bound for the tri-linear form $p(x, y, z) = z^T \mathrm{Circ}(x) y$. Namely there exists a polynomial in the $SL_n(\mathbf{C})$-orbit of p that has linear size! By Theorem 2.1.4:

$$z^T \mathrm{Circ}(x) y = z^T F_n \mathrm{diag}(DFT_n x) F_n^* y$$

so if we substitute $z^T := z^T F_n^*$, $y := F_n y$, $x := F_n x$, we get the polynomial

$$\sqrt{n} z^T \mathrm{diag}(x) y = \sqrt{n} \sum_{i=0}^{n-1} z_i x_i y_i.$$

This polynomial can be computed by a circuit of size $3n + O(\log\sqrt{n}) = O(n)$, by computing each of the n terms $x_i y_i z_i$ and add these, and next using $O(\log\sqrt{n})$ repeated additions to multiply by $\sqrt{n}$. The key point in this example is that we are dealing with a single output circuit. For example, using the repeated addition trick to multiply n outputs of a circuit by $\sqrt{n}$ would cost $\Theta(n\log n)$ in size, since you have to repeat for each output individually.

4.1 Definitions and Background

We next introduce some of the required concepts. We will provide proofs for completeness in case they are short.

4.1.1 Standard Gaussian vectors

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, see e.g. [BL02].

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 for U. This representation is independent of the choice of orthonormal basis.

We will use the following two Lemmas from Bürgisser and Lotz [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 4.1.1 ([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 $\|f\|^2$. 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 4.1.2 ([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 Σ of z to be entry-wise expectation of the outer product zz^*, i.e. $\Sigma = E[zz^*]$. Then we have*

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

for some absolute constant $\delta > 0$. More precisely, $\delta = 2^{-(\gamma+\sqrt{2\phi})}$ with $\gamma = \frac{1}{\sqrt{\pi}}\int_0^\infty t^{-\frac{1}{2}} e^{-t} \log t\,dt$, and $\phi = \frac{1}{2}\int_0^\infty e^{-\frac{t}{2}} \log^2 t\,dt$. (Here δ is approximately 0.02.)

4.1.2 Mean Square Volume & Matrix Rigidity

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. We let A_I stand for $A_{I,\{1,\ldots,n\}}$ and A^I for $A_{\{1,\ldots,m\},I}$. Pervasive in this work will be applications of the Binet-Cauchy Theorem, which states:

Theorem 4.1.3 (Binet-Cauchy Theorem) *Let A be an $m \times n$ matrix and let B be an $n \times m$ matrix with $n \geq m$. Then*

$$\det(AB) = \sum_{\substack{I \subseteq \{1,2,\ldots,n\} \\ |I|=m}} \det(A^I)\det(B_I).$$

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, BL03] 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.

Definition 4.1.1 ([Raz02, BL02]). Given $A \in \mathbf{C}^{m\times n}$, and r such that $1 \leq r \leq \min m,n$, define

$$\mathrm{vol}_r(A) = \max_{|I|=r}(\det(A_I A_I^*))^{1/2}, \tag{4.1}$$

$$\mathrm{vol}'_r(A) = \max_{|I|,|J|=r}(|\det(A_{I,J})|). \tag{4.2}$$

The centerpiece definition in [BL02, BL03], however, involves taking the Euclidean norm rather than the max-norm.

Definition 4.1.2 ([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

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

where I and J range over all r-subsets of $\{1,2,\ldots,n\}$.

These definitions are related by:

Lemma 4.1.4 ([BL02, BL03], respectively) *For A and r as above,*

$$\mathrm{vol}'_r(A) \leq \mathrm{msv}_r(A) \leq \binom{m}{r}^{1/2}\binom{n}{r}^{1/2}\mathrm{vol}'_r(A), \tag{4.3}$$

$$\mathrm{vol}_r(A) \leq \mathrm{msv}_r(A) \leq \binom{m}{r}^{1/2}\mathrm{vol}_r(A). \tag{4.4}$$

Proof. The inequalities given in (4.3) are immediate. For (4.4) use Theorem 4.1.3. □

An important fact is that mean square r-volume is invariant under unitary transformations. That is:

Proposition 4.1.5 *For $m \times n$ matrix A and any unitary matrices $U \in \mathbf{C}^{m \times m}$ and $V \in \mathbf{C}^{n \times n}$,*

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

Proof. By the Theorem 4.1.3:

$$\mathrm{msv}_r^2(A) = \sum_{|I|=r} \det(M_{I,I}),$$

where $M = AA^*$. Hence the right-side invariance $\mathrm{msv}_r(AV) = \mathrm{msv}_r(A)$, for any unitrary V is clear: $\mathrm{msv}_r^2(AV) = \sum_{|I|=r} \det(N_{I,I})$, for

$$N = (AV)(AV)^* = AVV^*A^* = AA^* = M.$$

For the left-side invariance, it is clear from the definition that for any matrix B,

$$\mathrm{msv}_r(B) = \mathrm{msv}_r(B^*).$$

Hence the left-side invariance follows from the right-side invariance by observing that

$$\mathrm{msv}_r(UA) = \mathrm{msv}_r((UA)^*) = \mathrm{msv}_r(A^*U^*) = \mathrm{msv}_r(A^*) = \mathrm{msv}_r(A).$$

□

So one can express $\mathrm{msv}_r(A)$ in terms of the *singular value decomposition* of A as folows. We first define:

Definition 4.1.3. The ith *singular value* $\sigma_i(A)$ is defined to be

$$\sigma_i(A) = \lambda_i(AA^*)^{1/2},$$

where $\lambda_i(AA^*)$ is the ith largest eigenvalue of AA^*.

The singular values of a matrix are non-negative real numbers. Recall the following theorem (See e.g. [Bha97]):

Theorem 4.1.6 (Singular Value Decomposition) *For any $m \times n$ matrix A with $m \geq n$, there exist unitary matrices $U \in \mathbf{C}^{m \times m}$ and $V \in \mathbf{C}^{n \times n}$, such that*

$$UAV = diag(\sigma_1, \sigma_2, \ldots, \sigma_n),$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$ are the singular values of A.

Hence by the unitrary invariance of mean-square-volume we have that

$$\begin{aligned}\mathrm{msv}_r^2(A) &= \mathrm{msv}_r^2(UAV)\\ &= \mathrm{msv}_r^2(\mathrm{diag}(\sigma_1,\sigma_2,\ldots,\sigma_n)\\ &= \sum_{|I|=r}\prod_{i\in I}\sigma_i^2\\ &= S_n^r(\sigma_1^2,\sigma_2^2,\ldots,\sigma_n^2),\end{aligned}$$

where S_n^r is the elementary symmetric polynomial of n variables of degree r.

There is also the following characterization of the singular values of a matrix (See [Bha97]):

Theorem 4.1.7 (Courant-Fisher minmax Theorem) *Let A be an $m\times n$ with $m\geq n$, matrix then for any $i=1,2,\ldots,n$,*

$$\sigma_i(A) = \max_{\substack{S\subseteq \mathbf{C}^n\\ dim(S)=i}}\ \min_{x\in S/\{0\}} \frac{||Ax||_2}{||x||_2},$$

where S ranges over all linear subspaces of dimension i.

From this it is immediately clear that for any matrix A, $\sigma_1(A)=||A||_2$. Also one has that $||A||_F^2=\sigma_1^2(A)+\sigma_2^2(A)+\cdots+\sigma_n^2(A)$.

As we have remarked above, msv_r is not preserved under transformations in $SL_n(\mathbf{C})$ (unless $r=n$). The following theorem states the use of the mean square volume measure for proving lower bounds.

Theorem 4.1.8 ([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 4.1.4 ([Raz02]). Let $A\in\mathbf{C}^{m\times n}$ be a matrix with with row vectors a_i, The *r-rigidity* of A is defined to be

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

where V ranges over all linear subspaces of $\mathbf{C}^n$, and $\mathrm{dist}(a,V)=\min_{v\in V}\|a-v\|_2$.

This notion relates to the r-volume measures defined above in the following way:

Lemma 4.1.9 ([Raz02]) *For any r, $vol_r(A)\geq\mathrm{rig}_r(A)^r$.*

If one considers an arbitrary topological sorting of the gates of a bounded coefficient linear circuit $f_1,f_2,\ldots,f_s$, then we can think of the f_i's defining an $s\times n$ matrix A. One can argue that each gate f_i can at most double the r-volume:

$$\mathrm{vol}_r(f_1,f_2,\ldots,f_i)\leq 2\cdot\mathrm{vol}_r(f_1,f_2,\ldots,f_{i-1}),$$

which implies $\mathrm{vol}_r(A)\leq 2^{s(\Gamma)}$, for any bounded-coefficient linear circuit Γ computing A. Combined with the above lemma this then yields the following theorem:

Theorem 4.1.10 ([Raz02]) *For $A \in \mathbf{C}^{m\times n}$, and $1 \le r \le m$, every linear bounded-constant circuit computing A has size at least $r \log \mathrm{rig}_r(A)$.*

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 4.1.11 ([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*

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

4.2 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.

Definition 4.2.1. 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 its 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 $\kappa(E) \le \kappa_1$.

Let us remark that well-conditioned matrices do *not* form a group. Unitary matrices have condition number 1, and do form a group. 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 the following easy proposition:

Proposition 4.2.1 *For any two $n \times n$ 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. Applying Theorem 4.1.6, 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

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

□

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 4.2.2 *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*

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

where c is some absolute constant.

Proof. By Theorem 2.1.4, we can write

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

where

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

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

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

where J ranges over all subsets of $\{1, \ldots, n\}$ of size r. By definition of standard Gaussian, we can write $a = V\beta$, 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

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

The last line follows because β is standard Gaussian distributed. We get that $\det(\Sigma) = |\det(W_J)|^2$. Applying Theorem 4.1.3 yields that

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

We *claim* now that $\det(V^*E^*EV) \geq \kappa_1^{-2r\frac{n-1}{n}}$, where $\kappa_1 > 0$ is a global constant. To prove the claim, observe that in terms of singular values $\sigma_i(EV)$ we have

$$\det(V^*E^*EV) = \prod_{i=1}^{r} \sigma_i(EV)^2.$$

By Theorem 4.1.7:

$$\sigma_r(EV) = \min_{||x||_2=1} ||EVx||_2.$$

Since V has orthonormal columns, for x with $||x||_2 = 1$, $||Vx||_2 = 1$. So for any x,

$$||EVx||_2 \geq \min_{||z||_2=1} ||Ez||_2 = \sigma_n(E).$$

For the matrix E we have

$$1 = \det(E^*E) = \prod_{i=1}^{n} \sigma_i(E)^2,$$

and by well-conditioning that $\frac{\sigma_1(E)}{\sigma_n(E)} \leq \kappa_1$, where κ_1 is an absolute constant. Hence we conclude that

$$\sigma_r(EV) \geq \sigma_n(E) \geq \kappa_1^{-\frac{n-1}{n}},$$

and hence that

$$\det(V^*E^*EV) \geq \kappa_1^{-2r\frac{n-1}{n}},$$

thus proving the claim.

Hence we conclude that there exists a set J such that

$$|\det(W_J)|^2 \geq \kappa_1^{-2r\frac{n-1}{n}} \tbinom{n}{r}^{-1}.$$

Applying Lemma 4.1.2 to the vector α_J, we get that with probability greater than $\frac{1}{2}$ that

$$\prod_{i \in J} |\alpha_i|^2 \geq \delta^r \det(\Sigma) \geq \delta^r \kappa_1^{-2r\frac{n-1}{n}} \tbinom{n}{r}^{-1},$$

where δ is an absolute constant. Hence

$$\mathrm{msv}_r^2(\mathrm{Circ}(Ea)) \geq n^r \delta^r \kappa_1^{-2r\frac{n-1}{n}} \tbinom{n}{r}^{-1} \geq n^r \delta^r \kappa_1^{-2r} 2^{-n}.$$

Hence by Proposition 4.2.1,

$$\mathrm{msv}_r^2(\mathrm{Circ}(Ea)D) \geq n^r \delta^r \kappa_1^{-4r} 2^{-n}.$$

Hence applying Theorem 4.1.8 we get:

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

where c is an absolute constant. □

Combining the above lemma with Lemma 4.1.11 in the same manner as in [BL02] yields the main theorem of this section.

Theorem 4.2.3 *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. 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 $\mathrm{Circ}(E^{-1}x)D^{-1}y$. 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 its corresponding linear form as a vector is $(c_1, \ldots, c_n)^T$. Let $R = \mathrm{rig}_{n-r}(f_1^T, \ldots, f_k^T)$. Observe that E^{-1} and D^{-1} have determinant 1 and are well-conditioned as well. By Lemmas 4.2.2 and 4.1.11, there exists an $a \in \mathbf{C}^n$ such that:

1. $s_{lin}^{bc}(\mathrm{Circ}(E^{-1}a)D^{-1}) \geq \frac{1}{2} r \log n - cn$, 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 $\mathrm{Circ}(E^{-1}a)D^{-1}$. 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 $\mathrm{Circ}(E^{-1}a)D^{-1}$. We can assume $k \leq n^2$, and by the rigidity bound of Theorem 4.1.10:

$$S(\Gamma) \geq s_{lin}^{bc}(f_1^T, \ldots, f_k^T) \geq (n - r) \log R - n. \tag{4.5}$$

So we obtain the inequality

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

which together with (4.5) yields $S(\Gamma) = \Omega(n \log n)$. □

To summarize, the main idea in the above proof 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.

4.3 Orbit circuits with exactly n multiplication gates

In previous sections we indicated why it is still difficult to prove super-linear lower bounds on $SL_n(\mathbf{C})$-orbits of natural functions, but we obtained such lower bounds when the matrices have bounded condition number. Now we show that if we restrict Γ to have only n multiplication gates, then a tight $\Omega(n \log n)$ lower-bound on the complexity of cyclic convolution applies, for arbitrary matrices in $SL_n(\mathbf{C})$ acting not only at the inputs but also at the outputs. Let $\times$ denote the entry-wise product of vectors, i.e. $(a \times b)_i = a_i b_i$, for each i.

Theorem 4.3.1 *For any $0 < \varepsilon < \frac{1}{2}$, for all but finitely many n, for any $n \times n$ matrices C, D, E such that*

$$E(Cx \times Dy) = \mathrm{Circ}(x)y,$$

one of the following conditions must hold:

1. *$|\det(C)|$ or $|\det(D)|$ is at least $n^{n(\frac{1-2\varepsilon}{4})}$, or*
2. *$|\det(E)|$ is at least $n^{\varepsilon n}$.*

We note that such a circuit exists, via Theorem 2.1.4. The proof works by showing that up to movable factors this representation is essentially unique.

Proof. Given that the range of $\mathrm{Circ}(x)y$ equals $\mathbf{C}^n$, we note that the matrix E must be non-singular. For $0 \le k \le n-1$, define an $n \times n$ matrix V^k by

$$(V^k)_{ij} = (C_i^T D_i)_{kj},$$

for $0 \le i, j \le n-1$, where C_i and D_i denote the ith row of C and D, respectively. Now we note some elementary lemmas:

Lemma 4.3.2 *For $0 \le k, j \le n-1$, $(EV^k)_j = e_{k+j \bmod n}$, where e_i denotes the ith standard basis vector, and $(EV^k)_j$ denotes the jth column of EV^k.*

Proof. $(EV^k)_j = E(V_j^k)$, where

$$V_j^k = \left((C_0^T D_0)_{kj}, \ldots, (C_{n-1}^T D_{n-1})_{kj}\right)^T.$$

For each i, $(C_i^T D_i)_{kj}$ is the coefficient of the term $x_k y_j$ computed at multiplication gate i, since there we compute polynomial $(C_i x)(D_i y)$. Hence $E(V_j^k)$ equals the n-vector of coefficients $(r_1, \ldots, r_n)$, where r_i equals the coefficient of $x_k y_j$ in $(\mathrm{Circ}(x)y)_i$, which in turn equals $e_{k+j \bmod n}$. □

Lemma 4.3.3 *For $0 \le k \le n-1$,*

$$V^k = E^{-1} \mathit{lshift}(I, k \bmod n),$$

where lshift(I,i) is the matrix obtained by wrap-around shifting the columns of I by i steps to the left.

Proof. Using Lemma 4.3.2, we get that

$$EV^k = e_{k \bmod n}, e_{k+1 \bmod n}, \dots, e_{k+n-1 \bmod n} = \mathit{lshift}(I, k \bmod n).$$

Since E must be invertible, we get the lemma. $\square$

Lemma 4.3.4 *For $0 \le k \le n-1$, $V^k = \mathit{lshift}(V^{k-1 \bmod n}, 1)$.*

Proof. Using the fact that $\mathit{lshift}(AB,i) = A \cdot \mathit{lshift}(B,i)$ and Lemma 4.3.3, we get that

$$\begin{aligned}
\mathit{lshift}(V^{k-1 \bmod n}, 1) &= \mathit{lshift}(E^{-1}\mathit{lshift}(I, k-1 \bmod n), 1) \\
&= E^{-1}\mathit{lshift}(\mathit{lshift}(I, k-1 \bmod n), 1) \\
&= E^{-1}\mathit{lshift}(I, k \bmod n) \\
&= V^k.
\end{aligned}$$

$\square$

Lemma 4.3.5 *For $0 \le i \le n-1$, the matrix $C_i^T D_i$ is a circulant matrix with*

$$(C_i^T D_i)_{st} = (C_i^T D_i)_{s-1 \bmod n, t+1 \bmod n},$$

for all $0 \le s,t \le n-1$.

Proof. Consider arbitrary $0 \le s,t \le n-1$. Then using Lemma 4.3.4 we get:

$$\begin{aligned}
(C_i^T D_i)_{st} &= V_{it}^s \\
&= V_{i,t+1 \bmod n}^{s-1 \bmod n} \\
&= (C_i^T D_i)_{s-1 \bmod n, t+1 \bmod n}.
\end{aligned}$$

$\square$

Proposition 4.3.6 *All entries of C and D must be nonzero.*

Proof. Suppose on row i of C there is a zero entry. Then $C_i^T D_i$ has one of its rows all zero. By Lemma 4.3.5 this implies that $C_i^T D_i$ has all entries zero. This means the output of the ith multiplication gate is always zero. Hence the output of the circuit is strictly contained in $\mathbf{C}^n$, which is a contradiction. For example, for $x = e_1$, $\mathrm{Circ}(x)y = Iy$. By symmetry, we conclude that also D must have all entries nonzero. $\square$

Lemma 4.3.7 *For $0 \le i \le n-1$, there exists an nth root of unity f such that for $0 \le j \le n-1$,*

$$C_{ij} = fC_{i,j+1 \ mod\ n}, \qquad and$$
$$D_{ij} = fD_{i,j+1 \ mod\ n}.$$

Proof. Observe that if some f satisfies the above, then $f^n = 1$. Fix $0 \le i \le n-1$. Let us use the short-hand c_j and d_j for the entries C_{ij} and D_{ij}, respectively, and we drop the mod n in the subscript, assuming all indexing is done mod n. By Lemma 4.3.5, for any s and t, and any number l, $c_s d_t = c_{s-l} d_{t+l}$. Fix $t = 0$ and $l = -1$. Since all entries of C and D are non-zero, we get for any s that

$$\frac{c_s}{c_{s+1}} = \frac{d_{n-1}}{d_0}.$$

Let $f = \frac{d_{n-1}}{d_0}$. For $0 \le j \le n-1$,

$$fc_{j+1} = \frac{c_j}{c_{j+1}} c_{j+1} = c_j.$$

Similarly we get for any $0 \le j \le n-1$, that

$$\frac{d_{j-1}}{d_j} = \frac{c_0}{c_{n-1}},$$

which implies the statement for the D_{ij}'s of the Lemma, and note the multiplier f is indeed the same for C and D. □

The above lemma tells us that for each row i there is a root of unity f and nonzero a_i and b_i so that $C_i = (a_i, fa_i, \ldots, f^{n-1}a_i)$, and $D_i = (b_i, fb_i, \ldots, f^{n-1}b_i)$. It is not too difficult to see that these multiplier must be distinct for different rows. Namely, if for $i \neq j$, rows i and j use the same multiplier, then $C_i = \lambda C_j$ and $D_i = \xi D_j$, for certain scalars λ and ξ. But then $(C_i x)(D_i y) = \lambda\xi(C_i x)(D_i y)$. In other words the ith and jth multiplication gate are restricted to be some fixed scalar multiple of each other. Hence the input to E is of dimension less than n, hence the output of the circuit has dimension $< n$, which is a contradiction.

So the full set $\{\omega^0, \omega, \ldots, \omega^{n-1}\}$ with $w = e^{2\pi i/n}$ is used. Without loss of generality we assume w^i is used for row i. Hence we get

$$\begin{aligned} C &= \mathrm{diag}(a_0, \ldots, a_{n-1}) DFT_n \\ D &= \mathrm{diag}(b_0, \ldots, b_{n-1}) DFT_n. \end{aligned} \tag{4.6}$$

From (4.6), and the fact that $\{r_i^T r_i : 0 \le i \le n-1\}$ is a linearly independent set with r_i equal the ith row of DFT_n, we obtain:

Proposition 4.3.8 *The set of polynomials $\{(Cx \times Dy)_i : 0 \le i \le n-1\}$ is linearly independent.*

The above proposition tells us that there is exactly one matrix E such that $E(Cx \times Dy) = \mathrm{Circ}(x)y$. It can be verified using Theorem 2.1.4 that the matrix

$$E = \frac{1}{n} DFT_n^* \mathrm{diag}(\frac{1}{a_0 b_0}, \ldots, \frac{1}{a_{n-1} b_{n-1}}).$$

fits the equation.

We now complete the proof of this section's main theorem. Let $\Delta = \mathrm{diag}(a_0 b_0, \ldots, a_{n-1} b_{n-1})$, $d_C = |\det(\mathrm{diag}(a_0, \ldots, a_{n-1}))|$, and $d_D = |\det(\mathrm{diag}(b_0, \ldots, b_{n-1}))|$. Then $|\det(\Delta)| = d_C d_D$. Fix $0 < \varepsilon < 1/2$. If $d_C \geq n^{-n(\varepsilon/2+1/4)}$, then

$$|\det(C)| = d_C |\det(DFT_n)| = d_C n^{n/2} \geq n^{n(1/4-\varepsilon/2)}.$$

Similarly, if $d_D \geq n^{-n(\varepsilon/2+1/4)}$, then $|\det(D)|$ is at least $n^{n(1/4-\varepsilon/2)}$. Otherwise $|\det(\Delta)|$ is at most $n^{-n(\varepsilon+1/2)}$. This implies that $|\det(E)|$ is at least $n^{\varepsilon n}$. □

As a corollary to the above theorem we get the following:

Corollary 4.3.9 *For any $n \times n$ matrices E, D and F with determinant equal to 1, any orbit circuit $F\Gamma(Ex, Dy)$ with exactly n multiplication gates computing* $\mathrm{Circ}(x)y$ *must have size at least* $\Omega(n \log n)$.

Proof. Let M_x and M_y be the linear maps computed at the input in the x and y variables, respectively, and let M_o be the linear map of the circuit at mapping the values from multiplication gates to output. These are all maps from $\mathbf{C}^n$ to $\mathbf{C}^n$. By Theorem 4.3.1, one of the three linear mappings, call it M, of the output circuit must have determinant of absolute value at least $n^{n/6}$. The map M can be written as a product of a determinant-1 matrix that does not count towards the circuit size, and another matrix N that is computed by gates. Hence using Theorem 2.1.1, the number of gates to compute N is at least $\log n^{n/6} = \Omega(n \log n)$. □

The above corollary implies a lower bound on bounded-coefficient complexity (when restricted to n multiplication gates) of the entire bilinear $SL_n(\mathbf{C})$-orbit of the mapping $\mathrm{Circ}(x)y$. Namely we have:

Corollary 4.3.10 *For any two $n \times n$ matrices E and D in $SL_n(\mathbf{C})$, the size of a bounded-coefficient circuit with n multiplication gates computing* $\mathrm{Circ}(Ex)Dy$ *must be* $\Omega(n \log n)$.

4.4 Orbits of ΣΠΣ- Formulae

In this section we extend our lower bounds from Chapter 3 to ΣΠΣ formulas with arbitrary linear transformations at the inputs. These linear transformations might themselves require n^2 formula size. More precisely, we consider orbit circuits of the form $C(Ex)$, where $E \in GL_n(\mathbf{C})$ and C is a ΣΠΣ-formula. To emphasize, constants on wires are unrestricted. Let $\ell_3^o(f)$ denote the smallest number of wires for a ΣΠΣ-formula C for which there exists invertible matrix E

such that $C(Ex) = f$. Regular ΣΠΣ-formula size, that is fixing E to be the identity map in the above, is denoted by $\ell_3(f)$.

Let us note that for polynomial f and affine subspace A of codimension κ, we can represent $f_{|A}$ by a substitution $f(Bx+b)$ for some matrix B of rank $n-\kappa$ and vector b. For a set of polynomials T, $\dim[\{t(Bx+b) : t \in T\}]$ is the same for all B of equal rank and fixed vector b.

Lemma 4.4.1 *Let $g \in \mathbf{C}[y_1,\dots,y_n]$ 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 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 $T \subseteq \mathrm{span}(S(Ex))$. Suppose the restriction to B is represented by substitution $(Bx+b)$. $E^{-1}B$ is also affine of codimension κ, and by the remark before this lemma,

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

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 4.4.2 *Let $f \in \mathbf{C}[x_1,\dots,x_n]$. 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*

$$\ell_3^o(f) \geq \min\left(\frac{\kappa^2}{d+2}, \frac{D}{\binom{\kappa+d}{d}}\right).$$

Proof. Suppose $f = C(Ex)$, where C is a ΣΠΣ formula with $\ell_3(f)$ many wires and E is some invertible matrix. Write Let $g = C(y)$. Observe that by Lemma 4.4.1 we have for any affine A of codimension κ,

$$\sum_{i=1}^{n} \dim[\partial_d(\frac{\partial g}{\partial y_i})_{|A}] \geq \dim[\partial_{d+1}(g)_{|A}] > D. \tag{4.7}$$

Let $M_1,\dots,M_s$ 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 $d_i = \text{fan-in}(M_i)$. 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. to variable y_k we get

$$\frac{\partial g}{\partial y_k} = \sum_{i=1}^{s}\sum_{j=1}^{d_i} c_{i,j,k}\frac{M_i}{l_{i,j}}. \tag{4.8}$$

Let $S = \{i | \dim(M_i^h) \geq \kappa\}$. If $|S| \geq \frac{\kappa}{d+2}$, then $\ell_3(f) \geq \frac{\kappa^2}{d+2}$. 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 $A = \{y | l_{i,j}(y) = 0, i \in S, j \in [d+2]\}$. By Proposition 3.1.1, we have that $\text{codim}(A) \leq \kappa$. Wlog. assume $\text{codim}(A) = \kappa$. For each $i \in S$, d+2 linear forms of M_i vanish on A. This implies that

$$\dim(\partial_d(\frac{M_i}{l_{i,j}})_{|A}) = 0.$$

For $i \notin S$, by Proposition 3.3.6,

$$\dim(\partial_d(\frac{M_i}{l_{i,j}})_{|A}) < \binom{\kappa+d}{d}.$$

Let $D_k = \dim(\partial_d(\frac{\partial g}{\partial y_k})_{|A})$. By equation (4.7), we have that $\sum_{k=1}^n D_k > D$. By Proposition 3.3.7 and equation (4.8),

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

Hence there must be at least $\frac{D_k}{\binom{\kappa+d}{d}}$ terms on the r.h.s., 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 $\sum_{i=1}^n \frac{D_i}{\binom{\kappa+d}{d}} > \frac{D}{\binom{\kappa+d}{d}}$. ☐

We compare the above with Theorem 3.3.3 and Shpilka and Wigderson's result that we quoted as Theorem 3.3.2. Let us define

$$\rho_{d,k}(f) = \min_{\text{codim}(A)=k} \dim[\partial_d(f)_{|A}]$$

Lemma 4.4.1 implies that for f in the $GL_n(\mathbf{C})$-orbit of g, i.e. $f = g(Ex)$, for some non-singular matrix E, that $\rho_{d,k}(f) = \rho_{d,k}(g)$. However, it does not hold in general that

$$\min_{\text{codim}(A)=k} \left(\sum_{i=1}^n \dim[\partial_d(\frac{\partial f}{\partial x_i})_{|A}] \right) = \min_{\text{codim}(A)=k} \left(\sum_{i=1}^n \dim[\partial_d(\frac{\partial g}{\partial x_i})_{|A}] \right).$$

This is the reason that we lose the "potential extra factor of n" arising from the summation in Theorem 3.3.3. Theorem 4.4.2 comes very close in its statement to Theorem 3.3.2. The only essential difference is the condition $\dim[\partial_d(f)_{|A}]$ versus $\dim[\partial_{d+1}(f)_{|A}]$. We will give an example in the applications that shows how this enables in certain cases for our Theorem 4.4.2 to outperform Theorem 3.3.2.

4.4.1 Lower Bounds

Theorem 4.4.3 *For* $1 \le d \le \log n$, $\ell_3^o(S_n^{2d}) = \Omega(\frac{n^{\frac{2d}{d+1}}}{d})$.

Proof. By Lemma 4.14 in [SW99] we have that 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 4.4.2 we get that

$$\begin{aligned}
\ell_3^o(S_n^{2d+2}) &\ge \min(\frac{\kappa^2}{d+2}, \frac{\binom{n-\kappa}{d+1}}{\binom{\kappa+d}{d}}) \\
&= \min(\frac{\kappa^2}{d+2}, \frac{\binom{n-\kappa}{d}}{\binom{\kappa+d}{d}} \frac{n-\kappa-d-1}{d+1}) \\
&\ge \min(\frac{\kappa^2}{d+2}, \frac{\binom{n-\kappa}{d}}{\binom{\kappa+d}{d}} \frac{n-2\kappa}{d+1})), \qquad (4.9)
\end{aligned}$$

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

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

Hence (4.9) is at least $\min(\frac{n^{\frac{2d+2}{d+2}}}{81(d+2)}, \frac{n^{\frac{2d+2}{d+2}}}{d+1}) = \Omega(\frac{n^{\frac{2d+2}{d+2}}}{d+2})$. □

Recall the product-of-inner-product polynomial:

$$PIP_n^2 = (\sum_{j=1}^{n} a_j b_j)(\sum_{i=1}^{n} c_i d_i).$$

We prove:

Theorem 4.4.4 $\ell_3^o(PIP_n^2) = \Omega(n^{4/3})$.

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 $\dim(\partial_2(PIP_n^2)_{|A}) \geq (n-\kappa)^2$. Applying Theorem 4.4.2 we get that $\ell_3^o(PIP_n^2) \geq \min(\frac{n^{4/3}}{3}, \frac{(n-n^{2/3})^2}{n^{2/3}+1}) = \Omega(n^{4/3})$. $\square$

The above is an example, where the different conditions of $\dim[\partial_d(f)_{|A}] > D$ versus $\dim[\partial_{d+1}(f)_{|A}] > D$ in the statements of Theorems 3.3.2 and 4.4.2 matter. Recall our previous remark that [SW99] yields only a trivial $\Omega(n)$ lower bound for this polynomial. More generally, we have the product of d inner products:

$$PIP_n^d = \prod_{i=1}^{d}(\sum_{j=1}^{n} a_j^i b_j^i),$$

for variables a_j^i, b_j^i with $i,j \in \{1,\ldots,n\}$.

Theorem 4.4.5 *For constant $d \geq 2$, $\ell_3^o(PIP_n^d) = \Omega(n^{\frac{2d}{d+1}})$.*

Compare with $\ell_3^*(PIP_n^d) = \Omega(n^{\frac{2d}{d+2}})$ in [SW99].

Theorem 4.4.6 $\ell_3(z^T Circ(x)y) = \Omega(n^{\frac{4}{3}})$.

Proof. Let $f = z^T\mathrm{Circ}(x)y$. Aply Theorem 3.3.3 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 $\ell_3(f) = \Omega(n^{4/3})$. $\square$

Note that $z^T\mathrm{Circ}(x)y$ 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 $O(n)$.

We cannot prove a non-linear lower bound on $\ell_3^o(z^T\mathrm{Circ}(x)y)$, because there exists a polynomial in the orbit of $z^T\mathrm{Circ}(x)y$ that has $O(n)$ $\Sigma\Pi\Sigma$-formula size! Namely, separately in each set of variables, apply DFT_n^{-1} to x, F_n to y and F_n^{-1} to z. By Theorem 2.1.4, $\mathrm{Circ}(x) = F_n\mathrm{diag}(\lambda)F_n^{-1}$ for $\lambda = DFT_n x$. Hence we get $z^T F_n^{-1}\mathrm{Circ}(DFT_n^{-1}x)F_n y = z^T\mathrm{diag}(x)y^T$.

The above is an example of a polynomial where the extra factor of n obtained by the summation in Theorem 3.3.3 matters: $\dim[\partial_2(f)_{|A}] = O(n)$, but $\sum_i \dim[\partial_1(\frac{\partial f}{\partial z_i})] \geq n(n-\kappa)$, for any affine A of codimension κ. This polynomial also provides us with a counter-example to the claim that for any $f = g(Ex)$,

$$\min_{\mathrm{codim}(A)=k}\left(\sum_{i=1}^{n}\dim[\partial_d(\frac{\partial f}{\partial x_i})_{|A}]\right) = \min_{\mathrm{codim}(A)=k}\left(\sum_{i=1}^{n}\dim[\partial_d(\frac{\partial g}{\partial x_i})_{|A}]\right).$$

If this were true, we could prove equally strong lower bounds for the $\Sigma\Pi\Sigma$-orbit model as obtainable with Theorem 3.3.3 for regular $\Sigma\Pi\Sigma$-formulas. However, this does not hold, and we had to weaken Theorem 3.3.3 somewhat, resulting in its analogy Theorem 4.4.2.

As a last application, we define

Definition 4.4.1. 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

$$LMM_d = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} (X^1 \cdot X^2 \ldots \cdot X^d)_{ij}.$$

We prove the following lower bound:

Theorem 4.4.7 *For constant $d \geq 1$, $\ell_3^o(LMM_{2d+1}) = \Omega(n^{4-\frac{4}{d+2}})$.*

Proof. 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 partial with respect to the 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 $\kappa = n^{\frac{2d+2}{d+2}}$ and applyingTheorem 4.4.2 yields the result. $\square$

4.5 Remarks

As a stepping stone towards proving lower bounds for unbounded constant circuits, we defined a computational model that allows for more unbounded constants than previously considered in the literature (e.g., see [BL02]), but that does this in some moderated sense. The model also serves the dual purpose of investigating the computational complexity of all mappings that are present in the G-orbit of a given bilinear map, for various matrix groups G under consideration. Given that taking $G = GL_n(\mathbf{C})$ results in a model that is at least as powerful as the unbounded constants case, the next natural thing we attempted was to lift the random substitutation technique of [BL02, Raz02] to the $SL_n(\mathbf{C})$-orbit model.

This turned out to be hard because of two conflicting issues. Namely, there is the apparent requirement in the random substitution technique to select the random input from a *subspace* U of some dimension εn with $\varepsilon < 1$, which seems to be about the only way to make the outputs of the linear forms on which substitution is performed "reasonably" bounded. Provided that is true, they can be replaced by "few enough" repeated additions, and this way a reduction to the (well understood) linear case is achieved. Unifying this modus operandi of the restriction technique with the abundance of ill-conditioned matrices present in $SL_n(\mathbf{C})$ is problematic.

Geometrically speaking only n-dimensional volumes retain the same volume under such transformation, but any lower dimensional volumes can be arbitrarily stretched or squashed. In any configuration of the argument we considered this becomes an issue. Either the msv_r-volume of the target linear form one reduces to is negatively impacted, or, attempting to salvage this, the outputs of linear forms on which one performs substitution are ill-behaved, and vice-versa.

We did manage to show that techniques from [BL02, Raz02] continue to stand while allowing the circuit to have for free at the inputs linear transformations in $SL_n(\mathbf{C})$ that have condition number $O(1)$. In particular unitary matrices present no problem. We also managed to show the desired result of proving an $\Omega(n \log n)$ size $SL_n(\mathbf{C})$-orbit model lower bound for circular convolution *assuming only n multiplication gates are used*.

We considered orbits in conjunction with $\Sigma\Pi\Sigma$-formulas. The fact that lower bounds for $*$-complexity are maintained unaltered under such an extension is trivial. Interestingly enough, we showed $+,*$-complexity lower bounds can also still be established, although in certain cases the results are weakened.

In the next two chapter we will focus on $DL_n(\mathbf{C})$-orbits, that is, allowing for free arbitrary *diagonal* matrices of determinant 1 at the inputs. Also these matrices can be arbitrary ill-conditioned, and hence will still provide a formidable proving ground.

We first will make an exposition of the complexity theoretic issues that are involved in interlude Chapter 5, outlining the global structure of the lower bound proof we are going to pursue. Then in Chapter 6 we will set up a framework that allows for a rigorous attack on the involved problems. Using Fourier analysis, in particular involving a discrete variant of the *Heisenberg uncertainty principle*, we will be able to establish some lower bounds for the circular convolution bilinear map. We will also establish a result about random Vandermonde matrices, and derive a circuit lower bound from that. Finally, some limitations will be explored using result known about the asymptotic eigenvalues of the *prolate matrix* [Sle78].

Chapter 5

Diagonal Orbits

Our aim is to extend the arguments by Raz and Bürgisser-Lotz [Raz02, BL02] with regard to the number of unbounded constants allowed in the circuit, and to give lower bounds on entire orbits $f(Dx,Ey)$, where f is a natural bilinear function like matrix multiplication or convolution, and D,E are matrices of unit determinant. We begin with the very special case where E is the identity and D is a diagonal matrix. Handling this case is not sufficient, but it brings out connections to major matrix problems about minors, in case of convolution about the discrete Fourier matrix DFT_n. The general two-sided case will be dealt with in Chapter 6. Accordingly, in this chapter and the next, we focus on circuits of the form

$$\Gamma_n(x_1 \cdot d_1^n, \ldots, x_n \cdot d_n^n, y_1, y_2, \ldots, y_n),$$

where $\{\Gamma_n\}_{n>0}$ is a family of bounded-coefficient bilinear circuits and

$$\{D_n = (d_1^n, d_2^n, \ldots, d_n^n)\}_{n>0},$$

is a family of n-tuples satisfying that for any n,

$$\prod_{i=1}^{n} d_i^n = 1.$$

These circuits compute bilinear mappings in the set of variables $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_n\}$. As done before for orbit circuits, for circuit size we only count the size of Γ_n. In other words, the constants d_i^n do not count against the size. They can be considered *unary helper gates*.

In this chapter we lay out the complexity theory side of the lower bound strategy. The next chapter attacks the mathematical problems involved, and establishes some lower bounds, and also indicates some limitations of the taken approach.

5.1 Strategy and Conditional Result

As in [BL02, Raz02], one of the inputs is going to be fixed by constants (we fix y), thereby reducing the bilinear case to a question about linear circuits. Once y is fixed, the outputs of the linear forms in y output constants that are used at the multiplication gates. These multiplications with constants can be replaced by performing repeated additions. In a way to be made more precise later, one can only do this, if the outputs of the linear forms in the y variables are "reasonably" bounded. If this is true, only few repeated addition will be needed, leaving the blow-up in size of the circuit limited. Also, with y fixed, the circuit computes a linear transformation in the x variables. If one manages to fix y so that the resulting linear map has provably high complexity, *while at the same time* leaving blow-up in size caused by the repeated addition to be limited, one would conclude the original circuit must have been of "high" complexity. For the purpose of bounding the magnitude of the linear forms when fixing y, we prove the following lemma.

Lemma 5.1.1 *Given $k \times n$ matrix F computed by b.c. linear circuit Γ with n inputs and k outputs, for all $0 \le l < n$, there exists $U \subseteq \mathbf{C}^n$ of co-dimension l such that for all $a \in U$*

$$\max_{1 \le i \le k} |(Fa)_i| \le ||a||_2 \cdot 2^{\frac{3s(\Gamma)+3n}{2l+2}}.$$

Proof. By the min-max charaterization of singular values (Theorem 4.1.7)

$$\sigma_n(F) = \min_{||a||_2=1} ||Fa||_2.$$

If $\sigma_n(F) < 1$, add n gates to the circuit that make a copy of the inputs. We obtain a circuit Γ' of at most $s(\Gamma)+n$ gates computing a $k' \times n$ matrix G with $\sigma_n(G) \ge 1$ and $k' \ge n$.

Consider G^*G. Using Theorem 4.1.3 (Binet-Cauchy), we get that

$$\det(G^*G) = \sum_{|T|=n} |\det(G_T)|^2 \le \binom{k'}{n} 2^{2s(\Gamma')}.$$

The last inequality follows from Morgenstern's Theorem 2.1.1. So

$$\det(G^*G) \le 2^{k'} 2^{2s(\Gamma')} \le 2^{3s(\Gamma')} \le 2^{3s(\Gamma)+3n}.$$

Also

$$\det(G^*G) = \prod_{i=1}^{n} \sigma_i(G)^2.$$

For arbitrary $0 \le l < n$,

$$\begin{aligned} \sigma_{l+1}(G)^{l+1} &\le \prod_{i=1}^{l+1} \sigma_i(G) \\ &\le \prod_{i=1}^{n} \sigma_i(G) \\ &\le 2^{\frac{3s(\Gamma)+3n}{2}}. \end{aligned}$$

So $\sigma_{l+1}(G) \leq 2^{\frac{3s(\Gamma)+3n}{2l+2}}$. By Theorem 4.1.7 (Courant-Fisher-Weyl min-max)

$$\sigma_{l+1}(G) = \min_{codim(U)=l} \max_{\text{unit } x \in U} ||Gx||_2.$$

Hence we conclude that there exists $U \subseteq \mathbf{C}^n$ of co-dimension l such that for all unit $x \in U$,

$$\max_{1 \leq i \leq k} |(Fx)_i| \leq \max_{1 \leq i \leq k'} |(Gx)_i| = ||Gx||_\infty \leq ||Gx||_2 \leq 2^{\frac{3s(\Gamma)+3n}{2l+2}}.$$

The statement of the lemma now follows by linearity. □

We compare the above with the proofs of Lemma 4.1 in [Raz02] and Lemma 4.2 in [BL03]. There the definition of rigidity is used to obtain a subspace U from which selecting a *random* input a yields a bound on the magnitudes of Fa with high probability. We obtain a subspace U such that *for all* unit $a \in U$ these magnitudes are bounded, alas with a slightly weaker bound. Namely, a standard Gaussian vector in an $n-l$ dimensional vector space has expected norm $\sqrt{n-l}$, but this factor crucially gets dampened in [BL03] and [Raz02]. Nevertheless, Lemma 5.1.1 will suffice for our purposes. We have the following conditional theorem:

Theorem 5.1.2 *Let $\{D_n = (d_1^n, d_2^n, \ldots, d_n^n)\}_{n>0}$ be a family of n-tuples satisfying that for any n, $\prod_{i=1}^n d_i^n = 1$. Suppose $\{\Gamma_n\}_{n>0}$ is a family of bounded-coefficient bilinear circuits such that for all n,*

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

Let

$$I_n = \{i : 0 \leq i \leq n-1 \text{ with } d_i^n < 1\},$$

and define $\ell_n = |I_n|$. If for every $\delta > 0$, there exists a $k_0 > 0$ so that for all but finitely many n, for any affine linear space U of codimension $\lfloor \frac{\ell_n}{k_0} \rfloor$, there exists $a \in U$ with $||a||_2 = 1$ such that Circ(a) *has an $\ell_n \times \ell_n$ minor M with rows I_n with*

$$|\det(M)| \geq 2^{-\delta n \log n},$$

then there exists $\gamma > 0$ such that for infinitely many n,

$$s(\Gamma_n) \geq \gamma n \log n.$$

Let us first make some preliminary remarks. Suppose that in the above $\ell_n = \Omega(n)$. This means there exists an $0 < \varepsilon_0 < 1$ so that for all but finitely many n, $\ell_n \geq \varepsilon_0 n$. In this case we think of the d_i^n that are larger than 1 as help gates as in [BL02]. There are at most $(1-\varepsilon_0)n$ many such help gates. Currently known techniques can already handle this amount of unbounded constants. Namely, Theorem 6.4 of [BL02] tells us that in this case $s(\Gamma_n) = \Omega(n \log n)$. The question we would like to address is whether we can manage to deal with $n - o(n)$ many unbounded constants in the circuit. This situation arises when $\ell_n = o(n)$.

Proof. (of Theorem 5.1.2) Wlog. we assume all d_i^n values are distinct. (If this is not true make infinitismal perturbations of the d_i^n and add n gates to correct these again.) For each n, let $i_{1,n},\dots,i_{n,n}$ be such that

$$d^n_{i_{1,n}} < d^n_{i_{2,n}} < d^n_{i_{\ell_n,n}} < 1 < d^n_{i_{\ell_n+1,n}} < \dots < d^n_{i_{n,n}}.$$

In case

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

then we can replace the constants which are bigger than 1 by bounded constant repeated additions. This takes at most $\sum_{j=\ell_n+1}^{n} \log d^n_{i_{jn}} = o(n\log n)$ 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 $s(\Gamma_n)=\Omega(n\log n)$.

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

$$\prod_{j=1}^{\ell_n} d^n_{i_{j,n}} < 2^{-\delta n\log n}, \tag{5.1}$$

Let us consider some large enough n for which (5.1) 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 $j>\ell$ and substitute $x_{i_j}=z_j/d_{i_j}$ otherwise. This yields a bounded coefficient bilinear circuit of size no bigger than $s(\Gamma)$, and it computes

$$(z_1,\dots,z_\ell)\mathrm{diag}(d_{i_1}^{-1},\dots,d_{i_\ell}^{-1})M,$$

where M is the $\ell\times n$ minor of $\mathrm{Circ}(y)$ corresponding to rows I_n.

Now set $r=n-\lfloor\frac{\ell}{k_0}\rfloor$, where k_0 is the constant that is assumed to exist by the statement of the theorem for $\frac{\delta}{2}$. Let $f_1,\dots,f_k$ be the linear forms in y of Γ. Lemma 5.1.1 provides us with a linear subspace U of dimension $n-\lfloor\frac{\ell}{k_0}\rfloor$ such that for any unit $a\in_R U$, we have that

$$\log\max_i |f_i(a)| \le \frac{3s(\Gamma_n)+3n}{2\lfloor \ell/k_0\rfloor+2}. \tag{5.2}$$

For any unit $a\in U$ and any $\ell\times\ell$ minor M_0 of $\mathrm{Circ}(a)$ with rows I we can obtain from Γ_n a bounded coefficient *linear* circuit computing the $\mathbf{C}^m\to\mathbf{C}^m$ map

$$(z_1,\dots,z_m)\mathrm{diag}(d_{i_1}^{-1},\dots,d_{i_\ell}^{-1})M_0,$$

by removing the outputs not corresponding to M_0 and replacing multiplications with $f_i(a)$ by $f_i(a)/\mu$ and correcting this by adding at most $\ell\log\mu$ repeated additions at the output gates, where $\mu = \max_i |f_i(a)|$. Hence the number of gates we added is at most

$$\begin{aligned} \ell \log \max_i |f_i(a)| &\leq \ell \frac{3s(\Gamma_n)+3n}{2\lfloor \ell/k_0 \rfloor + 2} \\ &\leq k_0 3s(\Gamma_n) + 3nk_0 \\ &\leq 4k_0 s(\Gamma_n). \end{aligned}$$

So the size of the resulting b.c. linear circuit is at most $(4k_0+1)s(\Gamma)$. However, by the condition of the theorem, and given that n is assumed to be large enough, the above can be done for a minor M_0 for which

$$|\det(M_0)| \geq 2^{-\frac{\delta}{2} n \log n},$$

This means that

$$|\det(\mathrm{diag}(d_{i_1}^{-1}, \ldots, d_{i_l}^{-1})M_0)| \geq 2^{\frac{\delta}{2} n \log n}.$$

However, by Morgenstern's bound (Theorem 2.1.1) any bounded coefficient circuit computing $\mathrm{diag}(d_{i_1}^{-1}, \ldots, d_{i_m}^{-1})M_0$ then requires at least $\frac{\delta}{2} n \log n$ gates. Hence $s(\Gamma_n) \geq \frac{\delta}{8k_0+2} n \log n$. □

5.2 Finding good minors

Using the notation of Theorem 5.1.2, and given our preliminary remark, we see that we are essentially left with establishing the following condition:

(Condition I) For every family $\{I_n \subseteq \{0, 1, \ldots, n-1\}\}_{n>0}$ with $\ell_n = |I_n| = o(n)$, and every $\delta > 0$, there exists a $k_0 > 0$ so that for all but finitely many n, for any affine linear space U of codimension $\lfloor \frac{\ell_n}{k_0} \rfloor$, there exists $a \in U$ with $||a||_2 = 1$ such that $\mathrm{Circ}(a)$ has an $\ell_n \times \ell_n$ minor M with rows I_n with

$$|\det(M)| \geq 2^{-\delta n \log n}.$$

By our preliminary remark, we know the conclusion of the theorem is true for $\ell_n = \Omega(n)$, without need to establish anything further. So actually, for complete coverage of all cases, we would want to establish the condition for functions ℓ_n that are *not* $\Omega(n)$, but we are already going to be content with the weaker theorem that would result from satisfying condition I for $\ell_n = o(n)$.

Let us remark the sets I_n and the subspace(s) U mentioned in the condition are *adversarial* in nature, they are determined by a hypothetical orbit circuit for circular convolution of size $o(n \log n)$, that we are trying to show does not exist. Hence the universal quantification over these quantities in the statement of condition I.

Given that Theorem 2.1.4 allows us to write

$$\mathrm{Circ}(a) = F_n \mathrm{diag}(DFT_n a) F_n^*,$$

it is no supprise that condition I is related to finding minors of DFT_n on a given set of rows I_n that have “reasonably” large determinant. We state the following condition:

(Condition II) For every family $\{I_n \subseteq \{0,1,\ldots,n-1\}\}_{n>0}$ with $\ell_n = |I_n| = o(n)$, and every $\delta > 0$, there exists a $k_0 > 0$ so that for all but finitely many n, for any $\lfloor \frac{\ell_n}{k_0} \rfloor$ columns $J_n \subset \{0,1,\ldots,n-1\}$, there exists an $\ell_n \times \ell_n$ minor M of DFT_n with rows I_n and columns disjoint from J_n with

$$|\det(M)| \geq 2^{-\delta n \log n}.$$

Theorem 5.2.1 *If Condition I holds, then so does Condition II.*

Proof. Suppose Condition I holds. Let $\{I_n \subseteq \{0,1,\ldots,n-1\}\}_{n>0}$ be given, and define $\ell_n = |I_n|$. Assume that $\ell_n = o(n)$ and let $\delta > 0$ be given. We want to argue it is now possible to select a $k_0 > 0$ so that for any family

$$\{J_n \subseteq \{0,1,\ldots,n-1\}\}_{n>0},$$

with $|J_n| = \lfloor \frac{\ell_n}{k_0} \rfloor$, for all but finitely many n, there exists $\ell_n \times \ell_n$ minor M of DFT_n with rows I_n and columns disjoint from J_n with

$$|\det(M)| \geq 2^{-\delta n \log n}.$$

For each n, define U_n to be the subspace of vectors v for which $(F_n v)_j = 0$ for all $j \in J_n$. This subspace has dimension $n - |J_n| = n - \lfloor \frac{\ell_n}{k_0} \rfloor$. By condition I, for any $\delta' > 0$,there exists $k_0' > 0$ so that, provided n is large enough, we have unit $a \in U_n$ such that $\mathrm{Circ}(a)$ has a square minor M with rows I_n such that

$$|\det(M)| \geq 2^{-\delta' n \log n}. \tag{5.3}$$

Let $Q(a) = \mathrm{Circ}(a)\mathrm{Circ}(a)^*$. Using Theorem 2.1.4 write

$$Q(a) = \mathrm{DFT}_n \mathrm{diag}(|\lambda_0|^2, |\lambda_1|^2, \ldots, |\lambda_{n-1}|^2)\mathrm{DFT}_n^*, \tag{5.4}$$

where $\lambda = F_n(a)$. Note that $||\lambda||_2 = 1$.

Using Theorem 4.1.3 (Binet-Cauchy Theorem) and (5.3) we get

$$\det(Q(a)_{I_n,I_n}) = \sum_{|S|=\ell_n} |\det(Circ(a)_{I_n,S}|^2 \geq 2^{-2\delta' n \log n}, \tag{5.5}$$

where in the sum S ranges over all subsets of size ℓ_n of $\{0,1,\ldots,n-1\}$. Also using Theorem 4.1.3 and now using (5.4) we can write

$$\det(Q(a)_{I_n,I_n}) = \sum_{|S|=\ell_n} |\det(DFT^n_{I_n,S})|^2 \prod_{s\in S} |\lambda_s|^2. \tag{5.6}$$

Since $||\lambda||_2 = 1$, for any S of size ℓ_n, $\sum_{s\in S} |\lambda_s|^2 \leq 1$. Using the arithmetic-geometric mean inequality, we then get

$$\prod_{s\in S} |\lambda_s|^2 \leq (\frac{1}{\ell_n})^{\ell_n}.$$

By our choice of U, the only terms in (5.6) that are possibly non-zero are those for those sets S that avoid J_n, namely $\prod_{s\in S}|\lambda_s|^2$ is zero for all others. Combining (5.5) and (5.6) we get that there exists some set S disjoint from J_n that has

$$|\det(DFT^n_{I_n,S})|^2 \prod_{s\in S}|\lambda_s|^2 \geq 2^{-2\delta' n\log n - n},$$

and hence

$$|\det(DFT^n_{I_n,S})|^2 \geq 2^{-2\delta' n\log n - n}\ell_n^{\ell_n}.$$

The above holds for any $\delta' > 0$, so with δ' chosen small enough we get that

$$2^{-2\delta' n\log n - n}\ell_n^{\ell_n} \geq 2^{-\delta n\log n},$$

This way we see Condition II is satisfied, provided Condition I holds. □

In other words Condition II is a necessary condition for establishing Condition I. In Chapter 6 we will see that Condition II would also be a sufficient condition for obtaining Condition I. However, now that we have extracted the more fundamental notion of finding "good" minors on the Fourier matrix, let us have a look at some issues that are involved in establishing Condition II.

As it turns out, Condition II is too strong to satisfy for *arbitrary* families of rows $\{I_n\}_{n>0}$ and columns $\{J_n\}_{n>0}$. To give an example of what can happen, suppose n is a square. Then it can be seen that any $\sqrt{n}\times\sqrt{n}$ minor of DFT_n that has rows which are multiples of $\sqrt{n}$ and avoids columns that are multiples of $\sqrt{n}$ is singular. For example, letting $\omega = e^{2\pi i/9}$, DFT_9 is given by:

$$\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\
1 & \omega & \omega^2 & \omega^3 & \omega^4 & \omega^5 & \omega^6 & \omega^7 & \omega^8\\
1 & \omega^2 & \omega^4 & \omega^6 & \omega^8 & \omega^1 & \omega^3 & \omega^5 & \omega^7\\
1 & \omega^3 & \omega^6 & 1 & \omega^3 & \omega^6 & 1 & \omega^3 & \omega^6\\
1 & \omega^4 & \omega^8 & \omega^3 & \omega^7 & \omega^2 & \omega^6 & \omega^1 & \omega^5\\
1 & \omega^5 & \omega^1 & \omega^6 & \omega^2 & \omega^7 & \omega^3 & \omega^8 & \omega^4\\
1 & \omega^6 & \omega^3 & 1 & \omega^6 & \omega^3 & 1 & \omega^6 & \omega^3\\
1 & \omega^7 & \omega^5 & \omega^3 & \omega^1 & \omega^8 & \omega^6 & \omega^4 & \omega^2\\
1 & \omega^8 & \omega^7 & \omega^6 & \omega^5 & \omega^4 & \omega^3 & \omega^2 & \omega^1
\end{pmatrix}$$

Selecting rows 0, 3 and 6:

$$\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\
1 & \omega^3 & \omega^6 & 1 & \omega^3 & \omega^6 & 1 & \omega^3 & \omega^6\\
1 & \omega^6 & \omega^3 & 1 & \omega^6 & \omega^3 & 1 & \omega^6 & \omega^3
\end{pmatrix}$$

and then removing columns 0, 3 and 6:

$$\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1\\
\omega^3 & \omega^6 & \omega^3 & \omega^6 & \omega^3 & \omega^6\\
\omega^6 & \omega^3 & \omega^6 & \omega^3 & \omega^6 & \omega^3
\end{pmatrix}$$

leaves a matrix with only two different kinds of columns, so any 3×3 minor of it will be singular. More generally, whenever $n = \ell \cdot k$, any $\ell \times \ell$ minor with rows $0, k, 2k, \ldots, (\ell - 1)k$ and columns avoiding $0, \ell, 2\ell, \ldots, (k-1)\ell$ can be seen to be singular. Consequently, one can observe that for $\ell_n = \omega(\sqrt{n})$ Condition II does not hold. In the next chapter we will therefore try to establish weaker versions of condition II, and derive (weakened) diagonal orbit lower bounds therefrom. The final lower bound theorem we will arrive at, while deriving some mathematical results that are interesting in their own right, will be the following result:

Main Result 5.2.1 *Let $\{D_n\}_{n>0}$ be a unit helper family, and suppose $\{\Gamma_n\}_{n>0}$ is a family of bounded-coefficient bilinear circuit such that for all n,*

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

Define $l_n = |D_n \cap (0,1)|$. We have that

1. *If $l_n = O(n^{\frac{1}{2}})$, then there exists $\gamma > 0$ so that $s(\Gamma_n) \geq \gamma n \log n$, for infinitely many n.*

2. *If $l_n = O(n^{\frac{3}{4}})$ and $\{D_n\}_{n>0}$ is asympotically contiguous, then then there exists $\gamma > 0$ so that $s(\Gamma_n) \geq \gamma n \log n$, for infinitely many n.*

3. *If $l_n = \Omega(n)$, then $s(\Gamma_n) = \Omega(n \log n)$.*

In the above, a family $\{D_n\}_{n>0}$ where each D_n is an n-tuple of distinct positive real numbers $(d_1^n, \ldots, d_n^n)$ such that $\prod_{i=1}^n d_i^n = 1$ is called a *unit helper family*. If for all but finitely many n, the entries in D_n of value less than one are contiguous (in the circular sense), we say that $\{D_n\}_{n>0}$ is *asymptotically contiguous*. In other words, the theorem proves a lower bound for orbit circuits of the form $\Gamma(Dx, y)$, where D is diagonal and with unit determinant, but with some further restrictions on how many helper constants are less than one, and how they are located relative to each other.

Curiously, in the above theorem it is not the unbounded constants that form a problem, but rather the seemingly innocent ones that are less than one, which the circuit could have supplied itself without problem. Note that the above theorem implies that we can handle, without further assumptions, any $n - o(\sqrt{n})$ many unbounded constants. At its most extreme this allows for $n - 1$ unbounded constants in the circuit, balanced against a single small helper constant that makes the product of all helper constants equal to one. This improves the εn many allowed unbounded constants for fixed $\varepsilon < 1$ from [BL02]. Although it must be said that we have strict requirements on where the constants are located in the circuit, and we have the requirement that their product is one. [BL02] has neither of these additional restrictions. Of course lifting the latter requirement puts one in the arena of the general unbounded constants case, which, even for linear circuits, has been a standing open problem in theoretical computer science for over 35 years.

5.3 Symmetry properties of cyclic convolution

We refer to Hungerford [Hun80] for the group theoretical notions used in the following. A curiousity is that in Chapter 4 we managed to lift the results of [BL02], amonst others, to orbit circuits of the form $\Gamma(U_0x, U_1x)$, where U_0 and U_1 are unitary. This includes the case where the free maps are permutation matrices. However, Theorem 5.2.1, or better said its proof, is incompatible with any such generalization. Of course, the two conditions of a unitary *and* diagonal matrix together, leave only the identity matrix, but more can be said. Namely, there is a certain lack of symmetry in the cyclic convolution map. In the following let S_n be the group of permutations on n-vectors. We think of each $\pi \in S_n$ to be a bijection $\pi : \mathbf{Z}_n \to \mathbf{Z}_n$, where $\mathbf{Z}_n = \{0, 1, \ldots, n-1\}$ is the additive group of integers modulo n.

Definition 5.3.1. Call a permutation $\pi \in S_n$ *retrievable* if there exist permutations π_1 and π_2 in S_n, such that

$$\pi_2[\pi(x)\mathrm{Circ}(\pi_1(y))] = x\mathrm{Circ}(y).$$

for n-vectors of variables $x = (x_0, x_1, \ldots, x_{n-1})$ and $y = (y_0, y_1, \ldots, y_{n-1})^T$.

In other words, a permutation is retrievable if application of it to the n-vector of variables x can be undone by applying a permutation to the n-vector y, and applying one to the result vector obtained by convolution of the permuted x and y vectors. Elementary reasoning yields the following:

Theorem 5.3.1 *For any n, the retrievable permutations form a group, and are precisely those permutation $\pi : \mathbf{Z}_n \to \mathbf{Z}_n$ for which there exists $b, g \in \mathbf{Z}_n$ with g relatively prime to n such that for each $i \in \mathbf{Z}_n$,*

$$\pi(i) = b + gi.$$

Proof. See Appendix B. □

The retrievable permutations form a subgroup R_n of S_n of size at most $n^2 - n$, hence there are in general vastly more unretrievable permutations than retrievable ones. So the circular convolution map enjoys nice symmetry properties, but, perhaps unexpectedly, is not “all symmetric”.

We conclude that for any $n > 3$, it is not in general possible to undo a permutation on the x variables by permuting the y variables, and then permuting the final result vector. If one could do this, then one could easily convert any circuit computing $\pi(x)\mathrm{Circ}(y)$ into one computing $x\mathrm{Circ}(y)$. Namely, simply permuting the y variables at the inputs and the outputs of the circuit with the π_1 and π_2 that work for π, and one is done. By taking inverses, this would mean that for any π, any circuit for $x\mathrm{Circ}(y)$ can be converted into a circuit computing $\pi(x)\mathrm{Circ}(y)$. Perfoming this conversion on an orbit circuit

$$\Gamma(x_1d_1, x_2, \ldots, x_nd_n, y) = x\mathrm{Circ}(y),$$

we would get a circuit Γ' such that

$$\Gamma'(x_{\pi(1)}d_1, x_{\pi(2)}, \ldots, x_{\pi(n)}d_n, y) = x\mathrm{Circ}(y).$$

which means we have a circuit Γ'' with

$$\Gamma''(x_1 d_{\pi^{-1}(1)}, x_2 d_{\pi^{-1}(2)}, \ldots, x_n d_{\pi^{-1}(n)}, y) = x\mathrm{Circ}(y).$$

In other words, we would have a means of permuting the helper constants on the variables. This would then allow us to at least establish item 2 of Theorem 5.2.1 without the contiguity requirement.

From the above we conclude that one cannot in general convert a circuit for $\pi(x)\mathrm{Circ}(y)$ into one computing $x\mathrm{Circ}(y)$ by permuting the y-inputs and outputs. However, something weaker would suffice for our purposes. Namely, if for every permutation π there exists a *reduction* that converts a circuit for $\pi(x)\mathrm{Circ}(y)$ into one for $x\mathrm{Circ}(y)$, using only $o(n\log n)$ additional circuit hardware, then one would obtain the same conclusion of establishing item 2 of Theorem 5.2.1 without the contiguity requirement. It is not clear whether this can be done.

5.4 Contiguity and Chordal Product

Given that we cannot establish Condition II in general, one natural scenario to consider is whether we can establish Condition II in case the set I_n is contiguous. Here we mean contiguous in the modular sense: $n-1$ and 0 are adjacent. In other words, I_n is contiguous if and only if it is of the form $\{b + r \bmod n : i \le r \le j\}$, for certain integers b, i and j. Establishing this weaker condition, would yield us a diagonal orbit lower bound for more restriced orbit circuits for which the helper variables that are less than 1 appear as a contiguous block, i.e. are all adjacent (again in the circular sense).

It is not hard to see that w.l.o.g. we can assume then that I_n consists of rows $0, 1, \ldots, \ell_n - 1$ of DFT_n. All $\ell_n \times \ell_n$ minors M with these rows are Vandermonde matrices of form $M = V(\omega_1, \omega_2, \ldots, \omega_{\ell_n})$ where the ω's are nth roots of unity. Using the determinant formula for a Vandermonde matrix, we have that $|\det(M)| = \mathcal{CP}(\omega_1, \omega_2, \ldots, \omega_{\ell_n})$, where we define for any finite set $P = \{p_1, p_2, \ldots, p_k\}$ of points on the unit circle in the complex plane their *chordal product*

$$\mathcal{CP}(P) = \prod_{1 \le i < j \le k} |p_i - p_j|.$$

Let $\Omega = \{\omega_0, \omega_1, \ldots, \omega_{n-1}\}$ be the nth roots of unity. Condition II now becomes:

(Condition II$'$) For $\ell_n = o(n)$, and every $\delta > 0$, there exists a $k_0 > 0$ so that for all but finitely many n, for any $\lfloor \frac{\ell_n}{k_0} \rfloor$ many roots of unity $J_n \subset \Omega$, there exists ℓ_n roots of unity $x_1, x_2, \ldots, x_{\ell_n} \in \Omega \setminus J_n$ such that

$$\mathcal{CP}(x_1, x_2, \ldots, x_{\ell_n}) \ge 2^{-\delta n \log n}.$$

Without the presence of the set of "off-limits" points J_n the $\mathcal{CP}(x_1, x_2, \ldots, x_{\ell_n})$ is maximized at $\ell_n^{\ell_n/2}$ by selecting the ℓ_n points with equal separation between adjacent points on the circle. Namely, Wlog. we select the ℓ_nth roots of unity. Hence $\mathcal{CP}(x_1, x_2, \ldots, x_{\ell_n}) = |\det(DFT_{\ell_n})| = \ell_n^{\ell_n/2}$. By the Hadamard Inequality (Theorem 2.1.3), this is the maximum magnitude of the determinant of any $\ell_n \times \ell_n$ matrix with unit entries.

The above means for example that for $\ell_n = O(\sqrt{n})$ it will be simple to satisfy Condition II$'$. Say $\ell_n \leq d\sqrt{n}$, for some constant $d > 0$, for all large enough n. For simplicity lets assume that ℓ_n divides n. Selecting k_0 so that $\lfloor \frac{\ell_n}{k_0} \rfloor < \frac{1}{d^2}\ell_n$ ensures that of the $\frac{n}{\ell_n}$ sets of ℓ_n equally spaced points, since $\frac{n}{\ell_n} \geq \frac{1}{d}\sqrt{n} \geq \frac{1}{d^2}\ell_n$, there must exist at least one that contains no off-limit point from J_n.

As we will see, we can actually establish condition II for $\ell_n = O(\sqrt{n})$, so there is no need for a contiguity requirement in this case at all. For general $\ell_n = o(n)$ there is no such simple argument as we described above. We are faced with the following problem:

Problem. *For some large n, consider the set $\Omega = \{\omega_0, \omega_1, \ldots, \omega_{n-1}\}$ of all nth roots of unity on the unit cirlce in the complex plane. Let $R \subseteq \Omega$ be a given set of roots that are "off-limits". For any ℓ, what is the optimal strategy to select ℓ roots of unity $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_\ell} \in \Omega \setminus R$ that maximizes $\mathcal{CP}(\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_\ell})$?*

Related to this question, what sets R in the above provide the worst-case scenario? That is:

Problem. *For any k, ℓ, for what kind of sets $R \subseteq \Omega$ of size k is*

$$\max_{\substack{S \subseteq \Omega / R \\ |S| = \ell}} \mathcal{CP}(S)$$

minimized, and what is this min-max value ?

We have some indication that sets R that are contiguous provide this worst-case scenario, but the question is related to some standing open problems [DS89] that turn out to be suprisingly hard to solve, as we will discuss in the next chapter.

For establishing item 2 of Theorem 5.2.1 we consider a randomized strategy: pick the ℓ_n points uniformly at random from the collection of points that are allowed. This strategy works fairly well. It enables us to get out desired lower bound for $\ell_n = O(n^{3/4})$.

For $\ell_n = n^{\varepsilon}$, with ε a constant arbitrarily close to 1, we give evidence that there is *no strategy at all* that enables us to satisfy Condition II$'$. We will give evidence that Condition II$'$ cannot be satisfied, even for $\varepsilon = 4/5 + \delta$, where $\delta > 0$ is constant. We do so by employing what is known about the asymptotic spectrum of the discrete prolate spheroidal wave functions [Sle78].

Chapter 6

Uncertainty Principles & Matrix Games

The *Heisenberg uncertainty principle* in quantum mechanics is widely known, even to the extent of having had a cultural impact. The principle is a theorem derivable from the axioms of quantum mechanics, and expresses the inherent impossibility of simultaneously knowing, to arbitrary precision, certain complementary observables in nature. For example, one cannot simultaneously, through measurement, determine both the position and velocity of some given elementary particle to arbitrary precision.

Physical interpretation aside, the uncertainty principle can be expressed quite generally as a mathematical statement about operators in a Hilbert space $\mathcal{H}$. Following Selig and Somaraju-Hanlen [Sel01, SH05], say $\mathcal{H}$ has inner product denoted by $\langle \cdot, \cdot \rangle$ and norm $|| \cdot || = \langle \cdot, \cdot \rangle^{1/2}$. For a linear operator $A : \mathcal{H} \to \mathcal{H}$ we denote its domain by $\mathcal{D}(A)$. Define the *normalized expected value of* A *with respect to* $f \in \mathcal{D}(A)$ by

$$\tau_A(f) = \frac{\langle Af, f \rangle}{\langle f, f \rangle}$$

and the *standard deviation of* A *with respect to* f by

$$\sigma_A(f) = ||(A - \tau_A(f))f||.$$

The uncertainty principle relates the standard deviations of two operators A and B to their commutator $[A,B]$, which is defined as $[A,B] = AB - BA$. An operator A is said to be symmetric if $\langle Ax, y \rangle = \langle x, Ay \rangle$ for every $x, y \in \mathcal{D}(A)$.

Theorem 6.0.1 (Uncertainty Principle, see [SH05]) *Let A and B be symmetric operators in some Hilbert space $\mathcal{H}$. Then*

$$\sigma_A(f)\sigma_B(f) \geq \frac{|\langle [A,B]f, f \rangle|}{2},$$

for all $f \in \mathcal{D}(AB) \cap \mathcal{D}(BA)$.

For the Hilbert space $L^2(\mathbf{R})$ of all square integrable functions $f : \mathbf{R} \to \mathbf{C}$, with inner product defined by

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)\overline{g(x)}dx,$$

the above implies the following classic uncertainty statement about the measures of concentration of a function $f \in L^2(\mathbf{R})$ and its Fourier transform $\hat{f} : \mathbf{R} \to \mathbf{C}$, defined by

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx.$$

Namely we have that

Theorem 6.0.2 (see [SH05]) *Let $f \in L^2(\mathbf{R})$ with $||f|| = 1$. Let*

$$x_a = \int_{-\infty}^{\infty} x|f(x)|^2 dx,$$

$$\omega_a = \int_{-\infty}^{\infty} \omega|f(\omega)|^2 dx,$$

$$\Delta x = \int_{-\infty}^{\infty} (x - x_a)^2 |f(x)|^2 dx, \text{ and}$$

$$\Delta\omega = \int_{-\infty}^{\infty} (\omega - \omega_a)^2 |f(\omega)|^2 d\omega.$$

Then

$$\Delta x \Delta\omega \geq \pi/2.$$

The above shows that for a function $f : \mathbf{R} \to \mathbf{C}$ one cannot simultaneously localize f and its Fourier transform $\hat{f}$ to arbitrary extent: the smaller the standard deviation Δx of f, the larger the standard deviation $\Delta\omega$ of $\hat{f}$ must be. Going back to physical interpretation briefly, in this scenario f could be the *wave function* of some particle (in one dimension, at some fixed time), in which case one obtains the probability of detecting the particle by square integration of the wave function. The position of the particle is a random variable, and the quantity x_a is the expected location of the particle. Δx is the standard deviation of this position random variable. As it turns out, $\hat{f}$ is the wave function in *momentum space*, that is ω_a and $\Delta\omega$ are the average and standard deviation of the momentum of the particle. The position and momentum arise from probablity distributions as one can witness them in reality by carrying out some large number of identically prepared experiments. The above gives limits on how much one can narrow down simultaneously the deviations for position and momentum.

There are severals settings in which one can observe the uncertainty phenomena. The above scenario is "continuous-to-continuous", i.e. the Fourier transform and its inverse move functions between continuous domains. Donoho and Stark [DS89] investigated several "discrete-to-discrete" analogues of the above uncertainty relation. That is, for n-vector x and its *discrete* Fourier transform $\hat{x} = DFT_n x$, they considered as measure of localization the support $supp(x)$, which is the total number of non-zeroes of x. They also defined a more quantatively subtle notion of ε-concentration of a vector x on a set of indices $T \subseteq \{0, 1, \ldots, n-1\}$,which is defined as the ℓ_2-norm of x restricted to T. For these two measures they proved inequalities in the spirit of Theorem 6.0.2, showing the limits on the simultaneous concentration achievable for any Fourier pair $(x, \hat{x})$.

Uncertainty relations of the kind obtained by [DS89] are closely related to properties of minors of the Fourier matrix DFT_n. For proving lower bounds in the orbit model for the circular convolution bilinear map $x^T \operatorname{Circ}(y)$, precise quantitative statements about these minors regarding the magnitude of their determinant is exactly what we need, as we saw in Chapter 5 with Conditions I and II expressed there.

In the following sections we will express particular *sufficient* conditions for yielding orbit model lower bounds in terms of certain *games* played on the DFT_n matrix. These games will take the place of Conditions I and II of Chapter 5. Using this linguistic tool will conveniently suppress some of the lengthy quantifier alternations in our statements that we would otherwise have.

To outline the idea, the game is played between a player and an adversary. The adversary chooses a set of rows R and a set of columns C. Then the player tries to select a minor of DFT_n with rows R avoiding columns C in order to maximize the determinant. We will establish connections between the existence of certain good strategies for this game to uncertainty type relations in the "discrete-to-discrete" setting. We will then use this to prove Theorem 5.2.1.

For the first item of this theorem we will involve an uncertainty relation proven in [DS89]. Unfortunately, this argument breaks down, for reasons indicated in Chapter 5, for ℓ_n beyond $O(n^{1/2})$.

In order to establish some further results, we make additional assumptions on the constant d_i present at the inputs. In case they are asymptotically contiguous, we can press on the statement of our theorem for larger ℓ_n up to $O(n^{3/4})$. Namely, in this case it will turn out that for lower bounds it is sufficient for the player to win the more relaxed version of the Fourier matrix game in which it is assumed that the set of rows R the adversary chooses in contiguous. Consequently, determinants of Vandermonde matrices will play a role.

The problem becomes the following: with some number ℓ_n of the nth roots of unity being disallowed by the adversary, how do we select ℓ_n other roots of unity $x_1, x_2, \ldots, x_{\ell_n}$ in order to maximize the determinant of the Vandermonde matrix $V(x_1, x_2, \ldots, x_{\ell_n})$ supported by those points? We will show a randomized strategy for the player that is sufficient for proving orbit model lower bounds in which we can tolerate up to $O(n^{3/4})$ roots being disallowed by the player. In order to achieve this result we prove a lower bound on the expected value of the determinant of the Vandermonde matrix $V(x_1, x_2, \ldots, x_{\ell_n})$ with nodes on the unit circle. This result is interesting in its own right, and may have further applications.

One application we give, is an uncertainty-type relation for a discrete analogue of the bandlimited functions. In the continous setting, a function $f : \mathbf{R} \to \mathbf{C}$ is called bandlimited if there exists $\Omega \in \mathbf{R}$ such that $\hat{f}(\omega) = 0$ for all $|\omega| > \Omega$. For bandlimited functions more intricate details are known about simultaneous concentration of f and $\hat{f}$ than the standard uncertainty principle. See for example [Sle78] for a study in the "continuous-to-discrete" domain.

Interestingly enough, [Sle78] will also give us some indications on the limits we can expect with our taken approach. Desirable would be to find player strategies that can tolerate *any* $\ell_n = o(n)$ number of roots being disallowed by the adversary. The worst-case scenario appears to arise when the adversary chooses these roots to be contiguous. When he/she does, we have some indication that there is no good strategy for the player (in a sense which we will make more precise later) once $\ell_n = \Omega(n^{4/5} \log^{1/5} n)$.

6.1 Minor Games on Matrices

Definition 6.1.1. We define the circulant game CIRC-Game(n,l,k,B) to be the following single-round game against an adversary agent:

Adversary: selects a linear subspace $U \subset \mathbf{C}^n$ of co-dimension k and l distinct rows $r_1, r_2, \ldots, r_l \in \{0, 1 \ldots, (n-1)\}$.
Player: selects $a \in U$ with $\|\|a\|\|_2 = 1$, and selects an $l \times l$ minor M of $Circ(a)$ with rows $r_1, r_2, ..., \ldots, r_l$.
Result: The player wins if and only if $\|\det(M)\| > B$.

Related to the above game is the following game on the DFT_n matrix:

Definition 6.1.2. We define the Fourier matrix game DFT-Game(n,l,k,B) to be the following single-round game against an adversary agent:

Adversary: selects l distinct rows $r_1, r_2, \ldots, r_l$ and k distinct columns $c_1, c_2, \ldots, c_k$ in $\{0, 1, \ldots, (n-1)\}$.
Player: selects an $l \times l$ minor M of the $n \times n$ Fourier matrix DFT_n with rows $r_1, r_2, \ldots, r_l$ and columns disjoint from $c_1, c_2, \ldots, c_k$.
Result: The player wins if and only if $\|\det(M)\| > B$.

We define DFT-Game$^*(n,l,k,B)$ and CIRC-Game$^*(n,l,k,B)$ to the same games as above, but with the relaxation that the adversary can choose only sets of rows R that are contiguous in the cyclic sense: $R = \{b + i \bmod n : 0 \leq i \leq l-1\}$ for some *base point b*.

For the contiguous circulant game it is immediately obvious that we can assume without loss of generality that the adversary chooses any particular contiguous set R of our preference, since for any two chosen sets R_1 and R_2, the matrices $Circ(a)_{R_1}$ and $Circ(a)_{R_2}$ just differ by a cyclic shift. We can make the same assumption without loss of generality for the contiguous Fourier game. Namely, for any l columns $C = \{c_1, c_2, \ldots, c_l\}$ and two contiguous sets R_1 and R_2 with base points b_1 and b_2 respectively, we have that

$$DFT_{R_1,C} = DFT_{R_2,C} \cdot \mathrm{diag}(\omega^{rc_1}, \omega^{rc_2}, \ldots, \omega^{rc_l}),$$

where $r = b_1 - b_2$, and $\omega = e^{2\pi i/n}$. Hence $|\det(DFT_{R_1,C})| = |\det(DFT_{R_2,C})|$.

We begin by proving a generalization of the phenomena we sketched in Chapter 5 with the DFT_9 example: for this matrix, any 3×3 minor with rows 0, 3, and 6 and columns avoiding 0, 3, and 6 is singular. In general we have the following:

Theorem 6.1.1 *If $n = l \cdot k$, then the adversary has a winning strategy for DFT-Game$(n,l,k,0)$.*

Proof. A winning strategy for the adversary is to take rows

$$r_i = ki,$$

for $i = 0, 1, \ldots, (l-1)$, and columns

$$c_i = li,$$

for $i = 0, 1, \ldots, (k-1)$.

Let A be the $l \times n$ minor of DFT_n with rows $r_0, r_1, \ldots r_{l-1}$. The rth column A_r of A equals $(1, \alpha^r, \alpha^{2r}, \ldots, \alpha^{(l-1)r})^T$, where $\alpha = e^{\frac{2\pi i}{n}k} = e^{\frac{2\pi i}{l}}$. Hence for any r,

$$A_r = A_{r+l \bmod n}.$$

With columns $0, l, 2l, \ldots, (k-1)l$ disallowed, there are therefore only $l-1$ distinct columns in the remaining set, so any $l \times l$ minor of A that avoids the disallowed columns will be singular.

□

Corollary 6.1.2 *If n is a square, then the adversary has a winning strategy for DFT-Game$(n, \sqrt{n}, \sqrt{n}, 0)$.*

So if $n = l \cdot k$, there is not much honor to achieve in general for the player as it comes to playing DFT-Game$(n, l, k, \cdot)$. This will also have a negative impact on the general lower bound result we are trying to prove, as we will see. It is the reason why in Theorem 5.2.1 for item 1 we stated a limitation of $\ell_n = O(n^{1/2})$. In case $k \cdot l < n$ however, this pathetic case does not apply, and the player does have a non-trivial strategy. For $k \cdot l$ below n, perturbation theory kicks in, and by applying the Binet-Cauchy Theorem one can guarantee the existence of a minor with a "reasonable" lower bound on the magnitude of its determinant. We have the following result:

Theorem 6.1.3 *The player has a winning strategy for DFT-Game(n, l, k, B), provided $k \cdot l < n$, and*

$$B < (n-kl)^{l/2} \binom{n-k}{l}^{-1/2}.$$

Proof. Suppose the adversary chooses l rows R and k columns C. Let $N = \{0, 1, \ldots, n-1\}$. Let $A = DFT_{R,N/C}$ and $B = DFT_{R,C}$. Then

$$AA^* = nI - BB^*$$

Both AA^* and BB^* are Hermitian, so by Theorem 2.1.2 (Weyl's Perturbation Theorem), provided $||BB^*||_2 \leq n$, for each i, the ith eigenvalue $\lambda_i(AA^*) \geq n - ||BB^*||_2$. We can write

$$BB^* = \sum_{i \in C} c_i c_i^*,$$

where c_i is the ith column of $DFT_{R,N}$. Since $||c_i c_i^*||_2 \leq ||c_i||_2^2 = l$, then by subadditivity of the ℓ_2-norm, $||BB^*||_2 \leq kl$. Hence

$$\det(AA^*) \geq (n-kl)^l.$$

By Theorem 4.1.3 (Binet-Cauchy Theorem),

$$\det(AA^*) = \sum_{|S|=l} |\det(A_{R,S})|^2.$$

Hence we conclude there exists S of size l such that

$$|\det DFT_{R,S}| \geq (n-kl)^{l/2}\binom{n-k}{l}^{-1/2}.$$

□

For our lower bound results for circular convolution, we require good strategies not for the Fourier matrix game, but rather for the circulant matrix game. Fortunately, these two games are closely related. In one direction we have the following theorem:

Theorem 6.1.4 *If the adversary has a winning strategy for DFT-Game(n,l,r,B), then it has a winning strategy for CIRC-Game$(n,l,r,\binom{n-r}{l}n^{-\frac{l}{2}}B)$. The same statement holds with Game replaced by Game*.*

Proof. Let R and C be the sets of l rows and r columns of the adversary's winning strategy in the fourier matrix game. Then for the circulant game the adversary chooses the set R for the rows, and takes U to be the subspace of vectors v for which $(F_n v)_i = 0$ for all $i \in C$. This subspace has dimension $n-r$.

Say the player picks unit $a \in U$, and say T is the set of columns of the minor he chooses. Using Theorem 2.1.4 write

$$\mathrm{Circ}(a) = \frac{1}{\sqrt{n}}\mathrm{DFT}_n \mathrm{diag}(\lambda)\mathrm{DFT}_n^*,$$

where $\lambda = F_n(a)$. Then $||\lambda||_2 = 1$.

Using Theorem 4.1.3 (Binet-Cauchy Theorem) we can write:

$$\det(Circ(a)_{R,T}) = \sum_{|S|=l}(\prod_{s\in S}\lambda_s)\det(DFT_{R,S})\det(\frac{1}{\sqrt{n}}DFT^*_{S,T}).$$

Since $||\lambda||_2 = 1$, for any S of size l, $\sum_{s\in S}|\lambda_s| \leq \sqrt{l}$. Using the arithmetic-geometric mean inequality, we then get

$$\prod_{s\in S}|\lambda_s| \leq (\frac{1}{\sqrt{l}})^l,$$

and note that there are at most $n-r$ nonzero λ_s because of the choice of U. By Theorem 2.1.3 (Hadamard inequality) we then have

$$|\det(\frac{1}{\sqrt{n}}DFT^*_{S,T})| \leq (\frac{\sqrt{l}}{\sqrt{n}})^l.$$

Since $\det(DFT_{R,S}) < B$ for any S disjoint from C, and $\prod_{i \in S} \lambda_i = 0$ for all other sets S, we get that

$$|\det(Circ(a)_{R,T})| < \sum_{|S|=l, S \cap C = \emptyset} (\frac{1}{\sqrt{l}})^l B (\frac{\sqrt{l}}{\sqrt{n}})^l = \binom{n-r}{l} n^{-l/2} B.$$

This proves the statement for the regular versions of the game. The statement for both versions of the relaxed game can be verified analogously. □

From this we see that the same pathetic case $n = k \cdot l$ arises for the circulant game. In this situation again there is not much glory to achieve for the player. Namely, we have the following corollary:

Corollary 6.1.5 *If $n = l \cdot k$, then the adversary has a winning strategy for CIRC-Game$(n,l,k,0)$. In particular, the adversary can win Circ-Game$(n, \sqrt{n}, \sqrt{n}, 0)$ in case n is a square.*

We can also prove a relation between the circulant and Fourier game in the reverse direction. The following lemma yields a way for the player to tranfer his strategy for the Fourier game to the circulant game. The strategy for the player in this case is to use some randomization: given the subspace U that the adversary selects, the player selects a standard Gaussian vector in U. Given that the player has a "good" strategy for the Fourier matrix game, this will combine to be a good strategy for the circulant game as well.

Theorem 6.1.6 *For any n,r,l with $l + r \leq n$, if the player has a winning strategy for DFT-Game(n,l,r,B), then the player has a winning strategy for CIRC-Game(n,l,r,B'), where*

$$B' = \frac{B\delta^{l/2}}{\sqrt{\binom{n}{r} 4^l (n-r)^l}},$$

and δ is a constant approximately 0.02. *More precisely $\delta = 2^{-(\gamma + \sqrt{2\phi})}$ with $\gamma = \frac{1}{\sqrt{\pi}} \int_0^\infty t^{-\frac{1}{2}} e^{-t} \log t dt$, and $\phi = \frac{1}{2} \int_0^\infty e^{-\frac{t}{2}} \log^2 t dt$. The same statement holds with Game replaced by Game*.*

Proof. Suppose the adversary chooses subspace U of dimension $n - r$ and a set of l rows R in the circulant game.

Consider standard Gaussian randomly selected $a \in_R U$, then $\lambda = F_n a$ is also standard Gaussian.

Write $\lambda = A\alpha$, where A is an $n \times (n-r)$ matrix that has orthonormal columns that span $F_n U$, and α is standard Gaussian in $\mathbf{C}^{n-r}$. Apply Theorem 4.1.3 (Binet-Cauchy):

$$\sum_{|R|=n-r} |\det(A_R)|^2 = \det(A^* A) = 1.$$

Hence there exists a set R of $n-r$ rows with $|\det(A_R)|^2 \geq \binom{n}{r}^{-1}$. Since the player can win DFT-Game(n,l,r,B), let T be a subset of R such that

$$|\det(DFT_{R,T})| > B.$$

Note that $\det(A_T A_T^*) \geq \binom{n}{r}^{-1}$. Namely $\det(A_T A_T^*) = \det(MM^*)$, where M is obtained by adding $n-r-l$ rows to A_T which are orthonormal and orthogonal to the span of the rows of A_T. Since each row r of A has $||r||_2 \leq 1$ we must have that $|\det(M)| \geq |\det(A_R)|$. That is, the $|\det(M)|$ is the maximum determinant one can get by appending $n-r-l$ rows of norm at most 1 to the l rows A_T.

The matrix $A_T A_T^*$ is the covariance matrix of centered Gaussian vector $(\lambda_i)_{i\in T}$. By Lemma 4.1.2, with probability greater than $\frac{1}{2}$ we have that

$$\prod_{i\in T} |\lambda_i|^2 \geq \delta^l \det(A_T A_T^*) \geq \delta^l \binom{n}{r}^{-1},$$

where δ is a constant approximately 0.02. More precisely, Lemma 4.1.2 gives $\delta = 2^{-(\gamma+\sqrt{2\phi})}$ with $\gamma = \frac{1}{\sqrt{\pi}} \int_0^\infty t^{-\frac{1}{2}} e^{-t} \log t dt$, and $\phi = \frac{1}{2}\int_0^\infty e^{-\frac{t}{2}} \log^2 t dt$.

Now let us bound the norm of the vector λ. We have that

$$E[||\lambda||_2^2] = E[||\alpha||_2^2] = (n-r)E[|\alpha_1|^2] = 2(n-r).$$

The last equality follows from Lemma 4.1.1. By the Markov inequality,

$$\Pr[||\lambda||_2^2 \leq 4(n-r)] \geq \frac{1}{2}.$$

From the above we conclude there must exist a vector $a \in U$ such that if we let $\lambda = F_n a$, then $||\lambda||_2^2 \leq 4(n-r)$ and simultaneously

$$\prod_{i\in T} |\lambda_i|^2 \geq \delta^l \binom{n}{r}^{-1}.$$

Say the player chooses $a' = \frac{a}{||a||_2}$, which is unit. Theorem 2.1.4 (Convolution Theorem) implies:

$$Circ(a') = DFT_n diag(\lambda') F_n,$$

where $\lambda' = F_n a'$. Let $D = Circ(a')Circ(a')^*$. Then

$$D = DFT_n diag(|\lambda_0'|^2, |\lambda_1'|^2, \ldots, |\lambda_{n-1}'|^2) DFT_n^*.$$

Using Theorem 4.1.3 (Binet-Cauchy), we can write

$$\begin{aligned} \det(D_{R,R}) &= \sum_{|S|=l} (\prod_{i\in S} |\lambda_i'|^2) |\det(DFT_{R,S})|^2 \\ &\geq (\prod_{i\in T} |\lambda_i'|^2) |\det(DFT_{R,T})|^2 \\ &> \frac{B^2 \delta^l}{\binom{n}{r} 4^l (n-r)^l}. \end{aligned}$$

Applying Binet-Cauchy once more, we have that

$$\det(D_{R,R}) = \sum_{|S|=l} |\det(Circ(a')_{R,S})|^2.$$

Hence there exists S such that

$$|\det(Circ(a')_{R,S})| > \frac{B\delta^{l/2}}{\sqrt{\binom{n}{r}4^l(n-r)^l}}.$$

This is the minor that the player chooses.

The above argument goes through in case of playing the contiguous games. In this case the R chosen by the adversary is contiguous, so to get the result it suffices for the player to invoke its winning strategy for DFT-Game$^*(n,l,r,B)$ instead.

□

As we can see in the above Lemma 6.1.6, there is some loss in the threshold B by which the player can win the game. In our application in Section 6.5 however, it will turn out that this loss is ignorable as a lower order term in our estimates. This gives us the convenience of focusing on the more fundamental notion of playing the Fourier matrix game.

6.2 Random Vandermonde Matrices

We are going to employ the probabilistic method to show the existence of good strategies for playing the contiguous Fourier matrix game. For the contiguous Fourier matrix game the essential question becomes the following:

Problem. *For some large n, consider the set $\Omega = \{\omega_0, \omega_1, \ldots, \omega_{n-1}\}$ of all nth roots of unity on the unit circle in the complex plane. Let $R \subseteq \Omega$ be a given set of roots that are "off-limits". For any ℓ, what is the optimal strategy to select ℓ roots of unity $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_\ell} \in \Omega \setminus R$ that maximizes the Vandermonde determinant:*

$$\prod_{1 \le s < t \le \ell} |\omega_{i_s} - \omega_{i_t}| \quad ?$$

Related to this question, what sets R in the above provide the worst-case scenario? That is:

Problem. *For any k, ℓ, for what kind of sets $R \subseteq \Omega$ of size k is*

$$\max_{\substack{S \subseteq \Omega \setminus R \\ |S| = \ell}} \prod_{p \neq q \in S} |p - q|$$

minimized, and what is this min-max value ?

6.2.1 Related Work

The above two questions are related to the following. Suppose $T \subseteq D := \{0, 1, \ldots, n-1\}$. What sets R minimize $||DFT^n_{T,R}||_2$? If we let $M = DFT^n_{T,R}$ and $N = DFT^n_{T,D\setminus R}$, then

$$MM^* + NN^* = nI.$$

So using the Theorem 2.1.2, if $||DFT^n_{T,R}||_2 < K$ we get that each eigenvalue $\lambda_i(NN^*) \geq n-K$, and consequently that $|\det(NN^*)| \geq (n-K)^l$, where $\ell = |T|$. Then applying Theorem 4.1.3 one gets that there exists an $\ell \times \ell$ minor of N with determinant of magnitude at least $\frac{(n-K)^\ell}{\binom{n-|R|}{\ell}}$. Donoho and Stark considered the *opposite* question of which sets T and R maximize $||DFT^n_{T,R}||_2$. They define the "index-limiting" operator $P_R = \mathrm{diag}(1_R)$, where 1_R is the $0,1$-valued n-vector that is 1 precisely for all indices in R, and the "frequence-limiting" operator $P_T = F^*_n \mathrm{diag}(1_T) F_n$. Note that $||P_T P_R||_2 = ||DFT_{T,R}||_2$. They conjecture:

Conjecture 2 ([DS89]). For interval T and set R with $|R| \cdot |T| = n$, $||P_T P_R||_2$ is maximized when R is also an interval.

Potentially, maximizing $||DFT^n_{T,R}||_2$ yields the converse effect of forcing $|\det(NN^*)|$ to be small, although one cannot directly conclude this from Theorem 2.1.2. Forcing $|\det(NN^*)|$ to be small also depresses the value

$$\max_{\substack{S \subseteq \Omega \setminus R \\ |S| = \ell}} \mathcal{CP}(S), \tag{6.1}$$

where $\mathcal{CP}(S)$ is the *chordal product of* S, which we defined in Section 5.4 by

$$\mathcal{CP}(S) = \prod_{p \neq q \in S} |p-q|.$$

So as answer to the second problem above, it appears plausible that the bad sets R that minimize (6.1) are when R is chosen to be an interval, i.e. if R is a set of indices that is contiguous in the modular sense. Computer runs seem to corroborate this idea, and in the analysis that follows such R indeed seem to form the major difficulty.

Also related to our work, is the question of the conditioning of a Vandermonde matrix. For real numbers $r_1, r_2, \ldots, r_\ell$, the Vandermonde matrix $V(r_1, r_2, \ldots, r_\ell)$ infamously can be highly ill-conditioned [Gau75]. For Vandermonde matrices with nodes in the complex plane, where the nodes are arranged to be nicely spread out the situation can be better. Ferreira [Fer99] gives some bounds for Vandermonde matrices with nodes on the unit circle in the complex plane that show the matrix can be quite well conditioned provided the nodes are spread around the cirle evenly.

We should also mention the powerful work done by Camdes, Romberg and Tao [CRT04]. They prove that for any T of size $O(\frac{n}{\log n})$, if one selects the set S by independently choosing for each column k to be in S with probability τ, where τ is some fixed constant, then with high probability for $M = DFT^n_{T,S}$, the determinant $\det(MM^*)$ is "not to small". Unfortunately, their *moment method* approach is not robust against the adversarial set R of points to avoid. At a critical juncture in their proof they rely on the *cancellation property* of the roots of unity, which states that for any r not divisible by n we have

$$\sum_{i=0}^{n-1} \omega^{ri} = 0,$$

where ω is any primitive nth root of unity. The presence of the set R implies that not all roots appear with equal probability, indeed some may appear with probability 0. Consequenctly, after taking expectations and doing a brute force application of the inclusion-exclusion principle, not all roots of unity are guaranteed to appear in the final expression to be cancelled. Hence the attempt to adapt their proof to our situation breaks down. Seen more holistically, since their proof makes no assumption about T, except on its size, the presence of the set R must make their proof break down, because of the phenomena we sketched in the introduction. Recall our observation that if n is a square, then there exist sets T and R of size $\sqrt{n}$ such that all minors with rows T and avoiding columns R are singular. The question is whether we can do better by assuming that T is contiguous. We will now turn to this question.

6.2.2 Randomized Selection Strategy

We first prove an estimate on a particular sum that is involved in the analysis.

Define the "ln-of-chord-length" function $f(t) = \ln|1-e^{it}|$, for $t \in \mathbf{R} \setminus \{k2\pi : k \in \mathbf{Z}\}$. Straightforward geometry gives us:

$$f(t) = \frac{1}{2}\ln(2-2\cos t),$$

which can be rewritten using the relation $\sin^2\frac{\alpha}{2} = \frac{1-\cos\alpha}{2}$ as

$$f(t) = \ln 2\sin\frac{t}{2} = \ln 2 + \ln|\sin\frac{t}{2}|.$$

We will also consider a discretized version of this function, which per abuse of notation will also be denoted by f. It will be clear from the context, whether f is referring to the discrete or continuous function.

Lemma 6.2.1 *Let $\varepsilon(t) = \ln|t| - f(t)$. Then for any t with $|t| < 1$,*

$$0 < \varepsilon(t) < \frac{t^2}{12}.$$

Proof. First of all for any t, $f(t) = \ln|1-e^{it}| < \ln|t|$. We thus see that $\varepsilon(t)$ is non-negative. For $t \in (0, 2\pi)$, we have for the error function $\varepsilon(t) = \ln|t| - f(t) = \ln\frac{|t|}{2\sin\frac{t}{2}}$. For $t > 0$, $\sin t \geq t - \frac{t^3}{6}$. So on this interval, $\varepsilon(t) \leq \ln\frac{t}{t-\frac{t^3}{24}} = -\ln(1-\frac{t^2}{24})$. For $\frac{-1}{24} < x < \frac{1}{24}$, $\ln(1+x) \geq x - \frac{x^2}{2}$. So for $0 < t < 1$,

$$0 < \varepsilon(t) < \frac{t^2}{24} + \frac{t^4}{1152} < \frac{t^2}{12}.$$

The lemma follows by symmetry of $f(t)$ and $\ln|t|$. □

Lemma 6.2.2 *Let $n \geq 7$, and let $\omega = e^{2\pi i/n}$. Define the discrete function $f(d) = \ln|1-\omega^d|$, for $d = 1,2,\ldots,n-1$. Then*

$$\sum_{d=1}^{n-1} f(d) \geq 2 - \ln 2 - \frac{2\pi^2}{3n^2}.$$

Proof. Using the fact (see e.g. [RW04], p. 182, equation 55) that

$$\int_0^{\pi\backslash 2} \ln(\sin x)dx = -\frac{\pi}{2}\ln 2,$$

we get that

$$\begin{aligned}
\int_0^{2\pi} f(t)dt &= 2\pi\ln 2 + \int_0^{2\pi} \ln\sin\frac{t}{2}dt \\
&= 2\pi\ln 2 + 2\int_0^{\pi} \ln\sin\frac{t}{2}dt \\
&= 2\pi\ln 2 + 4\int_0^{\pi\backslash 2} \ln\sin x dx \\
&= 0.
\end{aligned}$$

For $j = 0,1,,\ldots,n-1$, define interval $I_j = [j\frac{2\pi}{n},(j+1)\frac{2\pi}{n}]$. By the above,

$$\begin{aligned}
\frac{2\pi}{n}\sum_{d=1}^{n-1} f(d) &= \frac{2\pi}{n}\sum_{d=1}^{n-1} f(d) - \int_0^{2\pi} f(t)dt \\
&= \frac{2\pi}{n}f(1) - 2\int_0^{2\pi/n} f(t)dt + \frac{2\pi}{n}\sum_{d=2}^{n-1} f(d) - \int_{2\pi/n}^{(n-1)\frac{2\pi}{n}} f(t)dt. \qquad (6.2)
\end{aligned}$$

We will approximate $f(t)$ by $\ln t$ for t close to 0, and estimate the error incurred by this to bound the first two terms of (6.2). Using Lemma 6.2.1 , provided $n \geq 7$,

$$f(1) \geq \ln\frac{2\pi}{n} - \varepsilon(\frac{2\pi}{n}) \geq \ln\frac{2\pi}{n} - \frac{\pi^2}{3n^2}$$

and

$$\begin{aligned}
\int_0^{2\pi/n} f(t)dt &\leq \int_0^{2\pi/n} \ln t dt \\
&\leq [t\ln t - t]_0^{2\pi/n} \\
&= \frac{2\pi}{n}\ln\frac{2\pi}{n} - \frac{2\pi}{n}.
\end{aligned}$$

Hence (6.2) is at least

$$\begin{aligned}\frac{2\pi}{n}\ln\frac{2\pi}{n}-\frac{2\pi^3}{3n^3}-\frac{4\pi}{n}\ln\frac{2\pi}{n}+\frac{4\pi}{n}+\frac{2\pi}{n}\sum_{d=2}^{n-1}f(d)-\int_{2\pi/n}^{(n-1)\frac{2\pi}{n}}f(t)dt \quad &\geq \\ \frac{2\pi}{n}\ln\frac{n}{2\pi}+\frac{4\pi}{n}-\frac{2\pi^3}{3n^3}+\frac{2\pi}{n}\sum_{d=2}^{n-1}f(d)-\int_{2\pi/n}^{(n-1)\frac{2\pi}{n}}f(t)dt. \qquad &(6.3)\end{aligned}$$

We will now bound the last two term in the above expression. Let us first consider the case when n is even.

$$\begin{aligned}\frac{2\pi}{n}\sum_{d=2}^{n-1}f(d)-\int_{2\pi/n}^{(n-1)\frac{2\pi}{n}}f(t)dt &= \frac{2\pi}{n}\sum_{d=2}^{n-1}f(d)-\sum_{d=1}^{n-2}\int_{I_d}f(t)dt \\ &= \frac{2\pi}{n}\sum_{d=2}^{n/2}[f(d)+f(n+1-d)]-2\sum_{d=1}^{\frac{n}{2}-1}\int_{I_d}f(t)dt \\ &= \frac{2\pi}{n}\sum_{d=2}^{n/2}[f(d)+f(d-1)]-2\sum_{d=1}^{\frac{n}{2}-1}\int_{I_d}f(t)dt \\ &= \frac{2\pi}{n}\sum_{d=1}^{n/2-1}[f(d)+f(d+1)]-2\sum_{d=1}^{\frac{n}{2}-1}\int_{I_d}f(t)dt \\ &= \sum_{d=1}^{n/2-1}(\frac{2\pi}{n}[f(d)+f(d+1)]-2\int_{I_d}f(t)dt). \qquad (6.4)\end{aligned}$$

Since for $1\leq d\leq n/2-1$, $f(t)$ is strict monotone increasing, we know that

$$\frac{2\pi}{n}[f(d)+f(d+1)]-2\int_{I_d}f(t)dt\geq-\frac{2\pi}{n}[f(d+1)-f(d)].$$

Hence (6.4) is at least

$$\begin{aligned}\frac{2\pi}{n}\sum_{d=1}^{n/2-1}[f(d)-f(d+1)] &= \frac{2\pi}{n}[f(1)-f(\frac{n}{2})] \\ &\geq \frac{2\pi}{n}[\ln\frac{2\pi}{n}-\frac{\pi^2}{3n^2}-\ln 2].\end{aligned}$$

Hence (6.2) is at least

$$\frac{4\pi-2\pi\ln 2}{n}-\frac{4\pi^3}{3n^3}.$$

Hence we conclude that in case n is even, that

$$\sum_{d=1}^{n-1}f(d)\geq 2-\ln 2-\frac{2\pi^2}{3n^2}.$$

Let us now consider the case when n is odd. Then

$$\begin{aligned}
& \frac{2\pi}{n}\sum_{d=2}^{n-1} f(d) - \int_{2\pi/n}^{(n-1)\frac{2\pi}{n}} f(t)dt \\
= \quad & \frac{2\pi}{n}\sum_{d=2}^{n-1} f(d) - \sum_{d=1}^{n-2}\int_{I_d} f(t)dt \\
= \quad & \frac{2\pi}{n} f(\frac{n+1}{2}) + \frac{2\pi}{n}\sum_{d=2}^{(n-1)/2}[f(d)+f(n+1-d)] - 2\sum_{d=1}^{\frac{n-1}{2}-1}\int_{I_d} f(t)dt - \int_{I_{(n-1)/2}} f(t)dt \\
= \quad & \frac{2\pi}{n} f(\frac{n+1}{2}) - \int_{I_{(n-1)/2}} f(t)dt + \frac{2\pi}{n}\sum_{d=2}^{(n-1)/2}[f(d)+f(d-1)] - 2\sum_{d=1}^{\frac{n-1}{2}-1}\int_{I_d} f(t)dt \\
= \quad & \frac{2\pi}{n} f(\frac{n+1}{2}) - \int_{I_{(n-1)/2}} f(t)dt + \frac{2\pi}{n}\sum_{d=1}^{\frac{n-1}{2}-1}[f(d)+f(d+1)] - 2\sum_{d=1}^{\frac{n-1}{2}-1}\int_{I_d} f(t)dt \\
= \quad & \frac{2\pi}{n} f(\frac{n+1}{2}) - \int_{I_{(n-1)/2}} f(t)dt + \sum_{d=1}^{\frac{n-1}{2}-1}(\frac{2\pi}{n}[f(d)+f(d+1)] - 2\int_{I_d} f(t)dt). \qquad (6.5)
\end{aligned}$$

Since for $1 \le d \le (n-1)/2-1$, $f(t)$ is strict monotone increasing, we know that

$$\frac{2\pi}{n}[f(d)+f(d+1)] - 2\int_{I_d} f(t)dt \ge -\frac{2\pi}{n}[f(d+1)-f(d)].$$

Hence (6.5) is at least

$$\begin{aligned}
\frac{2\pi}{n} f(\frac{n+1}{2}) - \int_{I_{(n-1)/2}} f(t)dt + \frac{2\pi}{n}\sum_{d=1}^{\frac{n-1}{2}-1}[f(d)-f(d+1)] &= \\
\frac{2\pi}{n} f(\frac{n+1}{2}) - \int_{I_{(n-1)/2}} f(t)dt + \frac{2\pi}{n}[f(1)-f(\frac{n-1}{2})] &= \\
-\int_{I_{(n-1)/2}} f(t)dt + \frac{2\pi}{n} f(1) &\ge \\
\frac{2\pi}{n}[\ln\frac{2\pi}{n} - \frac{\pi^2}{3n^2} - \ln 2], &
\end{aligned}$$

and so we obtain the same bound as the n is even case. □

Lemma 6.2.3 *Let $n \ge 7$, and let $\omega = e^{2\pi i/n}$. Define the discrete function $f(d) = \ln|1-\omega^d|$, for $d = 1, 2, \ldots, n-1$, then*

$$\sum_{d=1}^{n-1} f(d) \le 2\ln\frac{n}{2\pi} + 2 + \ln 2 + \frac{\pi^2}{5832n^2}.$$

Proof. For $j = 0, 1, , \dots, n-1$, define interval $I_j = [j\frac{2\pi}{n}, (j+1)\frac{2\pi}{n}]$. As in the proof of Lemma 6.2.2 we can write

$$\begin{aligned}
\frac{2\pi}{n}\sum_{d=1}^{n-1} f(d) &= \frac{2\pi}{n}\sum_{d=1}^{n-1} f(d) - \int_0^{2\pi} f(t)dt \\
&= \frac{2\pi}{n} f(1) - 2\int_0^{2\pi/n} f(t)dt + \frac{2\pi}{n}\sum_{d=2}^{n-1} f(d) - \int_{2\pi/n}^{(n-1)\frac{2\pi}{n}} f(t)dt. \qquad (6.6)
\end{aligned}$$

Provided $n \geq 7$, we have by Lemma 6.2.1 that

$$\begin{aligned}
\int_0^{2\pi/n} f(t)dt &\geq \int_0^{2\pi/n} \ln t - \varepsilon(t)dt \\
&\geq [t\ln t - t]_0^{2\pi/n} - [\frac{t^3}{36}]_0^{2\pi/n} \\
&= \frac{2\pi}{n}\ln\frac{2\pi}{n} - \frac{2\pi}{n} - \frac{2\pi^3}{5832n^3}.
\end{aligned}$$

Hence (6.6) is at most

$$\frac{2\pi}{n} f(1) + \frac{4\pi}{n}\ln\frac{n}{2\pi} + \frac{4\pi}{n} + \frac{2\pi^3}{5832n^3} + \frac{2\pi}{n}\sum_{d=2}^{n-1} f(d) - \int_{2\pi/n}^{(n-1)\frac{2\pi}{n}} f(t)dt. \qquad (6.7)$$

We will now bound the last two term in the above expression. Let us first consider the case when n is even. As in the proof of Lemma 6.2.2 we can write

$$\frac{2\pi}{n}\sum_{d=2}^{n-1} f(d) - \int_{2\pi/n}^{(n-1)\frac{2\pi}{n}} f(t)dt = \sum_{d=1}^{n/2-1} \left(\frac{2\pi}{n}[f(d) + f(d+1)] - 2\int_{I_d} f(t)dt\right). \qquad (6.8)$$

Since for $1 \leq d \leq n/2 - 1$, $f(t)$ is strict monotone increasing, we know that

$$\frac{2\pi}{n}[f(d) + f(d+1)] - 2\int_{I_d} f(t)dt \leq \frac{2\pi}{n}[f(d+1) - f(d)].$$

Hence (6.8) is most

$$\begin{aligned}
\frac{2\pi}{n}\sum_{d=1}^{n/2-1} [f(d) - f(d+1)] &= \frac{2\pi}{n}[f(\frac{n}{2}) - f(1)] \\
&= \frac{2\pi}{n}[\ln 2 - f(1)].
\end{aligned}$$

Hence (6.6) is at most

$$\frac{4\pi}{n}\ln\frac{n}{2\pi} + \frac{4\pi}{n}(1 + \frac{1}{2}\ln 2) + \frac{2\pi^3}{5832n^3}.$$

Hence we conclude that in case n is even, that

$$\sum_{d=1}^{n-1} f(d) \leq 2\ln\frac{n}{2\pi} + 2 + \ln 2 + \frac{\pi^2}{5832n^2}.$$

Let us now consider the case when n is odd. As in Lemma 6.2.2 we can write

$$\begin{aligned}
& \frac{2\pi}{n}\sum_{d=2}^{n-1} f(d) - \int_{2\pi/n}^{(n-1)\frac{2\pi}{n}} f(t)dt \\
= \quad & \frac{2\pi}{n} f(\frac{n+1}{2}) - \int_{I_{(n-1)/2}} f(t)dt + \sum_{d=1}^{\frac{n-1}{2}-1} (\frac{2\pi}{n}[f(d)+f(d+1)] - 2\int_{I_d} f(t)dt). \qquad (6.9)
\end{aligned}$$

Since for $1 \leq d \leq (n-1)/2 - 1$, $f(t)$ is strict monotone increasing, we know that

$$\frac{2\pi}{n}[f(d)+f(d+1)] - 2\int_{I_d} f(t)dt \leq \frac{2\pi}{n}[f(d+1) - f(d)].$$

Hence (6.9) is at most

$$\begin{aligned}
\frac{2\pi}{n} f(\frac{n+1}{2}) - \int_{I_{(n-1)/2}} f(t)dt + \frac{2\pi}{n}\sum_{d=1}^{\frac{n-1}{2}-1} [f(d) - f(d+1)] &= \\
\frac{2\pi}{n} f(\frac{n+1}{2}) - \int_{I_{(n-1)/2}} f(t)dt + \frac{2\pi}{n}[f(1) - f(\frac{n-1}{2})] &= \\
-\int_{I_{(n-1)/2}} f(t)dt + \frac{2\pi}{n} f(1) &\leq \\
\frac{2\pi}{n}[\ln 2 - f(1)]. &
\end{aligned}$$

Hence we obtain the same bound as the n is even case. □

We now turn to the main result in this section. Given that in the contiguous version of the Fourier matrix game it does not really matter which block of rows the adversary chooses, we will focus on playing the game on the first l many rows. In this case any selected minor will be a Vandermonde matrix. In order to show existence of a good minor that avoids the set of columns chosen by the adversary, we will consider selecting a random such minor, and evaluate the expected value of its determinant.

More precisely, for complex numbers $z_0, z_1, \ldots, z_{l-1}$, denote by $V = V(z_0, z_1, \ldots, z_{l-1})$ the $l \times l$ Vandermonde matrix defined by $V_{ij} = z_i^j$ for $0 \leq i, j \leq l-1$, we have the following theorem:

Theorem 6.2.4 *For any n and l, r with $0 \leq r < \frac{n}{\pi}$ and $l + r \leq n$, Let $N = \{\omega^k | k = 0, 1, \ldots, n-1\}$, where $\omega = e^{2\pi i/n}$. Let R be an arbitrary subset of N of size r. Consider the process of picking*

$\{z_0,\dots,z_{l-1}\} \subset N\backslash R$ uniformly at random among all subsets of $N\backslash R$ of size l. Then for the Vandermonde matrix $V = V(z_0,z_1,\dots,z_{l-1})$ we have

$$E[\ln|\det V|] \geq \frac{(n-2r)\binom{l}{2}}{(n-r)(n-r-1)}(2-\ln 2-\frac{2\pi^2}{3n^2}) - \frac{\binom{l}{2}}{(n-r)(n-r-1)}(r^2\ln\frac{n}{r\pi}+r^2+\frac{r^4\pi^2}{36n^2}).$$

Proof.

$$\begin{aligned}
E[\ln|\det V|] &= E[\ln\prod_{i<j}|z_i-z_j|] \\
&= E[\sum_{i<j}\ln|z_i-z_j|] \\
&= \sum_{i<j}E[\ln|z_i-z_j|] && \text{(by linearity of E)} \\
&= \binom{l}{2}E[\ln|z_0-z_1|] && \text{(by symmetry).}
\end{aligned}$$

Let $\eta = E[\ln|z_0-z_1|]$. We can write the following expression for η:

$$\eta = \sum_{p\in N\backslash R}\sum_{q\in N\backslash R, q\neq p} \Pr[z_0 = p \text{ and } z_1 = q]\ln|p-q|.$$

Since $\{z_0,z_1\}$ is uniform among 2-subsets of $N\backslash R$, for any $p\neq q$,

$$\Pr[(z_0 = p \text{ and } z_1 = q) \text{ or } (z_0 = q \text{ and } z_1 = p)] = \binom{|N\backslash R|}{2}^{-1}.$$

Since the events $[(z_0 = p \text{ and } z_1 = q]$ and $[(z_0 = q \text{ and } z_1 = p]$ are disjoint and have equal probability, we can conclude that $\Pr[(z_0 = p \text{ and } z_1 = q] = \frac{1}{2}\binom{|N\backslash R|}{2}^{-1} = \frac{1}{(n-r)(n-r-1)}$. Define $f(k) = \ln|1-w^k|$, for $k = 1,2,\dots,n-1$, and let χ correspond to the characteristic function of $N\backslash R$. That is $\chi(i) = 1$ if $i\in N\backslash R$, and 0 otherwise. We have

$$\begin{aligned}
\eta &= \frac{1}{(n-r)(n-r-1)}\sum_{p\in N\backslash R}\sum_{q\in N\backslash R, q\neq p}\ln|p-q| \\
&= \frac{1}{(n-r)(n-r-1)}\sum_{i=0}^{n-1}\sum_{j=0, j\neq i}^{n-1}\chi(i)\chi(j)\ln|\omega^i-\omega^j| \\
&= \frac{1}{(n-r)(n-r-1)}\sum_{i=0}^{n-1}\sum_{d=1}^{n-1}\chi(i)\chi(i+d \bmod n)\ln|\omega^i-\omega^{i+d}| \\
&= \frac{1}{(n-r)(n-r-1)}\sum_{i=0}^{n-1}\sum_{d=1}^{n-1}\chi(i)\chi(i+d \bmod n)\ln|1-\omega^d|
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(n-r)(n-r-1)} \sum_{i=0}^{n-1} \sum_{d=1}^{n-1} \chi(i)\chi(i+d \bmod n) f(d) \\
&= \frac{1}{(n-r)(n-r-1)} \sum_{d=1}^{n-1} f(d) \sum_{i=0}^{n-1} \chi(i)\chi(i+d \bmod n) \\
&= \frac{1}{(n-r)(n-r-1)} \sum_{d=1}^{n-1} f(d)c(d),
\end{aligned}
$$

where we define

$$c(d) = \sum_{i=0}^{n-1} \chi(i)\chi(i+d \bmod n).$$

Now for any d,

$$c(d) = \sum_{i=0}^{n-1} \chi(i)\chi(i+d \bmod n) \geq n-2r,$$

since for fixed d, the number of indices i for which at least one of $\chi(i)$ and $\chi(i+d \bmod n)$ is zero is at most $2r$. Also we have that

$$
\begin{aligned}
\sum_{d=0}^{n-1} c(d) &= \sum_{d=0}^{n-1} \sum_{i=0}^{n-1} \chi(i)\chi(i+d \bmod n) \\
&= \sum_{i=0}^{n-1} \chi(i) \sum_{d=0}^{n-1} \chi(i+d \bmod n) \\
&= \sum_{i=0}^{n-1} \chi(i)(n-r) \\
&= (n-r)^2.
\end{aligned}
$$

So

$$\sum_{d=1}^{n-1} c(d) = (n-r)^2 - (n-r) = n^2 - 2rn + r^2 - n + r.$$

Since $c(d)$ is always at least $n-2r$, define an "excess" function $e(d)$ by

$$e(d) = c(d) - (n-2r).$$

The total excess equals

$$\sum_{d=1}^{n-1} e(d) = \sum_{d=1}^{n-1} c(d) - (n-1)(n-2r) = r^2 - r.$$

We get that

$$\sum_{d=1}^{n-1} f(d)c(d) = \sum_{d=1}^{n-1} f(d)[e(d) + (n-2r)]$$

$$\begin{aligned}
&= (n-2r)\sum_{d=1}^{n-1} f(d) + \sum_{d=1}^{n-1} f(d)e(d) \\
&\geq (n-2r)(2-\ln 2 - \frac{2\pi^2}{3n^2}) + \sum_{d=1}^{n-1} f(d)e(d),
\end{aligned}$$

where the last line follows from Lemma 6.2.2.

Note that for any d, $c(d) \leq n-r$, and thus $0 \leq e(d) \leq r$. We know that $\sum_{d=1}^{n-1} f(d)e(d)$ is smallest if the total excess $r^2 - r$ is placed at much as possible at places where $f(d)$ is the smallest. By the concavity of f, one can conclude that in case r is odd,

$$\begin{aligned}
\sum_{d=1}^{n-1} f(d)e(d) &\geq \sum_{d=1}^{(r-1)/2} f(d)r + \sum_{d=n-\frac{r-1}{2}}^{n-1} f(d)r \\
&= 2r\sum_{d=1}^{(r-1)/2} f(d) \\
&\geq 2r\frac{n}{2\pi}\int_0^{\frac{r-1}{2}\frac{2\pi}{n}} f(t)dt \\
&= \frac{rn}{\pi}\int_0^{\frac{(r-1)\pi}{n}} \ln t - \varepsilon(t)dt \\
&\geq \frac{rn}{\pi}[t\ln t - t - \frac{t^3}{36}]_0^{\frac{(r-1)\pi}{n}} \qquad \{\text{by Lemma 6.2.1}\} \\
&= r(r-1)\ln\frac{(r-1)\pi}{n} - (r-1)r - \frac{1}{36}\frac{r(r-1)^3\pi^2}{n^2} \\
&\geq r^2\ln\frac{r\pi}{n} - r^2 - \frac{r^4\pi^2}{36n^2},
\end{aligned}$$

and in case that r is even,

$$\begin{aligned}
\sum_{d=1}^{n-1} f(d)e(d) &\geq \sum_{d=1}^{r/2} f(d)r + \sum_{d=n+1-\frac{r}{2}}^{n-1} f(d)r \\
&\geq r\frac{n}{2\pi}\int_0^{\frac{r\pi}{n}} f(t)dt + r\frac{n}{2\pi}\int_0^{\frac{(r-2)\pi}{n}} f(t)dt + \\
&= \frac{rn}{2\pi}\int_0^{\frac{r\pi}{n}} \ln t - \varepsilon(t)dt + \frac{rn}{2\pi}\int_0^{\frac{(r-2)\pi}{n}} \ln t - \varepsilon(t)dt \\
&\geq \frac{rn}{2\pi}[t\ln t - t - \frac{t^3}{36}]_0^{\frac{r\pi}{n}} + \frac{rn}{2\pi}[t\ln t - t - \frac{t^3}{36}]_0^{\frac{(r-2)\pi}{n}} \qquad \{\text{by Lemma 6.2.1}\} \\
&\geq r^2\ln\frac{r\pi}{n} - r^2 - \frac{r^4\pi^2}{36n^2}.
\end{aligned}$$

Hence we finally conclude that

$$E[\ln\det V] \geq \frac{(n-2r)\binom{l}{2}}{(n-r)(n-r-1)}(2-\ln 2-\frac{2\pi^2}{3n^2}) - \frac{\binom{l}{2}}{(n-r)(n-r-1)}(r^2\ln\frac{n}{r\pi}+r^2+\frac{r^4\pi^2}{36n^2}).$$

□

As a special case it can be verified that the statement of Theorem 6.2.4 is also valid with $r=0$. In this case there are no roots of unity to be avoided. Combined with Lemma 6.2.3 we get:

Corollary 6.2.5 *For any n and any $l \leq n$, Let N be the set of nth roots of unity. Let $\{z_0,\ldots,z_{l-1}\} \subset N$ be a uniformly at random selected subset of size l. Then for the Vandermonde matrix $V = V(z_0,z_1,\ldots,z_{l-1})$ we have that*

$$E[\ln|\det(V)|] = \Phi\frac{\binom{l}{2}}{n-1},$$

where

$$2-\ln 2-\frac{2\pi^2}{3n^2} \leq \Phi \leq 2\ln\frac{n}{2\pi}+2+\ln 2+\frac{\pi^2}{5832n^2}.$$

Proof. Following the initial steps of the proof of Theorem 6.2.4 for $r=0$ one obtains

$$E[\ln|\det(V)|] = \frac{\binom{l}{2}}{n-1}\sum_{d=1}^{n-1} f(d).$$

Applying Lemma's 6.2.2 and 6.2.3 gives the result. □

Let us note however that in this case one knows the expected value of the determinant exactly:

Proposition 6.2.6 *For a random Vandermonde matrix V selected as in Corollary 6.2.5, we have that $E[|\det(V)|^2] = n^l\binom{n}{l}^{-1}$.*

Proof. Let $\mathcal{M}$ be the set of all $l \times l$ minors of DFT_n with rows $0,1,\ldots,l-1$. By Binet-Cauchy (Theorem 4.1.3):

$$\sum_{V\in\mathcal{M}} |\det(V)|^2 = n^l.$$

Hence for a uniformly at random selected $V \in_R \mathcal{M}$ we have that $E[|\det(V)|^2] = n^l\binom{n}{l}^{-1}$. □

Theorem 6.2.4 gives us a strategy for winning the contiguous version of the Fourier matrix game, which in turn, using Theorem 6.1.6, yields a strategy for winning the contiguous circulant game. This strategy will be the basis for the circuit lower bounds we will prove in Section 6.5.

Corollary 6.2.7 *For any n and any l,r with $1 \leq r < \frac{n}{\pi}$ and $l+r \leq n$, the player has a winning strategy for DFT-Game$^*(n,l,r,e^C)$, provided*

$$C < \frac{(n-2r)\binom{l}{2}}{(n-r)(n-r-1)}(2-\ln 2 - \frac{2\pi^2}{3n^2}) - \frac{\binom{l}{2}}{(n-r)(n-r-1)}(r^2 \ln\frac{n}{r\pi} + r^2 + \frac{r^4\pi^2}{36n^2}).$$

Proof. Recalling our remark after Definition 6.1.2, we can assume with loss of generality that the adversary chooses rows $R = \{0,1,\ldots,l-1\}$. Any $l \times l$ minor of DFT_n with rows R is a Vandermonde matrix. Let C be the set of columns the adversary chooses. Theorem 6.2.4 gives a lower bound on $E[\ln|\det(M)|]$ for randomly selected $l \times l$ minor M of DFT_n with rows R avoiding columns C. There must exist at least one minor M' that has $\ln|\det(M')| \geq E[\ln|\det(M)|]$. So the player chooses such a minor, for which we then have the lower bound on the absolute value of its determinant as stated in the corollary. $\square$

Let us express the above hiding some of the constants for later convenience:

Corollary 6.2.8 *For any n and any l,r with $1 \leq r < \frac{n}{\pi}$ and $l+r \leq n$, the player has a winning strategy for DFT-Game$^*(n,l,r,B)$ for some B where*

$$B \geq 2^{\Theta(\frac{l^2}{n} - \frac{l^2 r^2}{n^2}\log\frac{n}{r})}.$$

6.3 Discrete Uncertainty Principles

In this section we will establish a relation between the matrix games and various known discrete uncertainty relations. Let us begin with an alternative proof, which is new to our knowledge, of the Donoho-Stark discrete uncertainty principle.

Definition 6.3.1. For an n-vector f, define the *support of f* to be the set $\mathrm{supp}(f) = \{i : f_i \neq 0\}$.

The size of the support of a vector f is a crude measure of the amount of localization of a vector. Analogous to the Heisenberg uncertainty principle, we can prove that for this measure a vector f and its Fourier transform $\hat{f}$ cannot both be arbitrarily narrowly localized. More precisely, we have the following theorem:

Theorem 6.3.1 ([DS89]) *For any n-vector $f \neq 0$,*

$$|\mathrm{supp}(f)| \cdot |\mathrm{supp}(\hat{f})| \geq n, \tag{6.10}$$

where $\hat{f} = F_n f$ is the discrete Fourier transform of f.

Proof. Consider an arbitrary Fourier transform pair $(f,\hat{f})$ with $\hat{f} = F_n f$ and $f \neq 0$. Since

$$\mathrm{Circ}(f) = \sqrt{n}F_n^*\mathrm{diag}(\hat{f})F_n,$$

we have that

$$\mathrm{supp}(\hat{f}) = \mathrm{rank}(\mathrm{Circ}(f)).$$

Let R be the maximum number of zeroes following a non-zero entry in f (in the cyclic sense). Then $R \geq \frac{n}{|\mathrm{supp}(f)|} - 1$.

Namely, if this were not the case, then imagine partitioning the entries of f as follows: Start at an arbitrary nonzero position. Set $i = 1$. If there are no other zero positions then B_i equals this position. Otherwise, let B_i be this position together with all the zero positions that follow it (in the cyclic sense). Repeat this process for the next i. We obtain this way $B_1, B_2, \ldots, B_{|\mathrm{supp}(f)|}$ that partition all n entries of f. By the above then, for each i, $|B_i| \leq R+1 < \frac{n}{|\mathrm{supp}(f)|}$. So

$$|\bigcup_i B_i| < |\mathrm{supp}(f)| \cdot \frac{n}{|\mathrm{supp}(f)|} = n.$$

This is a contradiction, because $B_1, B_2, \ldots, B_{|\mathrm{supp}(f)|}$ partition the n entries of f.

The above implies the first $R+1$ rows of $\mathrm{Circ}(f)$ are independent, because they contain a square submatrix that is upper triangular (modulo cylic shifts). Hence $\mathrm{rank}(\mathrm{Circ}(f)) \geq R+1 \geq \frac{n}{|\mathrm{supp}(f)|}$. □

Interestingly enough, divisibility properties of n play an important role in the analysis. For example, Tao showed that, in case n is prime, the inequality (6.10) can be significantly improved. The proof relies on the well-known fact that for prime p the discrete Fourier transform matrix DFT_p is regular.

Definition 6.3.2. An $n \times n$ marix A is called *regular* if any square submatrix of A is non-singular.

Theorem 6.3.2 *For prime p, DFT_p is a regular matrix.*

The first proof of this fact is attributed to Chebotarëv, who proved it in 1926 (see [SJ96]). Although typical proofs of this fact are field theoretic in nature, Tao gives a proof by elementary means. Once one has established this fact the following can be proved quite readily:

Theorem 6.3.3 ([Tao91]) *For prime p, for any nonzero p-vector f and its Fourier transform $\hat{f} = F_p f$ we have that*

$$|\mathrm{supp}(f)| + |\mathrm{supp}(\hat{f})| \geq p+1.$$

Proof. Let $k = p - |\mathrm{supp}(\hat{f})|$. There are k zeroes in $\hat{f}$. Let $I \subseteq \{0, 1, \ldots, p-1\}$ be the indices of these zeroes. Suppose $|\mathrm{supp}(f)| \leq k$. Let $J \subseteq \{0, 1, \ldots, p-1\}$ be a set of size k that contains all indices of non-zero entries of f. We have that

$$(DFT^p_{I,J}) f_J = (DFT_p f)_I = 0,$$

but $f_J \neq 0$ since $f \neq 0$. This is a contradiction since $DFT^p_{I,J}$ is non-singular. Hence $|\mathrm{supp}(f)| > k = p - |\mathrm{supp}(\hat{f})|$. □

Actually, in the above proof we only used the fact that DFT_p is a regular matrix, so more generally we have:

Theorem 6.3.4 *Let A be an $n \times n$ regular matrix and consider pairs $(f, \hat{f} := Af)$ where $f \neq 0$. Then*

$$|\text{supp}(f)| + |\text{supp}(\hat{f})| \geq n + 1.$$

Corollary 6.3.5 *For any prime p, and any $l + k \leq p$, the player has a winning strategy for DFT-Game$(p, l, k, 0)$.*

In what follows, we will establish relations between our matrix games and uncertainty type relations. We will show that we can turn a refinement of the Donoho-Stark uncertainty relation into a game strategy, and also provide a tranferral in the converse direction. The game strategy obtained this way later will be used to prove our main lower bound theorem for orbit circuits. We also show that the strategy obtained in Corollary 6.2.7 can gives us an uncertainty type relation. This uncertainty relation will be for a discrete analogue of band-limited functions.

Definition 6.3.3. An n-vector f is called *l-index-limited* if

$$\text{supp}(f) \subseteq \{b + i \bmod \text{n} : 0 \leq i \leq l - 1\},$$

for some number b.

In other words a vector f is l-index-limited if its support is contained in a contiguous set (in the cyclic sense) of size l.

Let us start by making some preliminary observations about index-limited vectors f in conjunction with the support-size notion of localization. In the next section, we will turn to a more precise localization measure than $|\text{supp}(f)|$. For index-limited vectors one can easily prove a strengthening of the uncertainty inequality (6.10). We follow the same top-level idea used to prove Theorem 6.3.3.

Theorem 6.3.6 *For any n-vector $f \neq 0$ that is l-index-limited,*

$$|\text{supp}(\hat{f})| > n - l,$$

where $\hat{f} = F_n f$ is the discrete Fourier transform of f.

Proof. Consider an arbitrary Fourier transform pair $(f, \hat{f})$ and let $T = \{b + i \bmod \text{n} : 0 \leq i \leq l - 1\}$ be a contiguous set of indices containing $\text{supp}(f)$. Suppose $|\text{supp}(\hat{f})| \leq n - l$. Then we can find a set $S = \{s_1, s_2, \ldots, s_l\}$ of size l so that $\hat{f}_i = 0$ for each $i \in S$. In other words $DFT_{S,T} f_T = 0$ with $f_T \neq 0$. So $DFT_{S,T}$ is singular. However, $DFT_{S,T} = \text{diag}(\omega^{s_1 b}, \omega^{s_2 b}, \ldots, \omega^{s_l b}) V(\omega^{s_1}, \omega^{s_2}, \ldots, \omega^{s_k})$, that is a (nonsingular) diagonal matrix multiplied with a (nonsingular) Vandermonde matrix, and is hence not singular. $\square$

For $l \ll n - 2\sqrt{n} + 1$, the above guarantees $|\text{supp}(f)| + |\text{supp}(\hat{f})| \gg 2\sqrt{n}$, whereas Theorem 6.3.1 can only guarantee $|\text{supp}(f)| + |\text{supp}(\hat{f})| \geq 2\sqrt{n}$.

6.3.1 Uncertainty relations imply game strategies

We now turn to a less crude measure of localization than the support of a vector. Following [DS89]:

Definition 6.3.4. An n-vector f is ε-*concentrated* on a set T of indices if

$$\sqrt{\sum_{i \notin T} |f_i|^2} \leq \varepsilon.$$

Theorem 6.3.1 can be refined as follows:

Theorem 6.3.7 ([DS89]) *For any n-vector f with $||f||_2 = 1$ that is ε_T-concentrated on a set T and $\hat{f} = F_n f$ being ε_Ω-concetrated on a set Ω, we have that*

$$|T| \cdot |\Omega| \geq n(1 - (\varepsilon_T + \varepsilon_\Omega))^2. \tag{6.11}$$

Note that Brody and Meister [BM99] claim that the inequality (6.11) in the statement of the theorem can be improved to

$$S(\frac{|T| \cdot |\Omega|}{n}) \geq (1 - (\varepsilon_T + \varepsilon_\Omega))^2,$$

where $S(x)$ is defined as $S(x) = \frac{2}{\pi} Si(x) - \frac{1}{\pi} sin(x)$, and where $Si(x)$ is the sine-integral function: $Si(x) = \int_0^x \frac{sint}{t} dt$.

Counter-examples can be given to this claim for any t and u with $tu = n$ by taking $T = \{0, u, 2u, \ldots, u(t-1)\}$ and $\Omega = \{0, t, 2t, \ldots, t(u-1)\}$. It is well know [DS89] that the indicator for T transforms to the indicator of Ω when taking the Fourier tranform. In other words, there exist a Fourier transform pair $(f, \hat{f})$ with f 0-concentrated on T and $\hat{f}$ 0-concentrated on Ω. However $S(1) < 1$, so the above would claim this is impossible. We have been unable to verify the original intent of Brody and Meister's claim, and the authors have not responded to our queries.

Let us now use Theorem 6.3.7 in order to obtain a "fairly" good strategy for playing the Fourier matrix game. This will be the basis for proving part of the bilinear circuit lower bounds in section 6.5. For certain types of circuits Theorem 6.3.7 will be not be strong enough. This is where the game strategy obtained in Corollary 6.2.7 comes in. From this game strategy we will also be able to derive strengthened uncertainty relations for index-limited vectors.

Theorem 6.3.8 *For any l, r with $lr \leq n$ and $l + r \leq n$, the player has a winning strategy for DFT-Game(n, l, r, B). for any*

$$B < (\sqrt{n} - \sqrt{lr})^l \binom{n-r}{l}^{-1/2}.$$

Proof. Suppose the adversary chooses a set of rows R of size l and set of columns T of size r. Let M be the minor of F_n with rows R and columns T. By Theorem 6.3.7, for any unit vector f

that is 0-concentrated on T, $\hat{f} = F_n f$ is ε_R concentrated on R, where

$$\varepsilon_R \geq 1 - \sqrt{\frac{lr}{n}}.$$

Hence

$$||M||_2 = \max_{||a||_2=1} ||Ma||_2 \leq \sqrt{\frac{lr}{n}}.$$

Let N be the $l \times (n-r)$ minor of F_n corresponding to rows R and columns not in T. Since $NN^* + MM^* = I$, λ is an eigenvalue of MM^* if-and-only if $(1-\lambda)$ is an eigenvalue of NN^*. The singular values of M are the square roots of the eigenvalues of MM^*. Hence we conclude the smallest singular value of N is at least

$$\sigma_l^2(N) \geq 1 - \sqrt{\frac{lr}{n}},$$

and hence that

$$\sigma_l^2(\sqrt{n}N) \geq \sqrt{n} - \sqrt{lr}.$$

Therefore

$$\det(\frac{1}{n}NN^*) \geq (\sqrt{n} - \sqrt{lr})^{2l}.$$

By Theorem 4.1.3,

$$\det(\frac{1}{n}NN^*) = \sum_{|S|=l, S\cap T=\emptyset} |\det(DFT_{R,S})|^2.$$

Hence we conclude there exists a minor M_1 with rows R and columns avoiding T that has determinant at least

$$|\det(M_1)| \geq (\sqrt{n} - \sqrt{lr})^l \binom{n-r}{l}^{-1/2}.$$

□

Actually, Theorem 6.1.3 yields a slightly stronger strategy than the above theorem. For the types of lower bounds we will prove in Section 6.5 the slight numerical differences will turn out to be immaterial.

6.3.2 Games strategies imply uncertainty relations

Winning strategies against the adversary for the Fourier matrix game are useful for yielding discrete uncertainty relations. Similarly, winning strategies against the adversary in the contiguous Fourier matrix game imply uncertainty relations for index-limited vectors. The stronger the player's strategy, the stronger the uncertainty relation is obtained.

Lemma 6.3.9 *Suppose the player has a winning strategy for DFT-Game*(n,l,k,B). *Then for any set T of size l and any set Ω of size r with $r \le k$, if a unit n-vector f with Fourier transform $\hat{f} = F_n f$ is ε_T-concentrated on T, then $\hat{f}$ is ε_Ω-concentrated on Ω with*

$$\varepsilon_\Omega > (1-\varepsilon_T)\frac{B}{n^{l/2}}.$$

Proof. Consider an arbitrary Fourier transform pair $(f,\hat{f})$ with f ε_T-concentrated on arbitrary set T of size l. $T = \{b+i \bmod n : 0 \le i \le l-1\}$ Consider an arbitrary set of indices Ω of size r with $r \le k$. By the definition of the Fourier matrix game and using the fact that DFT_n is a symmetric matrix, there exists $l \times l$ minor V of DFT_n with columns T rows avoiding Ω such that

$$|\det(V)|^2 \ge B^2.$$

Since

$$|\det(V)|^2 = \det(VV^*) = \prod_{i=0}^{l-1} \lambda_i(VV^*) = \prod_{i=0}^{l-1} \sigma_i(V)^2,$$

we conclude that the smallest singular value $\sigma_l(V) \ge \frac{B}{\sigma_1^{l-1}}$. Being a minor of unitary matrix F_n, $\sigma_1(\frac{1}{\sqrt{n}}V) \le 1$, so $\sigma_1(V) \le \sqrt{n}$. So

$$\sigma_l(\frac{1}{\sqrt{n}}V) \ge \frac{B}{n^{l/2}}.$$

By the min-max characterization of singular values given in Theorem 4.1.7, we have for any $l \times l$ matrix A that

$$\sigma_l(A) = \inf_{x \ne 0} \frac{||Ax||_2}{||x||_2}.$$

Hence

$$||\hat{f}_\Omega||_2 \ge ||\frac{1}{\sqrt{n}}V(f_T)||_2 \ge \sigma_l(\frac{1}{\sqrt{n}}V)||f_T||_2 > (1-\varepsilon_T)\frac{B}{n^{l/2}}.$$

□

6.3.3 An uncertainty relation for index-limited vectors

Let us generalize the notion of an index-limited vector to work with our ε-concentration notion of localization:

Definition 6.3.5. An n-vector f is called ε, l*-index-limited* if there exists g with $||g||_2 \le \varepsilon$ such that $f - g$ is l-index-limited.

Analogously to Theorem 6.3.6, one would hope to be able to improve Theorem 6.3.7 when restricting to index-limited vectors. For example, it should be possible to obtain lower bounds on concentration for set T and Ω with $|T| \cdot |\Omega| > n$ when dealing with index-limited vectors, eventhough Theorem 6.3.7 is trivialized beyond this range. A complete analysis of this problem is still open. In order to make steps towards this goal, using Corollary 6.2.7, we now give an uncertainty type relation that does manage to express non-trivial lower-bounds on concentration for scenarios where $|T| \cdot |\Omega| > n$.

Lemma 6.3.10 *Suppose the player has a winning strategy for DFT-Game*$^*(n,l,k,B)$. *Then for any unit n-vector f that is ε, l-index-limited and any set Ω of size r with $r \leq k$, $\hat{f} = F_n f$ is ε_Ω-concentrated on Ω with*

$$\varepsilon_\Omega > (1-\varepsilon)\frac{B}{n^{l/2}} - \varepsilon.$$

Proof. Consider an arbitrary Fourier transform pair $(f, \hat{f})$ and let $T = \{b + i \bmod \mathrm{n} : 0 \leq i \leq l-1\}$ be a contiguous set of indices containing $\mathrm{supp}(f-g)$ with g some vector with $||g||_2 \leq \varepsilon$, and $||f||_2 = 1$. Consider an arbitrary set of indices Ω of size r with $r \leq k$. By definition of the relaxed Fourier game and the fact that the Fourier matrix is symmetric, there exists $l \times l$ minor V of DFT_n with columns T and rows avoiding Ω such that

$$|\det(V)|^2 \geq B^2.$$

Similarly as in the proof of Lemma 6.3.9, we get for the smallest singular value σ_l of $\frac{1}{\sqrt{n}}V$

$$\sigma_l(\frac{1}{\sqrt{n}}V) > \frac{B}{n^{l/2}}.$$

Let Ω' be the rows of V. Write

$$(F_n f)_{\Omega'} = (F_n(f-g) + F_n g)_{\Omega'} = \frac{1}{\sqrt{n}}V(f-g)_T + (F_n g)_{\Omega'}.$$

By the min-max characterization of singular values given by Theorem 4.1.7, we have that

$$||\frac{1}{\sqrt{n}}V(f-g)_T)||_2 \geq \sigma_l(\frac{1}{\sqrt{n}}V)||f-g||_2 > (1-\varepsilon)\frac{B}{n^{l/2}}.$$

Since $||(F_n g)_{\Omega'}|| \leq \varepsilon$, we get by the triangle inequality that

$$||\hat{f}_{\Omega'}||_2 > (1-\varepsilon)\frac{B}{n^{l/2}} - \varepsilon.$$

Since Ω' is disjoint from Ω we conclude $\hat{f}$ is ε_Ω concentrated on Ω with $\varepsilon_\Omega > (1-\varepsilon)\frac{B}{n^{l/2}} - \varepsilon$. □

We now state our uncertainty relation for index-limited vectors.

Corollary 6.3.11 *Suppose f is a unit n-vector that is ε, l-index-limited with Fourier transform $\hat{f} = F_n f$. Then for any set Ω of size r with $r \leq \frac{n}{\pi}$ and $l + r \leq n$, $\hat{f}$ is ε_Ω-concentrated on Ω with*

$$\varepsilon_\Omega \geq (1-\varepsilon)\frac{e^B}{n^{l/2}} - \varepsilon,$$

where

$$B = \frac{(n-2r)\binom{l}{2}}{(n-r)(n-r-1)}(2-\ln 2-\frac{2\pi^2}{3n^2}) - \frac{\binom{l}{2}}{(n-r)(n-r-1)}(r^2\ln\frac{n}{r\pi}+r^2+\frac{r^4\pi^2}{36n^2}).$$

Proof. This follows immediately from the player strategy shown to exist in Corollary 6.2.7 by applying Lemma 6.3.10. □

The lower-bound on concentration on Ω is fairly weak, but we should stress this bound is given for any conceivable set Ω, not just contiguous ones. It is conceivable that the bound can be significantly improved by directly analyzing the ℓ_2-norms of random Vandermonde matrices instead of their determinants.

Assuming Ω to be contiguous should make even further improvements possible. This would qualify for doing the discrete analogue of the work done by Slepian [Sle78]. A first step has been taken by Grunbaum [Grü81], but this still remains to be a major open problem.

The most notable fact is that our theorem still yields non-trivial lower bounds on concentration in case both $l, r \gg \sqrt{n}$, which is a breaking point for typical straightforward calculations. For example, Theorem 6.3.7 yields a trivial lower-bound of $\varepsilon_\Omega \geq 0$ in case $|T|\cdot|\Omega| \geq n$, even if $|T|$ is assumed to be contiguous.

6.4 The Circulant Game* - an ad hoc strategy

Before proving our circuit lower bounds, we pause for the following intermezzo, in which we consider an ad-hoc strategy for winning the contiguous version of the circulant game. The results of this section are stand-alone, and can be skipped by the impatient reader.

Definition 6.4.1. A vector space $U \subseteq \mathbf{C}^n$ is ε, l*-flat* with respect to given orthonormal basis $u_0, u_1, \ldots, u_{n-1}$ if for every nonzero $x \in U$, writing $x = \sum_{i=0}^{n-1} a_i u_i$, there exists $i \in \{0, 1, \ldots, n-1\}$ such that

$$|a_i| \leq \varepsilon||x||_2 + \sum_{j=1}^{l-1}\left(|a_{i-j \bmod n}| + |a_{i+j \bmod n}|\right).$$

In the following, if the basis is omitted when using this definition, it is understood we are considering flatness with respect to the standard basis. If a space U is not ε, l-flat, we say it is ε, l*-bumpy*, and in this case any nonzero vector $x \in U$ violating the above inequality is called an ε, l*-bumpy vector*.

If for vector x, we have that $|x_i| > \varepsilon||x||_2$ and the previous or next $l-1$ positions are 0, we say x has a *pure* ε, l*-halfbump*. Analogously to the above we define a vector space U to be *purely* ε, l*-half-flat* if it contains no pure ε, l-half-bumpy vectors.

Bumpiness is a projective notion in the following sense:

Proposition 6.4.1 *If x is an ε, l-bumpy vector then so is λx, for any nonzero $\lambda \in \mathbf{C}$. The same holds with "bumpy" replaced by "purely half-bumpy".*

Lemma 6.4.2 *If $U \subseteq \mathbf{C}^n$ is purely ε, l-halfbumpy, then there exists unit $x \in U$, such that for any contiguous set of l rows R, there exists a contiguous set of l columns T, such that*

$$|\det(\mathrm{Circ}(x)_{R,T})| > \varepsilon^l.$$

Proof. Consider any unit purely ε, l-half-bumpy vector x in U, which exists by Proposition 6.4.1. Write $x = (x_0, x_1, \ldots, x_{n-1})$ w.r. to the standard basis. Without loss of generality assume that for some i, $|c_i| > \varepsilon$ and $|c_{i-j \bmod n}| = 0$, for $j = 1, 2, \ldots l-1$. Also wlog. assume $R = \{0, 1, \ldots, l-1\}$. Let $T = \{i, i+1, \ldots, i+l-1\}$. Let $M = \mathrm{Circ}(x)_{R,T}$. Then M is upper triangular with x_i on the diagonal, so $|\det(M)| = |x_i|^l > \varepsilon^l$. $\square$

Lemma 6.4.3 *If $U \subseteq \mathbf{C}^n$ is ε, l-bumpy, then there exists a unit vector $x \in U$ such that for any contiguous set of l rows R, there exists contiguous set of l columns T, such that*

$$|\det(\mathrm{Circ}(x)_{R,T})| > \varepsilon^l.$$

Proof. Consider any unit ε, l-bumpy vector x in U, which exists by Proposition 6.4.1. Write $x = (x_0, x_1, \ldots, x_{n-1})$ w.r. to the standard basis. Without loss of generality assume that for some i, $|c_i| > \varepsilon + \sum_{j=1}^{l-1} |c_{i-j \bmod n}|$, and also wlog. assume $R = \{0, 1, \ldots, l-1\}$. Let $T = \{i, i+1, \ldots, i+l-1\}$. Let $M = \mathrm{Circ}(x)_{R,T}$. M has x_i on all diagonal entries, so using the Greshgorin disc theorem (see e.g. [Bha97]), for each eigenvalue, $|\lambda_k(M)| \geq |x_i| - \sum_{j=1}^{l-1} |x_{i-j \bmod n}| + |x_{i+j \bmod n}| > \varepsilon$. $\square$

Definition 6.4.2. Let

1. $\rho(n, l, k) = \inf\{\varepsilon : \forall U \subseteq \mathbf{C}^n$ of co-dimension k that is ε, l-bumpy$\}$, and
2. $\rho'(n, l, k) = \inf\{\varepsilon : \forall U \subseteq \mathbf{C}^n$ of co-dimension k that is purely ε, l-half-bumpy$\}$.

The above defines an interesting notion in its own right, but with regards to the circulant matrix games we immediately obtain the following:

Theorem 6.4.4 *The player has winning strategies for Circ-Game*$^*(n, l, k, B^l)$*, where $B = max(\rho(n, l, k), \rho'(n, l, k))$.*

Proof. Suppose the adversary chooses a set of l rows R and subspace $U \subset \mathbf{C}^n$ of dimension $n-k$. Then we know that U is at least $\rho(n, l, k), l$-bumpy. Hence by Lemma 6.4.3 the player can choose $x \in U$ and contiguous set of rows T so that $\det(\mathrm{Circ}(x)_{R,T}) > \rho(n, l, k)^l$. Also we know that U is at least $\rho(n, l, k)'$-half-bumpy. Hence by Lemma 6.4.2 the player can choose $x' \in U$ and contiguous set of rows T' so that $\det(\mathrm{Circ}(x)_{R,T'}) > \rho'(n, l, k)^l$. $\square$

Proposition 6.4.5 *For any n,k,l with $l-1 < n-k$, $\rho'(n,k,l) \geq 2^{-n+l-1}$.*

Proof. Consider arbitrary U of co-dimension k. We can add $l-1$ equations of the form $x_i = x_{i-1} = \ldots = x_{i-l+2} = 0$ to define a subspace U' of U of nonzero dimension. Pick a unit $x \in U'$. For purpose of contradiction assume that x is $2^{-n+l-1}, l$-flat. This means $|x_{i+1}| \leq 2^{-n+l-1}$, $|x_{i+2}| \leq 2^{-n+l}$, etc.. so

$$||x||_2 \leq \sum_{k=0}^{n-l} |x_{i+k+1 \bmod n}| \leq \sum_{k=0}^{n-l} 2^{-n+l-1+k} \leq 2^{-n+l-1}(2^{n-l+1}-1) < 1,$$

which is a contradiction. □

6.5 Bilinear Circuit Lower Bounds

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.

6.5.1 Strong asymptotic strategies

Definition 6.5.1. Let ℓ_n be a function with $2\ell_n \leq n$. We say that the player has a *strong asymptotic winning strategy for the contiguous (or regular circulant) game with respect to ℓ_n*, if for every $\delta > 0$ there exists a $k > 0$ such that for all but finitely many n, the player has a winning strategy for Circ-Game$^*(n, \ell_n, \lfloor \frac{\ell_n}{k} \rfloor, 2^{-\delta n \log n})$, or Circ-Game$(n, \ell_n, \lfloor \frac{\ell_n}{k} \rfloor, 2^{-\delta n \log n})$, respectively.

Similarly we define the notion of a strong asymptotic winning stategy for the Fourier matrix game and its contiguous version. We have shown there to be ways of transferring strategies in both directions between the Fourier matrix game and the circulant game (Theorems 6.1.6 and 6.1.4). Some loss in the strength of the strategies was involved, but when considering strong asymptotic strategies this loss is inconsequential. Namely, we have the following theorem:

Theorem 6.5.1 *Let ℓ_n be a function with $\ell_n = O(\frac{n}{\log n})$. The player has a strong asymptotic strategy for winning the circulant game with respect to ℓ_n if and only if it has a strong asymptotic strategy for winning the Fourier game w.r. to ℓ_n. The same statement hold for the contiguous versions of both games.*

Proof. Suppose the player has a strong asymptotic strategy w.r. to ℓ_n for winning the Fourier matrix game. So for every $\delta_0 > 0$, there exists a $k > 0$, such that for all but finitely many n, the player can win

$$\text{DFT-Game}(n, \ell_n, \lfloor \frac{\ell_n}{k} \rfloor, 2^{-\delta_0 n \log n}).$$

By Theorem 6.1.6 this means the player can win

$$\text{Circ-Game}(n, \ell_n, \lfloor \frac{\ell_n}{k} \rfloor, 2^{-\delta_0 n \log n} \cdot G),$$

where with $\delta \approx 0.02$ being the absolute constant of Theorem 6.1.6, the loss-factor G is given by

$$G = \frac{\delta^{\ell_n/2}}{\sqrt{\binom{n}{\lfloor \frac{\ell_n}{k} \rfloor} 4^{\ell_n} (n - \lfloor \frac{\ell_n}{k} \rfloor)}} \geq 2^{-O(n)}.$$

To summmarize, for some constant $c > 0$, we have that for any $\delta_0 > 0$, there exists k such that for all but finitely many n, the player can win

$$\text{Circ-Game}(n, \ell_n, \lfloor \frac{\ell_n}{k} \rfloor, 2^{-\delta_0 n \log n - cn}).$$

This implies he/she has a strong asymptotic winning strategy for winning the circulant game with respect to ℓ_n.

For the converse direction, suppose the player does not have an asymptotic winning strategy for the Fourier matrix game w.r.t. ℓ_n. So there exists a $\delta > 0$ such that for any k, there are infinitely many n, for which the adversary can win

$$\text{DFT-Game}(n, \ell_n, \lfloor \frac{\ell_n}{k} \rfloor, 2^{-\delta n \log n}).$$

Then by Theorem 6.1.4, the adversary can win

$$\text{Circ-Game}(n, \ell_n, \lfloor \frac{\ell_n}{k} \rfloor, 2^{-\delta n \log n} \cdot F),$$

where we can crudely bound the loss-factor F by

$$F = \binom{n - \lfloor \frac{\ell_n}{k} \rfloor}{\ell_n} n^{-\ell_n/2} = 2^{O(n)}.$$

To summarize, there exist a constant $c > 0$ and a constant $\delta > 0$, such that for all k, for infinitely many n, the adversary can pick ℓ_n rows and a subspace U of dimension $n - \lfloor \frac{\ell_n}{k} \rfloor$, such that any $\ell_n \times \ell_n$ minor M of $\text{Circ}(a)$ with rows as determined by the adversary has

$$|\det(M)| \leq 2^{-\delta n \log n + cn}.$$

So for any δ' that is infinitisimally smaller than δ, provided n is large enough one gets a straight $|\det(M)| \leq 2^{-\delta' n \log n}$ bound. This implies the player does not have a strong asymptotic strategy for the circulant game w.r.t. ℓ_n.

The statement for the relaxed versions of the games holds because our "transfer" Theorems 6.1.6 and 6.1.4 hold with both regular games replaced by their contiguous versions. □

6.5.2 Main Result

Definition 6.5.2. A family $\{D_n\}_{n>0}$ where each D_n is an n-tuple of distinct positive real numbers $(d_1^n, \ldots, d_n^n)$ such that $\prod_{i=1}^n d_i^n = 1$ is called a *unit helper family*. If for all but finitely many n, the entries in D_n of value less than one are contiguous (in the circular sense), we say that $\{D_n\}_{n>0}$ is *asymptotically contiguous*.

Lemma 6.5.2 *Let ℓ_n be a function satisfying $\ell_n = O(\sqrt{n})$. Then the player has a strong asymptotic winning strategy for the circulant game w.r.t. ℓ_n.*

Proof. Let $\delta > 0$ be given. Say $\ell_n \le c\sqrt{n}$ for all large enough n. Set $k = 4c^2$. By Theorem 6.3.8 the player can win DFT-game$(n, \ell_n, \lfloor \frac{\ell_n}{k} \rfloor, B)$ with

$$B := (\frac{1}{2}\sqrt{n})^{\ell_n} \binom{n - \lfloor \ell_n/k \rfloor}{\ell_n}^{-1/2}.$$

Then applying Theorem 6.1.6 we obtain a strategy for winning Circ-game$(n, \ell_n, \lfloor \frac{\ell_n}{k} \rfloor, D)$ with

$$D \ge (\frac{1}{2}\sqrt{n})^{\ell_n} \binom{n - \lfloor \ell_n/k \rfloor}{\ell_n}^{1/2} \varepsilon^{\ell_n/2} \binom{n}{\lfloor \ell_n/k \rfloor}^{-1/2} (n - \lfloor \ell_n/k \rfloor)^{-\ell_n/2},$$

where ε is a constant approximately 0.02. This is certainly at least $2^{-\delta n \log n}$, provided n is large enough. □

Lemma 6.5.3 *Let ℓ_n be a function satisfying $\ell_n = O(n^{3/4})$. Then the player has a strong asymptotic winning strategy for the contiguous circulant game w.r. to ℓ_n.*

Proof. Let $\delta > 0$ be given. Let k be a constant to be determined later. By Corollary 6.2.8, provided n is large enough, the player has a winning strategy for DFT-Game$^*(n, \ell_n, \lfloor \frac{\ell_n}{k} \rfloor, B)$ for some B where

$$B \ge 2^{\Theta(\frac{\ell_n^2}{n} - \frac{\ell_n^4}{k^2 n^2} \log \frac{kn}{\ell_n})}.$$

Now applying Theorem 6.1.6, we obtain a strategy for winning Circ-game$(n, \ell_n, \lfloor \frac{\ell_n}{k} \rfloor, D)$ with

$$D \ge 2^{\Theta(\frac{\ell_n^2}{n} - \frac{\ell_n^4}{k^2 n^2} \log \frac{kn}{\ell_n})} \varepsilon^{\ell_n/2} \binom{n}{\lfloor \ell_n/k \rfloor}^{-1/2} (n - \lfloor \ell_n/k \rfloor)^{\ell_n/2},$$

where ε is a constant approximately 0.02. We see that it is possible to set k large enough to make B at least $2^{-\delta n \log n}$ for all large enough n. □

Theorem 6.5.4 *Let $\{D_n\}_{n>0}$ be a unit helper family, and suppose $\{\Gamma_n\}_{n>0}$ is a family of bounded-coefficient bilinear circuits such that for all n,*

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

Define $\ell_n = |D_n \cap (0,1)|$. We have that

1. *If $\ell_n = O(n^{\frac{1}{2}})$, then there exists $\gamma > 0$ so that $s(\Gamma_n) \geq \gamma n \log n$, for infinitely many n.*

2. *If $\ell_n = O(n^{\frac{3}{4}})$ and $\{D_n\}_{n>0}$ is asympotically contiguous, then then there exists $\gamma > 0$ so that $s(\Gamma_n) \geq \gamma n \log n$, for infinitely many n.*

3. *If $\ell_n = \Omega(n)$, then $s(\Gamma_n) = \Omega(n \log n)$.*

Proof. Let us first prove the third item. Suppose $\ell_n = \Omega(n)$. Hence there exists an ε_0 with $1 > \varepsilon_0 > 0$ so that for all but finitely many n, $\ell_n \geq \varepsilon_0 n$. In this case we think of the d_i^n that are larger than 1 as help gates as in [BL02]. There are at most $(1-\varepsilon_0)n$ many such help gates. Theorem 6.4 of [BL02] yields that $s(\Gamma_n) = \Omega(n \log n)$.

Let us now focus on the first two items. 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

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

then we can replace the constants which are bigger than 1 by bounded constant repeated additions. This takes at most $\sum_{j=\ell_n+1}^{n} \log d^n_{i_{jn}} = o(n \log n)$ 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 item 1 and 2 of the theorem are satisfied.

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

$$\prod_{j=1}^{\ell_n} d^n_{i_{jn}} < 2^{-\delta n \log n}. \tag{6.12}$$

Let us consider some large enough n for which (6.12) 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 $j > \ell$ and substitute $x_{i_j} = z_j / d_{i_j}$ otherwise. This yields a bounded coefficient bilinear circuit of size no bigger than $s(\Gamma)$, and it computes

$$(z_1, \ldots, z_\ell)\mathrm{diag}(d_{i_1}^{-1}, \ldots, d_{i_\ell}^{-1})M,$$

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

Now set $r = n - \lfloor \frac{\ell}{k_0} \rfloor$, where k_0 is a constant to be determined later. Let $f_1, \ldots, f_k$ be the linear forms in y of Γ. Lemma 5.1.1 provides us with a linear subspace U of dimension $n - \lfloor \frac{\ell}{k_0} \rfloor$ such that for any unit $b \in_R U$, we have that

$$\log \max_i |f_i(b)| \leq \frac{3s(\Gamma_n) + 3n}{2\lfloor \ell / k_0 \rfloor + 2}. \tag{6.13}$$

We think of the subspace U and the set I as chosen by the adversary.

For any unit $b \in U$ and any $\ell \times \ell$ minor M_0 of $\mathrm{Circ}(b)$ with rows I we can obtain from Γ_n a bounded coefficient *linear* circuit computing the $\mathbf{C}^m \to \mathbf{C}^m$ map

$$(z_1, \dots, z_\ell)\mathrm{diag}(d_{i_1}^{-1}, \dots, d_{i_\ell}^{-1})M_0$$

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

Hence the number of gates we added is at most

$$\ell \log \max_i |f_i(b)| \le \ell \frac{3s(\Gamma_n) + 3n}{2\lfloor \ell/k_0 \rfloor + 2} \le k_0 3s(\Gamma_n) + 3nk_0 \le 4k_0 s(\Gamma_n).$$

So the size of the resulting b.c. linear circuit is at most $5k_0 s(\Gamma)$.

So provided the player has a strong asymptotic winning strategy with respect to ℓ_n for the circulant game, we know a constant k_0 can be chosen such that there exist unit $b \in U$ and M_0 with

$$|\det(M_0)| \ge 2^{-\frac{\delta}{2} n \log n},$$

and if we know in addition that I is contiguous, then only a strong asymptotic winning strategy for the contiguous circulant game is required for the same fact. This would imply that

$$|\det(\mathrm{diag}(d_{i_1}^{-1}, \dots, d_{i_\ell}^{-1})M_0)| \ge 2^{\frac{\delta}{2} n \log n}.$$

However, by Morgenstern's bound any bounded coefficient circuit computing $\mathrm{diag}(d_{i_1}^{-1}, \dots, d_{i_\ell}^{-1})M_0$ then requires at least $\frac{\delta}{2} n \log n$ gates. Hence $s(\Gamma_n) \ge \frac{\delta}{10k_0} n \log n$.

In case $\ell_n = O(n^{\frac{1}{2}})$, we know by Lemma 6.5.2 that the player has a strong asymptotic winning strategy w.r.t. ℓ_n for winning the circultant game, which establishes item 2 of the theorem.

In case $\ell_n = O(n^{\frac{3}{4}})$, we know by Lemma 6.5.3 that the player has a strong asymptotic winning strategy w.r.t. ℓ_n for winning the contiguous circulant game. So provided the helper family is asymptotically contiguous, the set I is contiguous, and this establishes item 3. □

The model of computation that we are considering is admittedly exotic, but it should be noted that the model allows for up to $n-1$ unbounded constants, which is more than the εn unbounded constants the help gates technique in [BL02] manages to handle, where $0 < \varepsilon < 1$ cannot depend on n. We do have a strong restriction on where the unbounded constants can appear in the circuit, and there is the restriction of their product being at most $\Theta(1)$. As we observed before, the orbit model has computational power somewhere in between the general unbounded-coefficient model and the bounded-coefficient model. However, it seems unlikely that the model we consider is as powerful as the general unbounded-coefficient case in which the helper constants d_i's are unrestricted.

Stepping away from the orbit model, what Theorem 6.5.4 establishes with respect to the standard bounded-coefficient model of computation, is a general lower bound for entire families of bilinear mappings that appear in the $SL_n(\mathbf{C})$-orbit of the circular convolution mapping. Namely, the following corollary is immediate:

Corollary 6.5.5 *Let $\{D_n\}_{n>0}$ be a unit helper family, and define $\ell_n = |D_n \cap (0,1)|$. If ℓ_n satisfies one of:*

1. *If $\ell_n = O(n^{\frac{1}{2}})$, or*
2. *If $\ell_n = O(n^{\frac{3}{4}})$ and $\{D_n\}_{n>0}$ is asymptotically contiguous, or*
3. *If $\ell_n = \Omega(n)$,*

then for any family of bounded-coefficient bilinear circuits $\{\Gamma_n\}_{n>0}$ that computes

$$\{x^T \mathrm{Circ}(d_1 y_1, d_2 y_2, \ldots, d_n y_n)\}_{n>0},$$

there exists $\gamma > 0$ so that for infinitely many n, $s(\Gamma_n) \geq \gamma n \log n$.

Both in Theorem 6.5.4 and its Corollary 6.5.5 a knowledge gap is present, informally speaking, for ℓ_n in between $O(n^{3/4})$ and $\Omega(n)$. In Section 6.6 we will give some evidence that, at least in our framework, we will not be able to close this gap. The analysis involves *discrete prolate spheroidal sequences* and their remarkable eigenvalue properties. First however, we will generalize Theorem 6.5.4 to two-sided orbits.

6.5.3 Two-Sided Diagonal Case

So far we have focused attention on diagonal orbit circuits in which only one side, which w.l.o.g. was assumed to be the x-side, has helper constants. We now generalize Theorem 6.5.4 to the scenario in which we have helper constants on both the x and y-side. Obviously in this more general case we will observe an analogous "knowledge-gap" as is present in Theorem 6.5.4, e.g., as it comes to dealing with ℓ_n that are not $O(\sqrt{n})$. We will show however that, provided we have on both of the input sides of the circuit any of the favorable situations that we did manage to handle before, then we can still establish the $n \log n$ lower bound.

Definition 6.5.3. Call a unit helper family $\{D_n\}_{n>0}$ *good* if for $\ell_n = |D_n \cap (0,1)|$ one of the following holds:

1. $\ell_n = O(\sqrt{n})$, or
2. $\ell_n = O(n^{3/4})$ and $\{D_n\}_{n>0}$ is asymptotically contiguous, or
3. for some $\varepsilon > 1/2$, for all but finitely many n, $\ell_n \geq \varepsilon n$.

We have the following theorem:

Theorem 6.5.6 *Let $\{D_n\}_{n>0}$ and $\{E_n\}_{n>0}$ be unit helper families that are both good, and suppose $\{\Gamma_n\}_{n>0}$ is a family of bounded-coefficient bilinear circuits such that for all n,*

$$\Gamma_n(x_1 \cdot d_1^n, \ldots, x_n \cdot d_n^n, y_1 \cdot e_1^n, \ldots, y_n \cdot e_n^n) = x^T \mathrm{Circ}(y).$$

Then there there exists $\gamma > 0$ such that for infinitely many n, $s(\Gamma_n) \geq \gamma n \log n$.

Proof. Let $\ell_n = |D_n \cap (0,1)|$. 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

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

then we can replace the constants which are bigger than 1 on the x-side by bounded constant repeated additions. This takes at most $\sum_{j=\ell_n+1}^{n} \log d^n_{i_{jn}} = o(n \log n)$ additional gates. Hence we would obtain a one-sided orbit bilinear circuits of size $s(\Gamma_n) + o(n \log n)$ that uses $\{E_n\}_{n>0}$ as helper constants only. Since $\{E_n\}_{n>0}$ is good, we obtain the conclusion of the theorem by application of Theorem 6.5.4.

Hence assume we have $\delta > 0$ and an infinity set N of input sizes such that for all $n \in N$,

$$\prod_{j=\ell_n+1}^{n} d^n_{i_{jn}} \geq 2^{\delta n \log n}.$$

Let $\ell'_n = |E_n \cap (0,1)|$. For each n, Let $j_{1n}, \ldots, j_{nn}$ be such that

$$e^n_{j_{1n}} < e^n_{j_{2n}} < \ldots < e^n_{j_{nn}}.$$

If on the subsequence N (per abuse, we treat N as an infinite sequence of increasing numbers) we have that

$$\log \prod_{k=\ell'_n+1}^{n} e^n_{j_{kn}} = o(n \log n),$$

that is, if for any $\eta > 0$, for all but finitely many $n \in N$,

$$\log \prod_{k=\ell'_n+1}^{n} e^n_{j_{kn}} \leq \eta n \log n,$$

then for each $n \in N$, we can replace the unbounded constants on the y-side by effectively $o(n \log n)$ repeated additions. Hence obtaining for each $n \in N$, a one-sided orbit bilinear circuits of size $s(\Gamma_n) + o(n \log n)$ that uses $\{D_n\}_{n>0}$ as helper constants only. Since $\{D_n\}_{n>0}$ is good, we obtain the conclusion of the theorem by now continuing as in the proof of Theorem 6.5.4.

Hence assume we have $\delta' > 0$ and an infinity subsequence N' of N, such that for all $n \in N'$,

$$\prod_{k=\ell'_n+1}^{n} e^n_{j_{kn}} \geq 2^{\delta' n \log n}.$$

Case I: Suppose on N', $\ell'_n = \Omega(\ell_n)$, i.e. suppose there exists $\eta > 0$, such that for all but finitely many $n \in N'$, we have that $\ell'_n \geq \eta \ell_n$.

Subcase A: If $\{D_n\}_{n>0}$ is good because of clause three of the definition, Then also $\{E_n\}_{n>0}$ is good because of clause three. So we have $\varepsilon, \varepsilon' > 1/2$ such that for all but finitely many n, $\ell_n \geq \varepsilon n$ and $\ell'_n \geq \varepsilon' n$. Thinking of the helper constants as help gates as in [BL02], in this case the circuit contains at most $(2-(\varepsilon+\varepsilon'))n$ unbounded constants. This is bounded away from n by a constant factor, and thus via Theorem 6.4 of [BL02] we obtain the statement of the theorem.

Subcase B: If $\{D_n\}_{n>0}$ is good because of clause one of the definition, i.e., $\ell_n = O(\sqrt{n})$, then by Lemma 6.5.2 we know the player has a asymptotic winning strategy for the circulant game w.r. to ℓ_n. Consider large enough $n \in N'$, 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 $j > \ell$ and substitute $x_{i_j} = z_j/d_{i_j}$ otherwise. This yields a one-sided orbit circuit Γ' of size no bigger than $s(\Gamma)$, for which

$$\Gamma'(z_1, z_2, \ldots, z_\ell, y_1 e_1, y_2 e_2, \ldots, y_n e_n) = (z_1, \ldots, z_\ell)\mathrm{diag}(d_{i_1}^{-1}, \ldots, d_{i_\ell}^{-1})M,$$

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

Now set $r = n - \lfloor \frac{\ell}{k_0} \rfloor$, where k_0 is a constant determined large enough so that for any subspace U of dimension r, there exists value b for y so that M (with $y := b$) has an $\ell \times \ell$ minor M_0 with $\det(M_0) \geq 2^{-\frac{8}{2} n \log n}$. Since we have an asymptotic winning strategy for winning the circulant game with respect to ℓ_n there exists such k_0. Observe that enlarging k_0 only makes the circulant game easier for the player. This will enable us to also satisfy the requirement that k_0 is chosen so that $\frac{1}{k_0} < \eta$. Hence, in this case $\ell' > \eta\ell > \frac{\ell}{k_0}$. Let $J = \{j_1, j_2, \ldots, j_{\ell'}\}$, i.e., J is the set of indices j where $e_j > 1$. Let V be the coordinate subspace determined by set of equation $y_j = 0$, for all $j \in J$. The dimension of V is ℓ'. Modify circuit Γ' into a bounded-coefficient bilinear circuit Γ'' by setting $y_j = 0$, for all $j \in J$ and pushing down e_i constants that are smaller than 1 onto the wires. For y restricted to V the output of Γ'' and Γ' are identical.

Let $f_1, \ldots, f_k$ be the linear forms in y of Γ''. We still consider these as being defined over all of the variables $y_1, y_2, \ldots, y_n$, eventhough only ℓ' many y variables are used. This way we can still consider them as defining n-input polynomial function. Lemma 5.1.1 provides us with a linear subspace U of dimension $n - \lfloor \frac{\ell}{k_0} \rfloor$ such that for any unit $b \in_R U$, we have that

$$\log \max_i |f_i(b)| \leq \frac{3s(\Gamma'') + 3n}{2\lfloor \ell/k_0 \rfloor + 2}. \tag{6.14}$$

Now since

$$\dim[U \cap V] \geq (n - \lfloor \frac{\ell}{k_0} \rfloor) + \ell' - n = \ell' - \lfloor \frac{\ell}{k_0} \rfloor > 0,$$

we know there exists unit $b \in U \cap V$. We fix this b for the y inputs. Now the outputs of the linear forms in y are just constants. Multiplication with these constants will be replaced by repeated additions just as was done in Theorem 6.5.4. To give the details, for the minor M_0 of $\mathrm{Circ}(b)$ with rows I we can obtain from Γ_n a bounded coefficient *linear* circuit computing the $\mathbf{C}^\ell \to \mathbf{C}^\ell$ map

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

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

Hence the number of gates we added is at most

$$\ell \log \max_i |f_i(b)| \leq \ell \frac{3s(\Gamma''_n)+3n}{2\lfloor \ell/k_0 \rfloor + 2} \leq 5k_0 s(\Gamma_n).$$

Since

$$|\det(\mathrm{diag}(d_{i_1}^{-1},\ldots,d_{i_\ell}^{-1})M_0)| \geq 2^{\frac{\delta}{2} n \log n},$$

we conclude by Theorem 2.1.1 that $s(\Gamma_n) \geq \frac{\delta}{10k_0} n \log n$.

Subcase C: If $\{D_n\}_{n>0}$ is good because of clause two of the definition, i.e $\ell_n = O(n^{3/4})$ and $\{D_n\}_{n>0}$ is asymptotically contiguous, we have an asymptotically winning strategy for the contiguous circulant game. The proof proceeds similarly as in Subcase B. Having only a strong strategy for the contiguous game is sufficient, since in the case the d_i constants that are cancelled form a contiguous block, and we therefore are working with minors of $\mathrm{Circ}(y)$ that are restricted to a contiguous block of rows.

Case II: Assume the opposite of Case I, i.e. assume for any $\eta > 0$, there are infinitely many $n \in N'$ such that $\ell'_n < \eta \ell_n$. Let N'' be infinite subsequence of N' for which this holds. On N'', $\ell_n = \Omega(\ell'_n)$. This case now follows similarly as in Case I, but with the x and y-sides interchanged and using N'' instead of N'. □

Let us note that at the current time item 3 of Definition 6.5.3 does not read $\ell_n \geq \Omega(n)$, as would be desirable, since this is what we did for the one-sided case. The reason being that Theorem 6.4 of [BL02] allows for up to εn unbounded constants present anywhere in the circuit, with $\varepsilon < 1$. However, it is not clear how to generalize this result to allowing up to $\varepsilon_1 n$ unbounded constants on one side of the circuit (say the linear in x part) *together with* another $\varepsilon_2 n$ constants on the other side (the linear in y part), where potentially $\varepsilon_1 + \varepsilon_2 > 1$. The [BL02] result can only be applied provided $\varepsilon_1 + \varepsilon_2 < 1$.

6.6 Closing the gap

Our original hope was to get item 2 of Theorem 6.5.4 to work for any $\ell(n) = o(n)$. Unfortunately, the following appears to be true:

Conjecture 3. There exists $\varepsilon < 1$, such that for $\ell(n) = \lfloor n^{\varepsilon} \rfloor$, the player does not have a strong asymptotic strategy for winning the contiguous version of Fourier matrix game w.r. to $\ell(n)$.

Actually we believe that the cut-off point lies somewhere for ε near $4/5$, which we will support using results obtained in [Sle78]. We state the following conjecture:

Conjecture 4. If $\ell_n = \Omega(n^{4/5} \log^{1/5} n)$ and $\ell_n = o(n)$, then

1. the player does not have a strong asymptotic strategy for winning the contiguous circulant game w.r. to ℓ_n, and

2. neither does the player have a strong asymptotic strategy for winning the contiguous Fourier game w.r. to ℓ_n.

Given that we have fairly efficient ways of transferring strategies between the Fourier matrix game and the circulant matrix game, it is no suprise that items 1 and 2 of conjecture 4 are closely related. Theorem 6.5.1 shows that items 1 and 2 are equivalent for $\ell_n = O(\frac{n}{\log n})$. It is also clear that conjecture 4 implies conjecture 3.

So, let us have a look at conjecture 4. To analyze this, suppose $\ell = o(n)$. Consider playing DFT-Game$^*(n, \ell, r = \lfloor \frac{\ell}{c} \rfloor, B)$, for some large enough n, where c is some constant. For convenience let us assume that r is odd. Let $N = \{0, 1, \ldots, n-1\}$. Suppose the adversary chooses rows $R = 0, 1, \ldots, l-1$ and set C of columns $0, 1, \ldots, (r-1)/2$ and $n-1, n-2, \ldots, n-(r-1)/2$. In this case, with $F = n^{-1/2} DFT_n$, letting $K = F_{R,C} F^*_{R,C}$, we have that the entries of K are given by the *Dirichlet kernel*. Namely, for $0 \leq s \neq t \leq \ell - 1$, letting $f = \frac{s-t}{n}$, we have

$$
\begin{aligned}
n \cdot K_{st} &= \sum_{k=-(r-1)/2}^{(r-1)/2} e^{2\pi i k f} = e^{-2\pi i f (r-1)/2} \sum_{k=0}^{(r-1)} e^{2\pi i k f} = e^{-2\pi i f (r-1)/2} \frac{1 - e^{2\pi i r f}}{1 - e^{2\pi i f}} \\
&= \frac{e^{-2\pi i f (r-1)/2} - e^{2\pi i f (r+1)/2}}{1 - e^{2\pi i f}} = \frac{e^{-\pi i f (r-1)} - e^{\pi i (r+1) f}}{1 - e^{2\pi i f}} = \frac{e^{-\pi i f (r-1)} - e^{\pi i (r+1) f}}{1 - e^{2\pi i f}} \\
&= \frac{e^{\pi i f}(e^{-\pi i r f} - e^{\pi i r f})}{e^{\pi i f}(e^{-\pi i f} - e^{\pi i f})} = \frac{e^{-\pi i r f} - e^{\pi i r f}}{e^{-\pi i f} - e^{\pi i f}} = \frac{-2i \sin(r f \pi)}{-2i \sin(f \pi)} \\
&= \frac{\sin(r(s-t)\frac{\pi}{n})}{\sin((s-t)\frac{\pi}{n})},
\end{aligned}
$$

where we can also take this formula to define K_{st} for $s = t$, provided it is understood that one takes the limiting value $K_{st} = r/n$ in this case.

Let $M = I - K$. We have that λ is an eigenvalue of M if and only if $1 - \lambda$ is an eigenvalue of K. If $\det(M) = 2^{-\omega(n \log n)}$, since M is also given by $M = F_{R,N/C} F^*_{R,N/C}$, then by Binet-Cauchy (Theorem 4.1.3),

$$
\det(M) = \sum_{\substack{S \subset N/C \\ |S| = \ell}} |\det(F_{R,S})|^2.
$$

So any $\ell \times \ell$ minor of DFT_n that avoids rows C has magnitude at most $n^{\ell/2} 2^{-\omega(n \log n)} = 2^{-\omega(n \log n)}$ for $\ell = o(n)$, which means the player does not have a strong asymptotic strategy. All eigenvalues of K are in the interval $[0, 1]$. This is because the largest singular value of $F_{R,C}$ is at most 1. Hence we have the same for M. To give an upper bound on $\det(M)$ it thus suffices to show the largest eigenvalues of K cluster very close to 1.

At this stage we introduce the *discrete prolate matrix* studied by Slepian [Sle78]. For *bandwidth parameter* W, he defines the $N \times N$ matrix:

$$\rho(N,W)_{st} = \frac{\sin 2\pi W(s-t)}{\pi(s-t)}, \text{ for } 0 \le s,t \le N-1,$$

where it is understood that for $s = t$ the value on the r.h.s. equals $2W$. Let us take $W = \frac{r}{2n}$ and $N = \ell$. Then

$$\rho(\ell, \frac{r}{2n})_{st} = \frac{\sin r(s-t)\frac{\pi}{n}}{\pi(s-t)}, \text{ for } 0 \le s,t \le \ell-1.$$

It is certainly clear that for fixed $0 \le s \ne t \le \ell - 1$, since $\ell = o(n)$,

$$\begin{aligned} \lim_{n\to\infty} \frac{K_{st}}{\rho(N,W)_{st}} &= \lim_{n\to\infty} \frac{\pi(s-t)}{n\sin((s-t)\frac{\pi}{n})} \\ &= \lim_{n\to\infty} \frac{\pi(s-t)}{n(s-t)\frac{\pi}{n}} \\ &= 1, \end{aligned}$$

and that on the diagonal both matrices have all entries equal to $\frac{r}{n}$.

6.6.1 Asymptotic Equivalence

Actually a much stronger relation holds between the matrix K and $\rho_{N,W}$. Considered as families of matrices depending on the parameter n, these families are *asymptotically equivalent*. We give the definition from [Gra02], modified to give some flexibility regarding the dimension of the matrices:

Definition 6.6.1. Two sequences of $\ell(n) \times \ell(n)$ matrices A_n and B_n are said to be *asymptotically equivalent* if there exists a bound K such that

1. for all n, $||A_n||_2, ||B_n||_2 < K$, and
2. $\lim_{n\to\infty} \frac{||A_n - B_n||_F}{\sqrt{\ell(n)}} = 0$.

Note that for an $\ell(n) \times \ell(n)$ matrix A, $||A||_2 \le ||A||_F \le \sqrt{\ell(n)}||A||_2$, so the second condition in the definition is weaker than straightforwardly requiring that $\lim_{n\to\infty} ||A_n - B_n||_2 = 0$. For asymptotically equivalent matrices their eigenvalues have the same distribution in the following strong sense. Namely, Theorem 2.4 from [Gra02] can be tweaked for our scenario to read:

Theorem 6.6.1 *Let A_n and B_n be asymptotically equivalent families of $\ell(n) \times \ell(n)$ Hermitian matrices. Let m and M be such that for each n, all the eigenvalues $\lambda_i(A_n)$ and $\lambda_i(A_n)$of A_n and*

B_n are in the interval $[m,M]$. Let $F(x)$ be an arbitrary function continous on $[m,M]$. Then

$$\lim_{n\to\infty} \ell(n)^{-1}\sum_{i=1}^{\ell(n)} F(\lambda_i(A_n)) = \lim_{n\to\infty} \ell(n)^{-1}\sum_{i=1}^{\ell(n)} F(\lambda_i(B_n)).$$

To give two examples, for F being the identity function, the above states that the averages of the eigenvalues, if convergent, converge to the same value. For $F(x) = \ln x$, provided eigenvalues are positive, one would obtain

$$\lim_{n\to\infty} \ln\det(A_n)^{1/\ell(n)} = \lim_{n\to\infty} \ln\det(B_n)^{1/\ell(n)}.$$

We will now prove that the (families of) matrices K and $\rho(\ell, \frac{r}{2n})$ are asymptotically equivalent.

Theorem 6.6.2 *If $\ell_n = o(n)$, then for any sequence r_n, the families of $\ell_n \times \ell_n$ matrices $\{K(n)\}_n$ and $\{\rho(l, \frac{r_n}{2n})\}_n$ defined by*

$$K(n)_{st} = \frac{\sin(r_n(s-t)\frac{\pi}{n})}{n\sin((s-t)\frac{\pi}{n})}$$

and

$$\rho(\ell_n, \frac{r_n}{2n})_{st} = \frac{\sin r_n(s-t)\frac{\pi}{n}}{\pi(s-t)}, \quad \textit{for } 0 \le s,t \le \ell_n - 1,$$

are asymptotically equivalent.

Proof. First of all, since $K(n) = F_{R,C}F^*_{R,C}$, by submultiplicativity of the ℓ_2-norm we know $||K(n)||_2 \le 1$. From [Sle78] we know that $||\rho(\ell_n, \frac{r_n}{2n})||_2 \le 1$. Let $D(n)_{st} = \rho(\ell_n, \frac{r_n}{2n}) - K(n)_{st}$. Remains to show that

$$\lim_{n\to\infty} \frac{||D(n)||_F}{\sqrt{\ell_n}} = 0.$$

We use Taylor expansions (see [RW04] p.197): for every t, there exists $0 < \theta < 1$ such that

$$\sin t = t - \frac{t^3}{6}\cos\theta t.$$

Consider fixed $0 \le s,t \le \ell_n - 1$ and n. Let $\alpha_n = \sin r_n(s-t)\frac{\pi}{n}$, and let $\beta_n = \pi(s-t)$. Using Taylor, write:

$$\begin{aligned}\sin\frac{\pi}{n}(s-t) &= \frac{\pi}{n}(s-t) - \frac{\pi^3}{6n^3}(s-t)^3\cos\theta\frac{\pi}{n}(s-t)\\ &= \frac{\beta_n}{n} - \frac{\gamma_n}{n}\end{aligned}$$

with $0 < \theta < 1$ depending on n and $s-t$ and

$$\gamma_n = \frac{\pi^3}{6n^2}(s-t)^3\cos\theta\frac{\pi}{n}(s-t).$$

We have that

$$\begin{aligned}
D(n)_{st} &= \frac{\alpha_n}{\beta_n} - \frac{\alpha_n}{\beta_n - \gamma_n} \\
&= \frac{\alpha_n(\beta_n - \gamma_n)}{\beta_n(\beta_n - \gamma_n)} - \frac{\alpha_n \beta_n}{\beta_n(\beta_n - \gamma_n)} \\
&= \frac{-\alpha_n \gamma_n}{\beta_n^2 - \beta_n \gamma_n} \\
&= \frac{-\alpha_n \frac{\pi^3}{6n^2}(s-t)^3 \cos\theta \frac{\pi}{n}(s-t)}{\pi^2(s-t)^2 - \frac{\pi^4}{6n^2}(s-t)^4 \cos\theta \frac{\pi}{n}(s-t)} \\
&= \frac{-\alpha_n \frac{\pi}{6n^2}(s-t) \cos\theta \frac{\pi}{n}(s-t)}{1 - \frac{\pi^2}{6n^2}(s-t)^2 \cos\theta \frac{\pi}{n}(s-t)}.
\end{aligned}$$

Since $\ell_n = o(n)$, if n is large enough the denominator in above expression is aribtrarily close to 1. The numerator has two oscillating factors, but converges to 0 as determined by the dominating $\frac{s-t}{n^2}$ factor. Hence there exists constant $c > 0$ so that for large enough n,

$$\begin{aligned}
\frac{||D(n)||_F^2}{\ell_n} &\leq \frac{1}{\ell_n} \sum_{s=0}^{\ell_n - 1} \sum_{t=0}^{\ell_n - 1} c \frac{(s-t)^2}{n^4} \\
&\leq \frac{c \ell_n^4}{\ell_n n^4} \\
&\leq \frac{c l_n^3}{n^4}.
\end{aligned}$$

Since $\ell_n = o(n)$, we get that

$$\lim_{n \to \infty} \frac{||D(n)||_F}{\sqrt{\ell_n}} \leq \lim_{n \to \infty} \frac{c^{1/2} \ell_n^{3/2}}{n^2} = 0.$$

□

Asymptotic equivalence provides us with some preliminary evidence of the close similarity of $K(n)$ and $\rho(\ell, \frac{r}{2n})$, but by itself is not strong enough to resolve Conjecture 4. The task at hand is to carry over the asymptotic eigenvalue analysis done for $\rho(\ell, \frac{r}{2n})$ to $K(n)$. We will give some experimental data that, together with what is already known about $\rho(\ell, \frac{r}{2n})$, suggest indeed one could prove the truth of Conjecture 4 by doing a precise asymptotic eigenvalue analysis of $K(n)$. Such an analysis however, is still an infamous open problem in Fourier analysis, as we will discuss (see also [AET99, Grü81, CX84]).

6.6.2 Experimental Data

Let us do an experimental comparison between $K(n)$ and $\rho(\ell, \frac{r}{2n})$. Define the function

$$Q(n) = \frac{\ln|\det(I - \rho(\ell_n, \frac{\ell_n}{2n}))|}{\ln|\det(I - K(n))|}, \tag{6.15}$$

where we fix some $0 < \delta < 1$ and set $\ell_n = \lfloor n^\delta \rfloor$. Figure 6.1 show the function $Q(n)$ for $\delta = 0.5$. The function appears to converge to a value just less than 1, suggesting that for any function $f(n)$,

$$|\det(I - \rho(\ell_n, \frac{\ell_n}{2n}))| = 2^{f(n)} \Longrightarrow |\det(I - K(n))| = 2^{\Theta(f(n))}. \tag{6.16}$$

Appendix A contains some additional data for different values of δ. For δ close to 1 computational precision becomes an issue, and the range for n must be chosen to be smaller for data to be reliable. Nevertheless, we believe the data suggests that implication (6.16) holds with $\ell_n = \lfloor n^\delta \rfloor$, for any $0 < \delta < 1$. The asymptotics of the eigenvalues of $\rho(N,W)$ are well-understood. This translates to statements about the determinant of $I - \rho(N,W)$, which can be seen to be smaller than $2^{-cn\log n}$ for any fixed $c > 0$, if $\ell_n = \Omega(n^{4/5} \log^{1/5} n)$. We will show this momentarily. If indeed implication (6.16) holds for any $0 < \delta < 1$, then this would prove Conjecture 4, and rule out strong asymptotic strategies for the player once $\ell_n = \Omega(n^{4/5} \log^{1/5} n)$.

6.6.3 Eigenvalues of $\rho(N,W)$

In [Sle78], the following asymptotic values for the eigenvalues of $\rho(N,W)$ are given. For large N and k with

$$k = \lfloor 2WN(1-\varepsilon) \rfloor, \text{ with } 0 < \varepsilon < 1,$$

we get

$$1 - \lambda_k(\rho(N,W)) \sim e^{-CL_4/2} e^{-L_3 N}, \tag{6.17}$$

where

$$A = \cos 2\pi W$$

and $A < B < 1$ is determined so that

$$\int_B^1 \sqrt{\frac{t-B}{(t-A)(1-t^2)}} dt = \frac{k}{N}\pi.$$

Furthermore, we have

$$C = \frac{4}{L_2}\left[\frac{N}{2}L_1 + (2 + (-1)^k)\frac{\pi}{4}\right]_{\mathrm{mod}2\pi},$$

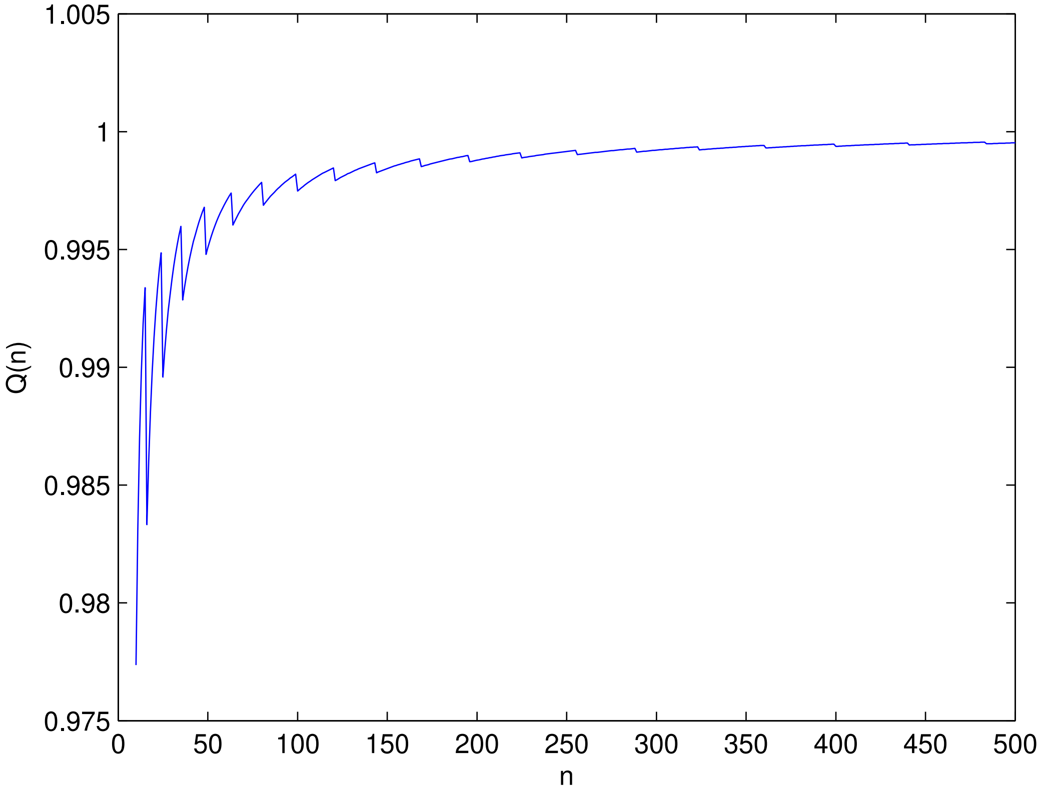

Figure 6.1: $Q(n)$ for $\delta = 0.5$.

where $[x]_{\text{mod}2\pi}$ is defined to be the number in $[0, 2\pi)$ congruent to x modulo 2π. The variables L_1, L_2, L_3 and L_4 determined by

$$L_1 = \int_B^1 P(t)dt \qquad L_2 = \int_B^1 Q(t)dt$$
$$L_3 = \int_A^B P(t)dt \qquad L_4 = \int_A^B Q(t)dt,$$

where

$$P(t) = \left| \frac{t-B}{(t-A)(1-t^2)} \right|^{1/2} \quad , \quad Q(t) = \left|(t-B)(t-A)(1-t^2)\right|^{-1/2}.$$

We apply this with $W = \frac{r}{2n} = \frac{\lfloor \ell/c \rfloor}{2n} \approx \frac{\ell}{2cn}$, $N = \ell$ and $\varepsilon = \frac{3}{4}$. We will assume n is some large enough number, and drop this index for the variables that depend on it. Note that in [Sle78] the bandwidth parameter W is taken to be fixed, but let us here provide evidence for our conjectures, modulo the assumption that this technicality can be resolved. We first perform a substitution $t = \sin\phi$ on the integral determining B. Define

$$f(t) = \sqrt{\frac{t-B}{t-A}}\frac{1}{\sqrt{1-t^2}}.$$

Then

$$\begin{aligned}\int_B^1 f(t)dt &= \int_{\arcsin B}^{\arcsin 1} f(\sin\phi)\cos\phi d\phi \\ &= \int_{\arcsin B}^{\pi/2} \sqrt{\frac{\sin\phi - B}{\sin\phi - A}}d\phi. \qquad (6.18)\end{aligned}$$

Note that since $A < B < 1$, we have that $\frac{\pi}{2} - \frac{\pi\ell}{cn} < \arcsin B < \frac{\pi}{2}$. Since $\ell = o(n)$ we can approximate (6.18) by

$$\frac{1}{2}(\frac{\pi}{2} - \arcsin B)\sqrt{\frac{1-B}{1-A}} \approx \frac{1}{\sqrt{2}}\sqrt{1-B}\sqrt{\frac{1-B}{1-A}} = \frac{1-B}{\sqrt{2-2A}}.$$

Approximating A by $1 - \frac{1}{2}(2\pi W)^2 = 1 - \frac{\pi^2\ell^2}{2c^2n^2}$, we get that

$$B \approx 1 - \frac{\pi^2\ell^2}{c^2n^2}(1-\varepsilon).$$

So for $\varepsilon = \frac{3}{4}$, $B \approx 1 - \frac{\pi^2\ell^2}{4c^2n^2}$, which is approximately in the middle of the interval $[A, 1]$. We will ignore the factor $e^{-CL_4/2}$ in (6.17), since this factor is certainly always less than 1. We will now give a lower bound on L_3:

$$\begin{aligned}L_3 &= \int_A^B \sqrt{\frac{B-t}{(t-A)(1-t^2)}}dt \\ &\geq \int_A^{\frac{A+B}{2}} \sqrt{\frac{B-t}{(t-A)(1-t^2)}}dt \\ &\geq \sqrt{\frac{B-A}{2}}\int_A^{\frac{A+B}{2}} \frac{1}{\sqrt{(t-A)(1-t^2)}}dt \\ &\geq \sqrt{\frac{B-A}{2}}\int_A^{\frac{A+B}{2}} \frac{1}{\sqrt{t-A}}dt \\ &= \sqrt{\frac{B-A}{2}}\int_0^{\frac{B-A}{2}} \frac{1}{\sqrt{\delta}}d\delta \\ &= \sqrt{\frac{B-A}{2}}2\sqrt{\frac{B-A}{2}} \\ &= B-A \\ &\approx \frac{\pi^2\ell^2}{4c^2n^2}.\end{aligned}$$

So we conclude that for the matrix $M' = I - \rho(N,W)$,

$$\lambda_k(M') \leq e^{-\ell L_3} = e^{-\frac{\pi^2 \ell^3}{4c^2 n^2}},$$

where $k = \lfloor WN/2 \rfloor \approx \frac{\ell^2}{4cn}$. Hence

$$\det(M') \leq e^{-\frac{\pi^2 l^5}{16c^3 n^3}} = e^{-\Theta(\frac{\ell^5}{n^3})}.$$

If this bound would carry over to the matrix M, which certainly seems plausible given the empirical evidence and also the asymptotic equivalence of the matrices K and $\rho(\ell, \frac{r}{2n})$, then in order for the player to have a strong asymptotic strategy, $\frac{\ell^5}{n^3}$ must be $o(n \log n)$. In other words, if $\frac{\ell^5}{n^3} = \Omega(n \log n)$, i.e., $\ell = \Omega(n^{4/5} \log^{1/5} n)$, then the player has no strong asymptotic strategy. This would then prove Conjecture 4.

Unfortunately, the asymptotic equivalence of K and $\rho(\ell, \frac{r}{2n})$ is by itself not strong enough to carry over eigenvalue results about the matrix M' to M while retaining the precise quantitive values provided by Slepian [Sle78]. We need to know about the precise rates of convergence. Also given our sensitive requirements on the clustering of eigenvalues of K near 1, that is, our need to observe eigenvalues that are *exponentially close* to 1, it seems difficult to carry over results about $\rho(\ell, \frac{r}{2n})$ to K using any standard perturbation techniques, such as Theorem 2.1.2. Namely, $||K - \rho(\ell, \frac{r}{2n})||_2$ does not converge to 0 exponentially fast. Note also that Theorem 6.6.2 actually shows for some constant $c > 0$, that

$$||D(n)||_2 \leq c^{1/2} \frac{\ell_n^2}{n^2}.$$

So for $\ell_n = o(n)$, we do have the ℓ_2-norm of the difference between K and $\rho(\ell, \frac{r}{2n})$ going to zero. The problem is that this convergence is not rapid enough: taking $s - t = 1$, we can see that $D(n)$ has entries that are roughly $\Omega(\frac{\ell_n}{n^2})$ and so certainly

$$||D(n)||_2 = \Omega(\frac{\ell_n}{n^2}).$$

Hence in an application of Theorem 2.1.2, the exponentially close clustering of eigenvalues of $\rho(\ell, \frac{r}{2n})$ near 1 would get lost in the approximation.

It appears that to know about the eigenvalues of K in the same precise manner that we know about the eigenvalues of $\rho(\ell, \frac{r}{2n})$, we need to carry out the *analogous analysis* as done by Slepian [Sle78]. However, as remarked before, this remains a major open problem [AET99]. A first step was taken by Grünbaum [Grü81] into resolving this issue. To give an idea, Slepian's results are based on the fact that $\rho(N,W)$ is closely related to the integral operator

$$L \equiv \int_W^W df' \frac{\sin N\pi(f - f')}{\sin \pi(f - f')}.$$

For L Slepian manages to give a diffential operator M that commutes with L. This implies that these operators have the same eigenfunctions. The eigenfunctions for L can be found by solving a diffential equation of Sturm-Liouville type. This then translates back to the eigenvectors and eigenvalues of $\rho(N,W)$.

For comparison, Grünbaum manages to give a tri-diagonal matrix M' that commutes with K. This then means K and M' have the same eigenvectors. Potentially, the eigenvectors of M' can be expressed in closed form by solving a difference equation, just like in the continuous scenario a diffential equation needed to be solved. This certainly is going to be a formidable task. Note that also some work towards this end has been done in [CX84], although at a more elementary level.

In any case, regardless of whether we can formally prove this, it seems inplausible that the knowledge-gap we observed in Theorem 6.5.4 and its Corollary 6.5.5 can be closed "all the way up to" $\ell_n = o(n)$ by the game strategy framework we devised. Our random Vandermonde matrix strategy gets us up to $\ell_n = O(n^{3/4})$. The above motivation leaves open the possibility one can perhaps push this up to $\ell_n = o(n^{4/5}\log^{1/5} n)$, but also suggests that at this point any DFT-Game strategy oriented argument will cease to work: at the $\ell(n) = n^{4/5+\delta}$ point the adversary appears to have the upper hand.

6.6.4 Equal Spacing Strategy and its limitations

The previous section gave evidence why it is plausible that for large enough $\varepsilon < 1$, there is no strong asymptotic strategy for the player with respect to $\ell(n) = \lfloor n^{\varepsilon} \rfloor$ in the contiguous Fourier matrix game. In this section, we will look at the particular scenario where the adversary chooses a contiguous block of disallowed columns, and where the player chooses his columns spaced at equal intervals in the remaining set of columns. This is a particularly instructive case to look at with regards to Conjecture 3. As noted before, we have some indication that this is the worst-case scenario as far as the adversary's choices are concerned. It will be interesting to see how well an intuitively good strategy, like spacing points equally in the allowed interval, fares in this case.

Instead of analyzing this scenario discretely, we will analyze the following continuous analogue. Let k be a constant, and suppose $\ell = o(n)$. Consider some large enough n. Say the adversary fixes an arbitrary *sector* S of the unit circle of angle $\frac{\ell}{kn}2\pi$. We will now try to find a set of ℓ points on the unit circle that are equally spaced in some sense and avoid the set S.

Let us start out with a set M of m equally spaced points on the unit circle. Let $R = M \cap S$. Say R has r points. Let $L = M/R$. We want L to have ℓ points, so assume a set M is chosen so $r = m - \ell$. Since the fraction of points of M that are in R will be proportional to the fraction that S is of the entire circle, we have that $\frac{r}{m} \sim \frac{\ell}{kn}$, so $r \sim \frac{\ell^2}{kn-\ell}$, and $m \sim \frac{\ell kn}{kn-\ell} = \frac{\ell}{1-\frac{\ell}{kn}}$.

For finite sets $A, B \in \mathbf{C}$, define

$$P_{AB} = \prod_{a\in A, b\in B, a\neq b} |a-b|.$$

We are interested in P_{LL}, since it relates to a Vandermonde determinant:

$$P_{LL} = |V(x_1, x_2, \ldots, x_l)|^2,$$

where $x_1, x_2, \ldots, x_n$ are the points in L. Observe that

$$P_{LL} = \frac{P_{LM}}{P_{LR}} \quad \text{and} \quad P_{LR} = \frac{P_{RM}}{P_{RR}},$$

so

$$P_{LL} = \frac{P_{LM}}{P_{RM}} P_{RR}.$$

Let x be a point contained in M. By symmetry, P_{xM} is the same for any point x of M. Now $P_{MM} = |\det(DFT_m)|^2 = m^m$, so $P_{xM} = m$. hence $P_{RM} = m^r$ and $P_{LM} = m^\ell$. Hence

$$P_{LL} = m^{l-r} P_{RR} = m^{2\ell-m} P_{RR}.$$

Taking the crude upper bound that any chord between points in R is of length at most $\frac{2\pi\ell}{kn}$, we get that

$$P_{RR} \leq \left(\frac{2\pi\ell}{kn}\right)^{r(r-1)}.$$

Hence (using that $m \leq 2\ell$ for large enough n)

$$\begin{aligned}
P_{LL} &\leq m^{2\ell-m} \left(\frac{2\pi\ell}{kn}\right)^{r(r-1)} \\
&\leq 2\ell^{2\ell} \left(\frac{2\pi\ell}{kn}\right)^{r(r-1)} \\
&= 2^{2\ell\log 2\ell + r(r-1)\log 2\pi\ell - r(r-1)\log kn} \\
&= 2^{-\Theta(\frac{\ell^4}{(kn)^2}\log kn) + \Theta(\frac{\ell^4}{(kn)^2}\log \ell) + \Theta(\ell\log\ell)}.
\end{aligned}$$

Hence if $\ell = \omega(n^{3/4})$, the dominant term $-\Theta(\frac{\ell^4}{(kn)^2}\log kn)$ will cause $-\log P_{LL}$ to be of growth order $\omega(n\log n)$.

Returning to the contiguous version of the Fourier matrix game, the only difference in the above scenario is that we cannot select arbitrary points on the unit circle, but must pick nth roots of unity. If n is large the player can select ℓ-many nth roots of unity that very closely approximate the equal spaced points in the set L. Our analysis indicates that such an equal spaced selection would not provide us with an asymptotically strong strategy for $\ell = \omega(n^{3/4})$, because in this case the resulting Vandermonde matrix has determinant of order $2^{-\omega(n\log n)}$.

From inspection of small cases one can deduce that the equal spacing strategy is not the optimal strategy against an adversary that chooses a contiguous block of columns. Slightly skewing the selected points towards the set of disallowed roots can yield a larger determinant. However, it appears unintuitive that by such minimal skewing one can produce an asymptotically strong strategy for *arbitrary* $\ell = o(n)$, given that the equal spacing stategy ceases to be useful at $\ell = \omega(n^{3/4})$.

It also should be emphasized that the equal spacing strategy works for $\ell = O(n^{3/4})$, but that this does not provide a simpler alternative for our random Vandermonde derived strategy. The equal spacing stategy assumes the set of disallowed columns to be contiguous, wheareas the random Vandermonde strategy get us up to $\ell = O(n^{3/4})$ with the disallowed ℓ many columns being in arbitrary configuration. To conclude, we have given evidence to support the claim that no strategy exists for the player for $\ell(n) = n^{4/5+\delta}$.

Chapter 7

Bounded Depth Circuits

In light of the inherent difficulty in proving general circuit lower bounds, various researchers have tried to make progress by adding one or more restricions to the computational model. One popular restriction has been the one in which the circuit is restricted to be of constant bounded depth. In this case arbitrary fan-in at gates is allowed in order to make the model nontrivial.

In boolean complexity the restriction to constant depth enables one to successfully prove exponential lower bounds [Ajt83, FSS81, Yao85, Hås89]. These papers constitute a body of work that is one of the shining gems of theoretical computer science. In the arithmetic world however, the situation is less bright. Currently only weak lower bounds, i.e., just barely non-linear, are known for constant depth circuits [RR03, Pud94].

Further progress has been made by adding additional restrictions to the computational model. Exponential lower bounds were proved for the size of monotone arithmetic circuits [SS77, MS80], and linear lower bounds are known for their depth [SS80, TT94]. In Chapter 3 we studied $\Sigma\Pi\Sigma$-formulas, which are of depth three. Over finite fields exponential lower bounds are known for $\Sigma\Pi\Sigma$-formulas, for example for computing the permanent and/or determinant polynomials [GK98, GR98]. Exponential lower bounds are known for *multi-linear* and *homogeneous* $\Sigma\Pi\Sigma$-formula [Nis91, NW96]. For unrestricted $\Sigma\Pi\Sigma$-formulas the only known lower bounds are the near-quadratic ones of [SW99], and the extensions of these results that we proved in Chapter 3. Note that Raz proved super-polynomial lower bounds on the size of general multi-linear formulas [Raz04a, Raz04b].

In this chapter we will proceed as follows. First we will prove two new versions of the classic "Derivative Lemma" of Baur-Strassen [BS82]. This lemma is used in combination with Strassen's degree method [Str73a, Str73b] to obtain *general* $\Omega(n \log n)$ arithmetical circuit lower bounds for single output functions. Originally Strassen's degree method works for proving lower bound on the size of straight-line programs computing several functions. The Derivative Lemma converts any straight-line program for a single function into one that computes the function together with all its partial derivatives with constant factor overhead, thereby enabling application of the degree method. Let us note that the lower bounds obtainable this way, for simple functions like $x_1^n + x_2^n + \ldots + x_n^n$ and less trivial functions like the determinant and the permanent, are the only general super-linear arithmetical circuit lower bounds known to date.

After exposition of our new versions of the Derivative Lemma, we will prove some lower bounds for a kind of bounded depth trilinear circuit, whose shape and form arises from application of our derivative lemmas. We call these kinds of circuits "interpolation circuits", and they compute linear combinations

$$\sum_{i=1}^{n} z_i p_i(x_1, x_2, \ldots, x_n) \tag{7.1}$$

of a collection of polynomials $p_1, p_2, \ldots, p_n$, where we will consider the coefficients z_i to be a "special" set of variables. These results take the ideas from [Lok95] a step further for our particular model.

Lokam considered bounded depth linear circuits with bounded coefficients, and bilinear formulas, which essentially are linear circuits of depth 2. We will prove size-depth trade offs for our special kind of bounded coefficient tri-linear circuit computing linear combinations of the form (7.1), where the polynomials p_i are bilinear polynomials of form $x^T Ay$.

Then in the last section, we will switch gears and prove a non-linear lower bound on the size of a bounded depth bilinear circuit computing circular convolution $x^T \mathrm{Circ}(y)$. To emphasize, the lower bound obtained there is without any restriction on the coefficients that are on the wires. We will employ a lemma from [RR03] about *superconcentrator* properties of the graph of a bilinear circuit, and we will combine this in a novel way with the *uncertainty principle* proved by Tao [Tao91], as it is known for cyclic groups of prime order, in order to obtain our lower bound.

7.1 Derivative Lemmas and Linear Interpolation

In this section inputs are not considered gates, fan-in is bounded by two and the size of circuits is measured by counting gates.

Definition 7.1.1. An **interpolation circuit** for computing polynomials $f_1, \ldots, f_m$ in variables $x_1, \ldots, x_n$ is defined to be an arithmetical circuit with inputs $x_1, \ldots, x_n$ and special inputs $b_1, \ldots, b_m$ that computes the linear combination $\sum_{i=1}^{m} b_i f_i$. Interpolation circuit size is defined by $i(f_1, \ldots, f_m) = s(\sum_{i=1}^{m} b_i f_i)$.

Our main interest is to consider interpolation circuits that have bounded coefficients. The reason is that interpolation circuits with bounded coefficients have, like the orbit models we defined before, computational power somewhere in between the bounded and unbounded coefficient model. An important technical detail is whether the circuit has access to a constant 1 gate. We will indicate explicitly by using superscipt ^{1} if that is the case. We use $f \leq^* g$ to indicate asymptotic ordering $f = O(g)$. Call a polynomial nontrivial if it is not equal to a variable or a constant. We have the following easy observations.

Proposition 7.1.1 *For any set of distinct nontrivial polynomials $f_1, \ldots, f_m$ we have that*

1. $i^{bc}(f_1, \ldots, f_m) \leq^* s^{bc}(f_1, \ldots, f_m)$.

2. $i^{bc,1}(f_1,\ldots,f_m) \leq^* s^{bc,1}(f_1,\ldots,f_m)$.

Applying the Baur-Strassen Derivative Lemma to a bounded coefficient interpolation circuit without access to 1, yields us a bounded circuit computing the separate functions with access to 1. Hence,

Proposition 7.1.2 *For any set of distinct nontrivial polynomials $f_1,\ldots,f_m$ we have that*

1. $s^{bc,1}(f_1,\ldots,f_m) \leq^* i^{bc}(f_1,\ldots,f_m)$.
2. $s^{bc,1}(f_1,\ldots,f_m) \leq^* i^{bc,1}(f_1,\ldots,f_m)$.

So we conclude that the bounded coefficient interpolation model with access to 1 is equally powerful as bounded coefficients with access to 1:

Corollary 7.1.3 $i^{bc,1}(f_1,\ldots,f_m) =^* s^{bc,1}(f_1,\ldots,f_m)$.

For linear circuits we can summarize the above situation as follows. We denote by $s^{bc,1}_{linear}$ the size of circuits that consists of addition gates computing homogeneous linear forms and addition gates computing constants, and allowing one multiplication gate at each output that multiplies a linear form and a constant gate. $i^{bc}_{bilinear}$ denotes the size of a bounded constant interpolation circuit which is bilinear. Observe that for a linear map $\lambda x.Ax$, $i^{bc}_{bilinear}(Ax) \leq^* s^{bc,1}_{linear}(Ax)$, because we can replace the multiplication gate with constant by performing repeated additions at the single output of the interpolation circuit. Conversely, $i^{bc}_{bilinear}(Ax) \geq^* s^{bc,1}_{linear}(Ax)$, by application of the Baur-Strassen Derivative Lemma, and then transferring constant multiplications to the outputs. Hence we have

Proposition 7.1.4 $s_{linear}(Ax) \leq^* s^{bc,1}_{linear}(Ax) =^* i^{bc}_{bilinear}(Ax) \leq^* s^{bc}_{linear}(Ax)$.

Examples can be given for which the interpolation model is more powerful than the bounded-coefficient model, when disallowing access to 1. For example $s^{bc}_{linear}(2^n x_1,\ldots,2^n x_n) = \Omega(n^2)$, whereas $i^{bc}_{bilinear}(2^n x_1,\ldots,2^n x_n) = O(n)$. The $i^{bc}_{bilinear}$-model can play a similar role as the orbit model in future research, namely provide an intermediate goal for proving lower bounds, somewhere in between the bounded and unbounded constant model.

Theorem 7.1.5 *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 $b_0,b_1,\ldots,b_n$ computing*

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

for all $j = 1\ldots m$.

Proof. 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, add a new input x_{n+1}, and 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 following 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 input variable x_{n+1} with the gate h. That is, add the gate h, and 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 performing the substitution

$$b_{n+1} \leftarrow \sum_{i=1}^{n} b_i \frac{\partial h}{\partial x_i}.$$

Since for any $j = 1 \ldots m$,

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

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 the bounded-constant circuit Γ. We conclude that Γ''' has size at most $5(s-1)+4 \leq 5s$, and that it is a bounded-constant circuit. □

Corollary 7.1.6 *In the statement of Theorem 7.1.5, if Γ does not use a constant 1 input gate, then neither does the constructed circuit.*

The above property is violated by the Baur-Strassen lemma. To give an example, a bounded coefficient bilinear circuit computing $x^T Ay$ is turned by that construction (when just constructing ∂x_i's) into a bounded-coefficient circuit computing Ay , but using a constant 1 input gate to build up constants, which get used at multiplication gates. This is an unfortunate fact, because current volume and spectral techniques, in particular Morgenstern's Theorem, for proving lower bounds on linear circuits get defeated by such usage of constants. Note that [NW95] overlooked this fact, and that the proof their "Corollary 3" is wrong. In this example, our proof of Theorem 7.1.5 simply reproduces a bounded-coefficient bilinear circuit computing $b^T Ay$. Applying the corollary for $m = 1$ yields the following:

Corollary 7.1.7 $i^{bc}(f, \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}) \leq^* s^{bc}(f)$.

We can also obtain a "transpose" of the above theorem.

Theorem 7.1.8 *Given a bounded coefficient circuit Γ computing $f_1, \dots, f_m$ at (non-input) gates of fanout zero in variables $x_1, \dots, x_n$ of size s, we can construct a bounded-coefficient circuit of size at most $5s$ with extra inputs $b_1, \dots, b_m$ 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 $j = 1 \dots n$, whenever these are not identically zero.*

Proof. We use induction to 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, add the new input x_{n+1}, and 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 $f'_1, \dots, f'_m$. By induction, we obtain a bounded coefficient circuit Γ'' with inputs $x_1, \dots, x_{n+1}$ and $b_1, \dots, b_m$ 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 $j = 1 \dots n+1$ 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 following equality for any $i = 1 \dots m$ and $k = 1 \dots n$:

$$\frac{\partial f_i}{\partial x_k} = \frac{\partial f'_i}{\partial x_k}[x_{n+1} \leftarrow h] + \frac{\partial f'_i}{\partial x_{n+1}}[x_{n+1} \leftarrow h] \cdot \frac{\partial h}{\partial x_k}.$$

Let Γ''' be the circuit obtained from Γ'' by replacing input variable x_{n+1} with the gate h. That is, add the gate h, and 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 computing $\sum_{i=1}^m b_i f'_i[x_{n+1} \leftarrow h] = \sum_{i=1}^m b_i f_i$ and for each $j = 1 \dots n+1$,

$$g_j = \sum_{i=1}^m b_i \frac{\partial f'_i}{\partial x_j}[x_{n+1} \leftarrow h].$$

By the chain rule, whenever x_j is not present in h, which is for all but at most two indices $j \in \{1, \dots, n\}$, $g_j = \sum_{i=1}^m b_i \frac{\partial f_i}{\partial x_j}$. For the remaining indices j, add gates to compute

$$g_j + g_{n+1} \cdot \frac{\partial h}{x_j} \quad = \quad \sum_{i=1}^m b_i \frac{\partial f'_i}{\partial x_j}[x_{n+1} \leftarrow h] + \sum_{i=1}^m b_i \frac{\partial f'_i}{\partial x_{n+1}}[x_{n+1} \leftarrow h] \cdot \frac{\partial h}{x_j}$$

$$= \sum_{i=1}^{m} b_i \left(\frac{\partial f'_i}{\partial x_j}[x_{n+1} \leftarrow h] + \frac{\partial f'_i}{\partial x_{n+1}}[x_{n+1} \leftarrow h] \cdot \frac{\partial h}{x_j} \right)$$

$$= \sum_{i=1}^{m} b_i \frac{\partial f_i}{\partial x_j}.$$

This can be done using at most 3 gates. Hence the final circuit has at most $5(s-1)+3 \leq 5s$ gates. □

7.1.1 Closed Form Bilinear Derivative Lemma

For a general homogeneous bilinear circuit computing the bilinear form $f = x^T Ay$ corresponding to a matrix A, as we noted in the previous section, application of the Baur-Strassen construction to obtain $(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n})$, which are the linear forms given by Ay, yields a circuit in which each gate computes a linear form in y, but using constant gates and allowing computed constants to multiply linear forms. This is unfortunate since for such circuits currently there are no lower bound techniques known. Hence there is no straightforward reduction of proving lower bounds for bilinear forms via the Baur-Strassen derivative lemma to the linear case. This contrasts with the successful Raz/Bürgisser-Lotz strategy for bounded-coefficient circuits, whose extension we studied in previous chapters. The culprit that causes the Bauer-Strassen construction to introduce these undesired multiplications with build-up constants can be seen to be linear part of the bilinear circuit *below the multiplication gates*. Here we will show that if this lower layer is a not a circuit but a *formula*, then we do have a derivative-lemma construction that leaves a homogeneous linear circuit with only addition gates.

In case the lower layer is a formula, we can assume wlog. that this lower layer consists of a single unbounded fan-in addition gate summing the outputs of all multiplication gates. Namely, multiplication gates with fan-out bigger than one can be duplicated so all multiplication gates have fan-out one, and this can be done with constant factor overhead. Next all constant on these fan-out wires can be pushed upward, resulting in a lower layer that just adds up the multiplication gates. Hence we can state our theorem as follows:

Theorem 7.1.9 *Suppose we have a linear circuit $C_1(x_1, x_2, \ldots, x_n)$ computing homogeneous linear forms $l_1(\vec{x}), l_2(\vec{x}), \ldots, l_k(\vec{x})$ and a circuit $C_2(y_1, y_2, \ldots, y_n)$ computing homogeneous linear forms $r_1(\vec{y}), r_2(\vec{y}), \ldots, r_k(\vec{y})$. Let f be a bilinear form given by*

$$f = \sum_{i=1}^{k} l_i(\vec{x}) r_i(\vec{y}).$$

Then we can construct a homogeneous linear circuit computing $\partial_x f := (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n})$ of size $O(s_1 + s_2)$, where s_1 and s_2 are the sizes of C_1 and C_2, respectively.

Proof. For each $i \in \{1, 2, \ldots, k\}$ write

$$l_i(\vec{x}) = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n,$$

with $a_{i1}, a_{i2}, \ldots, a_{in} \in \mathbf{C}$. Then

$$\begin{aligned}\frac{\partial f}{\partial x_s} &= \sum_{i=1}^{k} \frac{\partial l_i(\vec{x}) r_i(\vec{y})}{\partial x_s} \\ &= \sum_{i=1}^{k} \frac{\partial l_i(\vec{x})}{\partial x_s} r_i(\vec{y}) \\ &= \sum_{i=1}^{k} a_{is} r_i(\vec{y})\end{aligned}$$

In other words, defining the $k \times n$ matrix $A = (a_{pq})_{1 \le p \le k, 1 \le q \le n}$,

$$\partial_x f = [r_1(\vec{y}), r_2(\vec{y}), \ldots, r_k(\vec{y})]A,$$

so that

$$(\partial_x f)^T = A^T [r_1(\vec{y}), r_2(\vec{y}), \ldots, r_k(\vec{y})]^T.$$

Now circuit $\mathcal{C}_1$ computes $\lambda\vec{x}.A\vec{x}$. Since the circuit size of a matrix A and its transpose A^T are the same, we obtain a homegenous linear circuit $\mathcal{C}_3$ with k inputs and n outputs computing A^T. By the above we thus get a homogeneous linear circuit for $\partial_x f$ by composing circuits $\mathcal{C}_2$ and $\mathcal{C}_3$: first $r_1(\vec{y}), r_2(\vec{y}), \ldots, r_k(\vec{y})$ are computed by $\mathcal{C}_2$ and then these are taken as inputs to $\mathcal{C}_3$. Doing so, the n ouputs of $\mathcal{C}_3$ will yield $\partial_x f$. □

7.2 Bounded Depth Bilinear Interpolation Circuits

In this section we are going to consider bilinear interpolation 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 linear mappings computing separately for input vectors x and y. Both these mapping are computed by depth $d-1$ circuits. Multiplication gates are allowed, but are restricted to have exactly one of its inputs taken to be a z variable. We think of the z variables as if they were constants taken from the underlying field $\mathbf{C}$.

Say the outputs of these circuits are $l_1(x), ..., l_k(x)$ and $r_1(y), \ldots, r_k(y)$. These are actually linear in x or y, but may contain higher powers of z variables.

Then there are k multiplication gates computing $m_i = \ell_i(x) r_i(y)$ for $1 \le i \le k$. Finally there is a single unbounded fan in addition gate, taking inputs from all multiplication gates. 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 $(p)_{ij} =$ the coefficient of the monomial $x_i y_j$. Linear forms $\ell_i(x)$ and $r_i(y)$ are identified with row vectors. Under this identification we can thus say that each multiplication gate m_i computes $\ell_i^T r_i$. The function computed by the circuit is required to be of the form

$$\sum_{k=1}^{m} z_k (x^T A_k y) = x^T (\sum_{k=1}^{m} z_k A_k) y \quad ,$$

for certain complex $n \times n$ matrices A_k. In this situation, we say the circuit is an *interpolation circuit for computing matrices* $A_1, A_2, \ldots, A_m$. The idea is that, by varying the assignments of complex numbers to z, we can compute any of the bilinear forms $x^T My$, for any matrix M obtained as the linear combination $M = z_1A_1 + z_2A_2 + \ldots + z_mA_m$.

7.2.1 Preliminaries and Related Work

Definition 7.2.1 ([Lok95]). Let $1 \le r \le n$, for an $n \times m$ matrix A we define its ℓ_2-*r*-*rigidity* to be

$$\Delta_A^2(r) = min\{||A - B||_F^2 : B \text{ is an } n \times m \text{ matrix of rank at most } r\},$$

where $||A - B||_F$ denotes the Frobenius norm.

Lokam defined the ℓ_1-norm $||C||_1$ of a circuit to be the sum of the absolute values of all constants on the wires of C. For a matrix A, defining $||C^{[d]}(l_A)||_1$ to be the minimum ℓ_1-norm of a linear circuit of depth d computing the linear mapping l_A, he proved:

Lemma 7.2.1 ([Lok95]) *For any* $r \ge 1$*,*

$$||C^{[d]}(l_A)||_1 \ge r\left(\frac{\Delta_A^2(r)}{n}\right)^{1/2d}.$$

This results was later improved by Pudlák [Pud98] to

$$||C^{[d]}(l_A)||_2^2 \ge dn|\det A|^{2/dn},$$

where the ℓ_2-norm of a circuit is defined analogously to the ℓ_1-norm of a circuit.

One class of matrices for which we have good bounds on their ℓ_2-rigidity are Hadamard matrices.

Definition 7.2.2. An $n \times n$ matrix H is called a *generalized Hadamard matrix* if $HH^* = nI_n$.

When the entries of the matrix H are restricted to be ± 1 one gets the standard definition of a Hadamard matrix. As an example, the Fourier matrix DFT_n is a generalized Hadamard matrix. One has:

Theorem 7.2.2 ([Lok95]) $\Delta_H^2(r) = n(n - r)$.

Denoting by $C_1^{[d]}(l_A)$ the minimum number of wires of any depth d linear circuit with constants on the wires of norm at most 1 that computes $\lambda x.Ax$, one then has by Lokam's result that for any generalized Hadamard matrix H, $C_1^{[d]}(l_H) = \Omega(n^{1+\frac{1}{2d}})$, and by Pudlák's improvement $C_1^{[d]}(l_H) = \Omega(n^{1+\frac{1}{d}})$. Lokam also considered bilinear formulas, as introduced in [NW95], corresponding to a matrix A, which are formulas of form

$$b_A(x,y) = \sum_{i=1}^{m} x^T q_i p_i^T y,$$

where p_i and q_i are column vectors. The size $L(b_A)$ of the formula b_A is taken to be the total number of non-zero entries in the q_i and p_i vectors. These formulas are essentially depth 2 linear circuits: $s(b_A) = \Theta(C^{[2]}(l_A))$ [NW95]. So one gets for bilinear formula with bounded coefficients a lower bound $L_1^b(b_H) = \Omega(n^{5/4})$ via Lokam and $L_1^b(b_H) = \Omega(n^{3/2})$ via Pudlák's result, where H is a generalized Hadamard matrix. Lokam results yield the original bound proved in [NW95], and Pudlák's bound improves it.

7.2.2 Our Result

Theorem 7.2.3 *Let C be an interpolation circuit of structure as defined above with multiplication layer at depth d that computes $A_1,\ldots,A_m$. Then for $1 \le r \le n$, the number of wires of C that do not fan out from z variables is at least*

$$r\left(\sum_{i=1}^{m} \Delta_{A_i}^2(r)\right)^{1/(2d-1)} n^{-2/(2d-1)}.$$

Proof. Let C be given as indicated. Fix $1 \le r \le n$. Let S equal the number of wires of C that do not fan out of z variables. We call a gate or non z-variable special if the number of wires fanning out from it is at least S/r. Note there can be at most r special gates. No multiplication gate or the output gate can be special.

We now will consider what happens to a matrix A_i that is computed, in the sense that we defined, as we remove a special gate g. That is, temporarily fix $z_i = 1$ and $z_k = 0$ for $k \neq i$. Let $l_1, l_2, \ldots, l_k$ be the linear forms in x and $r_1, r_2, \ldots, r_k$ be the linear forms in y computed by the circuit, after this assignment. The output of the circuit with this assignment to z will be the bilinear form $x^T A_i y$. Now remove g and consider the modified output $x^T A_i^{new} y$. We will have six cases to consider.

Case 1: g is an input variable x_j. In this case we remove the wires fanning out from x_j. That means that for each i, $\ell_i^{new} = \ell_i$ with jth entry set to zero. Hence for each i, $m_i^{new} = m_i$ with row j zeroed out. Since each output A_i is simply a linear combination of the matrices m_i, we get $A_i^{new} = A_i$ with the jth row zeroed out, i.e. A_i gets modified by subtracting a matrix of rank 0 or 1.

Case 2: g is an input variable y_j. Similarly as above we can conclude each output gets modified by subtracting a matrix of rank at most one.

Case 3: g is a multiplication gate $m_i = \ell_i^T r_i$. The output gets modified by subtracting a scalar multiple of m_i. Observe that $\mathrm{rank}(m_i) \le 1$. So the output gets again modified by subtraction a matrix of rank at most one.

Case 4: g is an addition gate linear in x. Suppose gate g computes the linear form l. Then for each i, $\ell_i^{new} = \ell_i - \gamma_i l$, for certain scalars γ_i. Hence for each i, $m_i^{new} = (\ell_i^{new})^T r_i =$

$\ell_i^T r_i - \gamma_i l^T r_i$. 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 one. Hence again we have that each output is modified by a matrix of rank at most one.

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 matrix of rank at most one.

Case 6: g is a multiplication gate that has one of its inputs being a z variable. With z being assigned to, we can consider this gate to be an addition gate, so 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 that for each i, if we set all z's to be zero except $z_i = 1$, then the output of the circuit is a bilinear form $x^T(A_i - B_i)y$, where B_i is some matrix with rank at most r.

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:

$$\Phi = \sum_{s=1}^m \sum_{i=1}^n \sum_{j=1}^n |(A_s - B_s)_{ij}|^2. \tag{7.2}$$

For a given pair (x_i, y_j), there are at most $(S/r)^d \cdot (S/r)^{d-1}$ pairs of paths starting in x_i and y_j and that come together in the same multiplication gate. Then from that gate there is a single edge to the output. We can estimate (7.2) by summing over all these pairs of paths and over all assignments to z that set exactly a single $z_i = 1$. One pair of paths can contribute to *at most one* of the $A_i - B_i$. Namely, if the pair contains two multiplication gates with special z_i and z_j input with $i \neq j$, then contribution to $A_j - B_j$ and $A_i - B_i$ is zero, since in either case the other variable is set to zero. Since any constant on a wire 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

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

Observe that

$$\Phi = \sum_{s=1}^m ||A_s - B_s||_F^2 \geq \sum_{s=1}^m \Delta_{A_i}^2(r),$$

from which the theorem readily follows. □

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

Proof. By Theorem 7.2.2, we know that for a Hadamard matrix H, $\Delta_r^2(H) \geq n(n-r)$. Applying the above Theorem one gets that the number of wires not fanning out of z variables is at least

$$r\left(\sum_{i=1}^{m} \Delta_{A_i}^2(r)\right)^{1/(2d-1)} n^{-2/(2d-1)} \geq r\left(n \cdot n(n-r)\right)^{1/(2d-1)} n^{-2/(2d-1)} \tag{7.3}$$

$$= r(n-r)^{1/(2d-1)}. \tag{7.4}$$

Setting $r = n/2$ then yields the corollary. □

7.3 Bilinear circuits with unbounded coefficients of depth $O(1)$

In [RR03] a super-linear lower bound is proved on the number of edges of any bilinear circuit with *arbitrary coefficients* and *constant* depth computing matrix multiplication. Their result gives a lower bound on the number of edges present in the circuit below the multiplication gates. In other words, the bilinear circuit gets to perform two linear transformations at the inputs in the two different variable sets *free of charge*. In our orbit-related terminology, the circuits are taken to be of the form $\Gamma(Ex, Dy)$, where E and D are *arbitrary* matrices of *arbitrary* dimension. The proof technique is graph theoretic in nature. It make use of certain superconcentrator properties any circuit computing matrix product must posess.

In this section we will verify that this proof technique can also be successfully applied to the circular convolution function $x^T \mathrm{Circ}(y)$ which has been the main focus of our attention in previous chapters. Interestingly enough, we will essentially reduce the problem to a question about the superconcentrator properties of the discrete Fourier transform. Recall the definition:

Definition 7.3.1. An n-superconcentrator is a directed acyclic graph $G = (V, E)$ with n input nodes $I_G \subseteq V$ and n output nodes $O_G \subseteq V$ such that for every m, for every sets $X \subset I_G$, and $Y \subset O_G$, there exist m vertex disjoints paths from X to Y.

In can be seen that for prime p, any linear circuit computing DFT_p is a p-superconcentrator. Namely, it is well-known that any minor of DFT_p is non-singular [Tao91]. If there would exist any sets $X \subset I_G$ and $Y \subset O_G$ of size m such that there are strictly fewer than m vertex disjoint paths from X to Y, then the corresponding minor $DFT_{X,Y}^p$ would be singular.

We will not directly use this fact, but rather use the *discrete uncertainty principle* proved by Tao [Tao91], which was stated in Theorem 6.3.3. Nevertheless, the proof of this uncertainty principle relies on the fact that all minors of DFT_p are non-singular, for prime p, so superconcentrator properties of DFT_p are involved, albeit indirectly.

We now introduce some prerequisites taken from [RR03]. We will need some definitions about slow-growing functions and a lemma.

7.3.1 Prerequisites

Definition 7.3.2. For a function $f : \mathbf{N} \to \mathbf{N}$, define $f^{(i)}$ to be the composition of f with itself i times:

1. $f^{(0)}$ is the identity function,
2. $f^{(i)} = f \circ f^{(i-1)}$, for $i > 0$.

Futhermore, for f such that $f(n) < n$, for all $n > 0$, define

$$f^*(n) = \min\{i : f^{(i)} \leq 1\}$$

As in [RR03], the following set of extremely slow-growing functions $\lambda_d(n)$ will be used to express the lower bounds. Each $\lambda_d(n)$ is a monotone increasing function tending to infinity.

Definition 7.3.3. Let

1. $\lambda_1(n) = \lfloor \sqrt{n} \rfloor$,
2. $\lambda_2(n) = \lceil \log n \rceil$,
3. $\lambda_d(n) = \lambda^*_{d-2}(n)$, for $d > 2$.

For a directed acyclic graph G, V_G denotes the set of all nodes, I_G those with in-degree 0, and O_G those with out-degree 0. The depth of G is the length in edges of the longest path from I_G to O_G. Raz and Shpilka prove the following combinatorial lemma:

Lemma 7.3.1 ([RR03]) *For any $0 < \varepsilon < \frac{1}{400}$ and any layered directed acyclic graph G of depth d with more than n vertices and less than $\varepsilon \cdot n \cdot \lambda_d(n)$ edges, the following is satisfied:*

For some k with $\sqrt{n} \leq k = o(n)$, there exist subsets $I \subset I_G$, $O \subset O_G$, and $V \subset V_G$ for which $|I|, |O| \leq 5\varepsilon \cdot d \cdot n$ and $|V| = k$, and such that the total number of directed paths from $I_G \backslash I$ to $O_G \backslash O$ that do not pass through nodes in V is at most $\varepsilon \cdot \frac{n^2}{k}$.

7.3.2 Circuits for Circular Convolution

The circuits we will consider in this section are of the following form. They are bounded depth bilinear circuits with arbitrary fan-in and fan-out with arbitrary constants on the wires. We will assume our circuits are layered. We will give lower bounds on the number of edges present in the circuit below the multiplication gates. In other words, these circuits get two arbitrary linear transformations at the inputs for free. For use in this section only, we define:

Definition 7.3.4. For a bounded depth bilinear circuit C we define its size $s(C)$ to be the number of edges in the circuit between the multiplication gates and the outputs, and define by its depth $d(C)$ to be the length of a longest path in edges from a multiplication gate to an output.

We begin with the following easy proposition:

Proposition 7.3.2 *Any bilinear circuit of depth 1 computing circular convolution $x^T \mathrm{Circ}(y)$ has size $s(C) \geq n^2$.*

Proof. A circuit of depth 1 has a very simple structure. There are some number r of multiplication gates M_r computing products $M_r = L_r(x)R_r(y)$, where $L_r(x)$ and $R_r(y)$ are linear forms. Then there is one layer of output gates, each gate computing summation over some set of input multiplication gates.

We will argue that each output gate must be connected to at least n multiplication gates. For purpose of contradiction suppose that this is not the case. Say some output gate O_i takes input from $< n$ multiplication gates. Consider the subspace of dimension at least 1 defined by equations $L_j(x) = 0$, for each multiplication gate j attached to output O_i. We can select a non-zero vector a from this space such that for any assigment $y = b$,

$$(a^T \mathrm{Circ}(b))_i = 0.$$

This yields a contradiction, for example we can take b^T to be equal to a^* shifted by i, then $(a^T \mathrm{Circ}(b))_i = ||a||_2^2$, which is non-zero, since a is a non-zero vector. □

We now prove our main result for arbitrary constant bounded depth.

Theorem 7.3.3 *There exists $\varepsilon > 0$ such that if p is a prime number, any layered bilinear circuit with inputs $x = (x_0, x_1, \ldots, x_{p-1})$ and $y = (y_0, y_1, \ldots, y_{p-1})$ of depth d computing cyclic convolution $x^T \mathrm{Circ}(y)$ has size $s(C) \geq \varepsilon p \lambda_d(p)$.*

Proof. Consider the circuit computing

$$x^T \mathrm{Circ}(y) = x^T F_p \mathrm{diag}(DFT_p(y)) F_p^*.$$

We first apply substitutions $x^T := x^T F_p^*$ and $y = \frac{1}{n} DFT_P^* y$ at the inputs. This does not alter the circuit below the multiplication gates, but now we have a circuit computing

$$x^T \mathrm{diag}(y) F_p^*.$$

Let G be the directed acyclic graph of depth d given by the part of circuit below the multiplication gates. The set I_G is the collection of multiplication gates $M_i = L_i(x)R_i(y)$, where $L_i(x)$ and $R_i(y)$ are linear forms. Take $O_G = \{1, 2, \ldots, p\}$ to be the set of outputs of the circuit. Let $\varepsilon > 0$ be some small enough constant to be determined later. Trivially G has at least p vertices. Suppose that G has strictly fewer than $\varepsilon p \cdot \lambda_d(p)$ edges. Lemma 7.3.1 applies, and we obtain sets $I \subset I_G$, $O \subset O_G$ and $V \subset V_G$ such that

1. $|I|, |O| \leq 5\varepsilon dp$,
2. $|V| = k$, with $\sqrt{n} \geq k = o(p)$, and
3. the total number of directed paths from $I_G \backslash I$ to $O_G \backslash O$ that do not pass through nodes in V is at most $\varepsilon \frac{p^2}{k}$.

For each output node $i \in O_G \backslash O$, define $P(i)$ to be the number of multiplication gates in $I_G \backslash I$ for which there exists a directed path that bypasses V and reaches node i. Let R be a set of $r = 10k$ output gates with lowest $P(i)$ values. By averaging we get that

$$\sum_{r \in R} P(r) \leq \frac{r}{|O_G \backslash O|} \sum_{r \in O_G \backslash O} P(r) \leq \frac{r}{p - 5\varepsilon dp} \cdot \frac{\varepsilon p^2}{k} = \frac{10\varepsilon p}{1 - 5\varepsilon d}.$$

Let I' be the set of all multiplication gates in $I_G \backslash I$ for which there exist directed paths to nodes in R that bypass V. We can conclude that

$$|I'| \leq \frac{10\varepsilon p}{1 - 5\varepsilon d}.$$

Define a linear subspace W by the set of equations

$$R_i(y) = 0 \text{ for all } i \in I \cup I'.$$

For any fixed substitution for $y \in W$ the resulting circuit has all of the gates computing linear function in the x variables. Relative to a fixed choice for y, define linear subspace W_y by equations $g_v(x) = 0$ for all $v \in V$, where $g_v(x)$ denotes the linear form computed at gate v. Note that $\dim(W) \geq p - 5\varepsilon dp - \frac{10\varepsilon p}{1-5\varepsilon d}$ and $\dim(W_y) \geq p - k$, for each y. Now we have arranged that for each $y \in W$, and each $x \in W_y$,

$$(x^T diag(y) F_p^*)_r = 0, \tag{7.5}$$

for each $r \in R$.

In order to reach a contradiction, we will now argue that it is possible to select $y \in W$ and $x \in W_y$ such that some output in R is non-zero.

First of all, fix a vector $y \in W$ that has at most $5\varepsilon dp + \frac{10\varepsilon p}{1-5\varepsilon d}$ zeroes: this can be done because $\dim(W) \geq p - 5\varepsilon dp - \frac{10\varepsilon p}{1-5\varepsilon d}$. Let A be the set of indices i for which $y_i = 0$. Let $m = |A|$. Let W'_y be a subspace of W_y of dimension 1 obtained by adding equations to the defining set of W_y as follows. For the first stage add $x_i = 0$ for each $i \in A$. In a second stage, start adding equations that require $x_i = 0$ for $i \notin A$, until the dimension has been cut down to 1. Since we are starting out with a space of dimension $p - k$, after the first stage, the dimension will be cut down to at most $p - k - m$, so we will be able to add $x_i = 0$ in the second stage for at least $p - k - m - 1$. many i with $i \notin A$. Provided ε is small enough, since $k = o(n)$, $k + m$ will be less than a small fraction of p, so we are guaranteed that we can indeed complete this process still leaving a subspace of non-trivial dimension. Select an arbitrary non-zero vector x from W'_y. Observe that of the $p - m$ indices i not in A, x_i is non-zero for at most $k + 1$ entries, and that x_i is zero for all $i \in A$. So x_i is zero for each i for which $y_i = 0$. Since x itself is a nonzero vector there must be some place i where x_i and y_i are both nonzero.

Let $f = x^T diag(y)$ and $\hat{f} = fF_P^*$. We thus conclude that f is a non-zero vector, but that $|\text{supp}(f)| \leq k+1$.

By the discrete uncertainty principle for cyclic groups of prime order [Tao91], stated in Theorem 6.3.3, we have that

$$\text{supp}(f) + \text{supp}(\hat{f}) \geq p+1.$$

Hence the output vector of the circuit $\hat{f}$ is non-zero in at least $p+1-(k+1) = p-k$ places. Since R is of size $10k$, by the pigeonhole principle, there must be some output in R that is non-zero. This is in contradiction with equation (7.5). $\square$

Chapter 8

Conclusions

Given the inherent hardness in proving lower bounds for Boolean circuits, we embarked upon a study of arithmetical circuits. They bring the promise, more readily than Boolean circuits, of involving sophisticated concepts from algebra and algebraic geometry in a successful lower bound proof.

We continued the investigation of $\Sigma\Pi\Sigma$-formulas started by Shpilka and Wigderson [SW99]. There we presented a new technique for proving lower bounds by introducing the notion of *resistance* of a polynomial. Using this notion we proved tight lower bounds, amongst others, on the sum of nth powers polynomial $f = \sum_{i=1}^{n} x_i^n$. For any d, there are only n many dth order partial derivatives for this polynomial, which makes it hard to derive lower bounds using the partial derivatives technique from [SW99].

The partial derivatives technique yields lower bounds on multiplicative complexity only. In Chapter 3, we showed how this method can be extended to give lower bounds on total complexity, utilizing a closed form Baur-Strassen style derivative lemma for the $\Sigma\Pi\Sigma$ case. We have shown that this yields stronger lower bounds than those from [SW99], especially for low-degree polynomials. In certain cases, this improvement manages to lift trivial $\Omega(n)$ lower bounds, derived using the partial derivatives technique, to non-linear results. For instance, we showed for the elementary symmetric polynomial of degree 4 that $\ell_3(S_n^4) = \Omega(n^{4/3})$, and for the product-of-inner-product polynomial that $\ell_3(PIP_n^2) = \Omega(n^{4/3})$.

Both the partial derivatives technique and our resistance technique are limited to yielding quadratic lower bounds only. Such is tolerable when dealing with families of polynomials that indeed have $O(n^2)$ size $\Sigma\Pi\Sigma$-formulas, like $\sum_{i=1}^{n} x_i^n$ and (using Ben-Or's interpolation result) the elementary symmetric polynomials, but shows a severe gap in our knowledge when dealing with families of polynomials that are believed to be much more complex. As originally remarked in [SW99], currently we know of no super-polynomial lower bounds for the depth-three $\Sigma\Pi\Sigma$-formula model over fields of characteristic zero. For example, one would like to establish such bounds for the determinant and permanent polynomials. This contrasts with the situtation for Boolean circuits, for which we know exponential lower bounds for constant depth circuits [Ajt83, FSS81, Yao85, Hås89]. Future work on $\Sigma\Pi\Sigma$-formulas should be directed towards closing this discrepancy.

Open Problem 5. Prove a super-polynomial lower bound on the $\Sigma\Pi\Sigma$-formula size for an explicit function, e.g., the determinant or permanent, over a field of characteristic zero.

Suspiciously absent in current lower bound techniques for $\Sigma\Pi\Sigma$-formulas are random restriction type arguments, whereas all the results of [Ajt83, FSS81, Yao85, Hås89] proceed using random restrictions. Note that Raz manages to use random restrictions in conjunction with a partial derivatives based technique in his work on *multilinear* arithmetical formulas [Raz04a, Raz04b].

In Chapter 4 we investigated bilinear circuits with complex coefficients of $O(1)$ bounded magnitude. These circuits form a logical next place to investigate, given that linear circuits with bounded coefficients are essentially understood [Mor73], and given that unbounded coefficient linear circuits have confounded any form of non-trivial lower bound, even after 35 years of intense research activity.

We introduced the bilinear orbit circuit model. For $GL_n(\mathbf{C})$-orbits this model is at least as powerful as the unbounded coefficient case, but for $SL_n(\mathbf{C})$ it provided a challenging computational model to prove lower bounds for. The only known techniques for proving lower bounds for bounded coefficient bilinear circuits of [BL02, Raz02] fail to stand in this model, due to possible ill-conditioning of the free maps. The model was introduced because it allows a moderated study of a computation model in which more unbounded coefficients can be present than current techniques allow for. Secondly, lower bounds for the orbit circuit complexity of a single polynomials $p(x,y)$ translate to sweeping lower bounds on entire orbits of $p(x,y)$.

Our study was focused on the circular convolution mapping $\lambda x,y.\mathrm{Circ}(x)y$. We showed that if the free maps have condition number $O(1)$, then the the proof of [BL02] can be adapted to show that circular convolution still requires $\Omega(n\log n)$ size. Future work could be directed towards lifting this restriction, and prove general $SL_n(\mathbf{C})$-orbit lower bounds, but there are difficulties abound.

Namely, there is the apparent requirement in the random substitution technique to select the random input from a *subspace* U of some dimension εn with $\varepsilon < 1$, which seems to be about the only way to make the outputs of the linear forms on which substitution is performed "reasonably" bounded. Provided that is true, they can be replaced by "few enough" repeated additions, and this way a reduction to the (well understood) linear case is achieved. Unifying this modus operandi of the restriction technique with the abundance of ill-conditioned matrices present in $SL_n(\mathbf{C})$ is problematic. Geometrically speaking, only n-dimensional volumes retain the same volume under such transformations, but any lower dimensional volume can be arbitrarily stretched or squashed. In any configuration of the argument we considered this becomes an issue. Either the msv_r-volume of the target linear form one reduces to is negatively impacted, or, attempting to salvage this, the outputs of the linear forms on which one substitutes are ill-behaved, or vice-versa.

We managed to prove tight $\Omega(n\log n)$ size $SL_n(\mathbf{C})$-orbit lower bounds for circular convolution in case the circuit has precisely n multiplication gates. The proof shows that in this case the convolution theorem circuit, which uses the discrete Fourier transform and its inverse, is essentially unique.

We also considered orbits in conjunction with $\Sigma\Pi\Sigma$-formulas. The fact that lower bounds for $*$-complexity are maintained under such an extension is trivial. Interestingly enough, we showed some of our lower bounds on $+,*$-complexity can still be established, although some of them needed to be weakened.

Given the difficulties proving lower bound on $SL_n(\mathbf{C})$-orbit circuits, any attempt to lift the $O(1)$ condition number assumption perhaps is best attacked by first considering the diagonal $DL_n(\mathbf{C})$-orbit model as an important test case. Diagonal matrices of unit determinant can still be arbitrarily ill-conditioned. We managed to prove both a "one-sided" and "two-sided" diagonal orbit lower bound, modulo some extra assumptions about the amount and placement of helper constants less than 1 (see Theorems 6.5.4 and 6.5.6). We did so by introducing a novel game to be played on the DFT_n matrix, in which an adverary selects some rows that must be included and some columns that must be avoided. Then the goal was to find a minor satisfying these restrictions with maximum determinant. We related this game to several discrete uncertainty principles. In the contiguous case of playing this game, i.e., where an interval of rows is chosen, this led us to a randomized game strategy. We defined for any finite set $P = \{p_1, p_2, \ldots, p_k\}$ of points on the unit circle in the complex plane their *chordal product*

$$\mathcal{CP}(P) = \prod_{1 \le i < j \le k} |p_i - p_j|,$$

and asked the fundamental question:

Open Problem 6. For some large n, consider the set $\Omega = \{\omega_0, \omega_1, \ldots, \omega_{n-1}\}$ of all nth roots of unity on the unit cirlce in the complex plane. Let $R \subseteq \Omega$ be a given set of roots that are "off-limits". For any ℓ, what is the optimal strategy to select ℓ-many nth-roots of unity $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_\ell} \in \Omega \setminus R$ that maximizes $\mathcal{CP}(\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_\ell})$?

We approached the above problem by simply selecting the ℓ roots of unity uniformly at random. This yielded a result (Theorem 6.2.4) about random Vandermonde matrices with nodes on the unit cirle, which appears of independent mathematical interest. This strategy fares fairly well. In the terminology of Theorem 6.5.4, for $\ell_n = O(n^{3/4})$ this strategy was sufficiently strong to imply circuit lower bounds.

Related to the question of what is the optimal strategy, is the question what sets R in the above provide the worst-case scenario? That is:

Open Problem 7. For any k, ℓ, for what kind of sets $R \subseteq \Omega$ of size k is

$$\max_{\substack{S \subseteq \Omega / R \\ |S| = \ell}} \mathcal{CP}(S)$$

minimized, and what is its value ?

We have some indication that sets R that are contiguous provide this worst-case scenario, but the question is related to some long standing open problems [DS89] that turn out to be suprisingly hard to solve.

During our investigation, we also encountered an interesting numerical problem that is interesting for purely mathematical reasons. Suppose we define the following sequence of points $\{p_m\}_{m\geq 1}$ on the unit circle: $p_1 = 1$, and for $m > 1$, p_m is the first point q (if it exists) in counter-clockwise rotation around the unit circle after p_{m-1} such that $\prod_{i=1}^{m-1} |q - p_i| = 1$. This sequence arose in trying to devise a strategy that packs in points in a greedy manner, by adding a point each time, but ensuring that the added point has good chord-product with the previously added points. For those purposes, we also considered the modification of the above problem in which there was some given sector on the unit circle off-limits.

In any case, the interesting feature is that the sequence $\{p_m\}_{m\geq 1}$ appears to be infinite, and appears to enjoy a nice $\Theta(m^{1/2})$ growth (when seen in radians). It would be nice to give a closed form expression for the points in this sequence. The sequence for the modified problem is a little more erratic, but also appears to be infinite for "reasonable" disallowed sectors.

Beyond the $\ell_n = O(n^{3/4})$ growth rate a better strategy is required than random selection, but as we posed in Conjecture 4, we do not believe there exists a strategy that can deal with arbitrary $\ell_n = o(n)$. Conjecture 4 can be settle if one manages to carry over the asymptotic eigenvalue analysis of the prolate matrix of [Sle78] to the discrete-to-discrete case. We have made the conjecture plausible both from an empirical and theoretical standpoint. Carrying out the discrete analogy of the eigenvalue analysis of [Sle78] however, will be no easy task. See e.g. [Grü81, CX84, AET99]. In any case, this is an interesting problem in Fourier analysis, but from the theoretical computer science point of view, it would be more interesting to see whether one can devise an alternative lower bound argument that circumvents the issue.

As far as the contiguity assumption is concerned, one way to remove it, would be by giving a *reduction* that converts a circuit for $\pi(x^T)\mathrm{Circ}(y)$ into one for $x^T\mathrm{Circ}(y)$, using only $o(n\log n)$ additional circuit hardware. It is not clear whether this can be done. We showed that one certainly cannot in general convert a circuit for $\pi(x^T)\mathrm{Circ}(y)$ into one computing $x^T\mathrm{Circ}(y)$ by permuting the y-inputs and outputs. This would only work for permutations of form $\pi(i) = b + gi$, where g is a generator of the additive group of integers modulo n (See Appendix B).

In any case, if it is true that in the unbounded coefficient model the size of any bilinear circuit computing $x^T\mathrm{Circ}(y)$ is $\Omega(n\log n)$, then it is also true that any orbit circuit $\Gamma(Dx, Ey)$ with D and E diagonal and of unit determinant has size $\Omega(n\log n)$. We have managed to prove the latter under some additional restrictions, but still left to be resolved is the situation for general diagonal maps of determinant one:

Open Problem 8. Prove that any bilinear orbit circuit $\Gamma(Dx, Ey)$, where D and E are diagonal with unit determinant that computes circular convolution $x^T\mathrm{Circ}(y)$, must have size $s(\Gamma) = \Omega(n\log n)$.

The presence of arbitrary diagonal matrices D and E of unit determinant defeats any of the known volumetric techniques [BL02, Raz02]. Such is the case essentially because the matrix D can be *highly* ill-conditioned, making it hard to find "good" minors (in the sense of having large determinant) of the matrix $\mathrm{Circ}(a)$ that are in the "right" place. For the result in [BL02], it is sufficient to argue the *existence* of a good minor, whereas in the orbit model one seems to be forced to argue existence of good minors in a certain place of the matrix. The results we obtained still manage to strengthen [BL02]. Provided we made some extra assumptions about D, we could indeed locate such good minors of $\mathrm{Circ}(a)$ in the required place, the way our argument demanded.

We have tried to push the restrictions on D as far as possible, but for the kind of volumetric technique we were pursuing, we met a roadblock in trying to win our matrix games under extreme circumstances, because of phenomena related to the prolate spheroidal wave functions in [Sle78].

The *real* question is how far any kind of volumetric technique will carry in the orbit model. It seems non-volumetric techniques are called for, but that might be tantamount to proving lower bounds in the unbounded coefficient model. As a main goal in our orbit model setup, still open is the following problem:

Open Problem 9. Prove that any bilinear orbit circuit of form $\Gamma(Dx, Ey)$ (or $\Gamma(Dx, y)$), where D and E have unit determinant, computing circular convolution $x^T \mathrm{Circ}(y)$ has size $s(\Gamma) = \Omega(n \log n)$.

For that matter, up to now we have concentrated on circular convolution, but more generally it would be desirable to solve:

Open Problem 10. Prove a non-linear lower bound on the size of any bilinear orbit circuit of form $\Gamma(Dx, Ey)$ (or $\Gamma(Dx, y)$), where D and E have unit determinant for computing some *explicitly* defined bilinear map.

Then there is of course the holy-grail of proving lower bounds for the undbounded coefficient model for bilinear or low degree functions, which is equivalent to proving lower bounds in the orbit model for arbitrary diagonal maps. Even stronger than that (given that the linear maps do not count against the size), one may try to solve:

Open Problem 11. Prove a non-linear lower bound on the size of any bilinear orbit circuit of form $\Gamma(Dx, Ey)$ (or $\Gamma(Dx, y)$), where D and E are *arbitrary* $n \times m$ matrices for computing some *explicitly* defined bilinear map.

Finally, in Chapter 7 we considered bounded depth bilinear circuits and introduced interpolation circuits. We proved a Baur-Strassen style derivative lemma for this model, which has the added advantage that it does not introduce additional constants, as the regular derivative lemma notoriously does. We gave a closed form derivative lemma for a special kind of bilinear circuits, whose bottom layer is a formula. Results of [Lok95] were extended to a special kind of bilinear circuit. Finally, we proved a non-linear lower bound for bilinear circuits (with *unbounded* coefficients) computing circular convolution in case the input size n is a prime number. We did this using in the discrete uncertainty principle for cyclic groups of prime order [Tao91], and combining it with a "superconcentrator-lemma" of [RR03]. It would be interesting to see whether we can remove the assumption that n is prime. This might be hard, because only if n is prime do we know that DFT_n is a regular matrix, and thus that any linear circuit for it must be a superconcentrator.

Open Problem 12. Can one prove a non-linear lower bound for a bilinear circuit computing $\mathrm{Circ}(x)y$ in case the input size n is composite?

To summarize, we extended the partial derivatives method for $\Sigma\Pi\Sigma$-formulas. Some contributions were made to Fourier analysis and the theory of random matrices. We introduced the usage of uncertainty principles for proving lower bounds, in particular the strengthened uncertainty principle for cyclic groups of prime order [Tao91]. We extended the bilinear lower bounds of [BL02, Raz02]. Overall we have deepened the lower bound results of several published papers [SW99, BL02, Raz02, RR03, Lok95], and we have delineated mathematical obstacles to proving more general lower bounds.

Bibliography

[AET99] P. McCorquodale A. Edelman and S. Toledo. The future fast Fourier transform? *SIAM J. Sci. Comput.*, 20(3):1094–1114, 1999.

[Ajt83] M. Ajtai. Σ_1^1 formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[BCS97] P. Bürgisser, M. Claussen, and M.A. Shokrollahi. *Algebraic Complexity Theory*. Springer Verlag, 1997.

[Ben83] M. Ben-Or. Lower bounds for algebraic computation trees. In *Proc. 15th Annual ACM Symposium on the Theory of Computing*, pages 80–86, 1983.

[BGS75] T. Baker, J. Gill, and R. Solovay. relativizations of the P=NP? question. *SIAM J. Comput.*, 4:431–442, 1975.

[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.

[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.

[BLY92] A. Björner, L. Lovász, and A. Yao. Linear decision trees: volume estimates and topological bounds. In *Proc. 24th Annual ACM Symposium on the Theory of Computing*, pages 170–177, 1992.

[BM99] D.C. Brody and B. Meister. Discrete uncertainty relations. *J. Phys. A: Math. Gen.*, 32:4921–4930, 1999.

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

[Bür98] Peter Bürgisser. On the structure of Valiant's complexity classes. In *15th Annual Symposium on Theoretical Aspects of Computer Science*, volume 1373 of *lncs*, pages 194–204, Paris France, 25–27 February 1998. Springer.

[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.

[CLO92] D. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms*. Springer Verlag, 1992.

[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.

[CT65] J.W. Cooley and J.W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, 19:297–301, 1965.

[CX84] C. Chamzas and W.Y. Xu. On the periodic discrete prolate spheroidal sequences. *SIAM J. App. Math.*, 44:1210–1217, 1984.

[DS89] D.L. Donoho and P.B. Stark. Uncertainty principles and signal recovery. *SIAM J. App. Math.*, 49:906–931, 1989.

[Fer99] P. J. S. G. Ferreira. Superresolution, the recovery of missing samples, and Vandermonde matrices on the unit circle. In *Proc. Workshop on Sampling Theory and App.*, pages 216–220, 1999.

[FSS81] M. Furst, J. Saxe, and M. Sipser. Parity, circuits, and the polynomial-time hierarchy. In *Proc. 22nd Annual IEEE Symposium on Foundations of Computer Science*, pages 260–270, 1981.

[Gau75] W. Gautschi. Norm estimates for inverses of Vandermonde matrices. *Numer. Math.*, 23:337–347, 1975.

[GK98] D. Grigoriev and M. Karpinski. An exponetial lower bound for depth 3 arithmetic circuits. In *Proc. 13th Annual ACM Symposium on the Theory of Computing*, pages 577–582, 1998.

[Got66] D.H. Gottlieb. A certain class of incidence matrices. *American Mathematical Society*, 17:1233–1237, 1966.

[GR98] D. Grigoriev and M. Razborov. Exponetial complexity lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. In *Proc. 39th Annual IEEE Symposium on Foundations of Computer Science*, pages 269–278, 1998.

[Gra02] R.M. Gray. Toeplitz and circulant matrices: A review. Technical report, Stanford University, 2002.

[Grü81] F. A. Grünbaum. Eigenvectors of a Toeplitz matrix: discrete version of the prolate spheroidal wave functions. *SIAM J. Alg. Disc. Meth.*, 2:136–141, 1981.

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

[Hås86] J. Håstad. Almost optimal lower bounds for small-depth circuits. In *Proc. 18th Annual ACM Symposium on the Theory of Computing*, pages 6–20, 1986.

[Hås88] J. Håstad. *On the Computational Limitations of Small-Depth Circuits*. MIT Press, Cambridge, MA, 1988.

[Hås89] J. Håstad. Almost optimal lower bounds for small-depth circuits. In S. Micali, editor, *Randomness and Computation*, volume 5 of *Advances in Computing Research*, pages 143–170. JAI Press, Greenwich, CT, USA, 1989.

[Hun80] T.W. Hungerford. *Algebra*. Graduate Texts in Mathematics. Springer Verlag, 1980. 8th printing.

[IM02] K. Iwama and H. Morizumi. An explicit lower bound of $5n - o(n)$ for boolean circuits. In *International Symposium on Mathematical Foundations of Computer Science*, pages 353–364, 2002.

[Koi96] Pascal Koiran. Hilbert's Nullstellensatz is in the polynomial hierarchy. *Journal of Complexity*, 12(4):273–286, December 1996.

[Lok95] S. Lokam. Spectral methods for matrix rigidity with applications to size-depth trade-offs and communication complexity. In *Proc. 36th Annual IEEE Symposium on Foundations of Computer Science*, pages 6–15, 1995.

[Lok01] S. Lokam. Spectral methods for matrix rigidity with applications to size-depth trade-offs 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.

[MS80] M. Jerrum M. Snir. Some exact results for straight-line computation over semi-rings. Technical report, University of Edinburg, 1980. Research Report CRS-58-80.

[MS01] K. Mulmuley and M. Sohoni. Geometric complexity theory, P vs. NP, and explicit obstructions. In *Proceedings, International Conference on Algebra and Geometry, Hyderabad, 2001*, 2001.

[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.

[Nis91] N. Nisan. Lower bounds for non-commutative computation: extended abstract. In *Proc. 23rd Annual ACM Symposium on the Theory of Computing*, pages 410–418, 1991.

[NS82] M.A. Naimark and A.I. Stern. *The Theory of Group Representations*. Springer-Verlag, New York, NY, 1982.

[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.

[NW96] N. Nisan and A. Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6:217–234, 1996.

[Pud94] P. Pudlák. Communication in bounded-depth circuits. *Combinatorica*, 14:203–216, 1994.

[Pud98] P. Pudlák. A note on the use of the determinant for proving lower bounds on the size of linear circuits. Technical Report ECCC TR98-42, Electronic Colloquium in Computational Complexity, 1998.

[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.

[Raz04a] R. Raz. Multilinear formulas for permanent and determinant are of super-polynomial size. In *Proc. 36th Annual ACM Symposium on the Theory of Computing*, 2004. to appear; also ECCC TR03-067.

[Raz04b] R. Raz. Separation of multilinear circuit and formula size. In *Proc. 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 344–351, 2004.

[RR97] A. Razborov and S. Rudich. Natural proofs. *J. Comp. Sys. Sci.*, 55:24–35, 1997.

[RR03] A. Shpilka R. Raz. Lower bounds for matrix product, in bounded depth circuits with arbitrary gates. *SIAM Journal on Computing*, 32(2):488–513, 2003.

[RW04] L. Rade and B. Westergren. *Mathematics Handbook for Science and Engineering, 5th ed.* Springer Verlag, 2004.

[Sel01] K.K. Selig. Uncertainty principles revisited. Technical Report http://ww-lit-ma.tum.de/veroeff/quel/010.47001.pdf, Technische Universitat Munchen, 2001.

[SH05] R. Somaraju and L.W. Hanlen. Uncertainty principles for signal concentrations, 2005. at arxiv.org/cs.IT/0512030.

[Shp01] A. Shpilka. Affine projections of symmetric polynomials. In *Proc. 16th Annual IEEE Conference on Computational Complexity*, pages 160–171, 2001.

[SJ96] P. Stevenhagen and H.W. Lenstra Jr. Chebotarëv and his density theorem. *Mathematical Intelligencer*, 18(2):26–37, 1996.

[Sle78] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty - v: The discrete case. *Bell System Technical Journal*, 57(5):1371–1430, 1978.

[SS77] E. Shamir and M. Snir. Lower bound on the number of multiplication and the number of additions in monotone computations. Technical report, IBM Thomas J. Watson Research Center, 1977. Research Report RC6757.

[SS80] E. Shamir and M. Snir. On the depth complexity of formulas. *Math. Sys. Thy.*, 13:301–322, 1980.

[Str73a] V. Strassen. Berechnung und Programm II. *Acta Informatica*, 2:64–79, 1973.

[Str73b] V. Strassen. Die Berechnungskomplexetät von elementarysymmetrischen Funktionen und von Interpolations-Koeffizienten. *Numer. Math.*, 20:238–251, 1973.

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

[Tao91] T. Tao. An uncertainty principle for cyclic groups of prime order. Technical report, California Institute of Technology, 1991. arXiv.math.CA/0308286 v6.

[TT94] M. Tompa and P. Tiwari. A direct version of shamir and snir's lower bounds on monotone circuit depth. *Inf. Proc. Lett.*, 49(5):243–248, 1994.

[Val79a] L. Valiant. Completeness classes in algebra. Technical Report CSR-40-79, Dept. of Computer Science, University of Edinburgh, April 1979.

[Val79b] L. Valiant. The complexity of computing the permanent. *Theor. Comp. Sci.*, 8:189–201, 1979.

[Yao85] A. Yao. Separating the polynomial-time hierarchy by oracles. In *Proc. 26th Annual IEEE Symposium on Foundations of Computer Science*, pages 1–10, 1985.

Appendix A

The following figures refer to the function $Q(n)$ defined in equation (6.15).

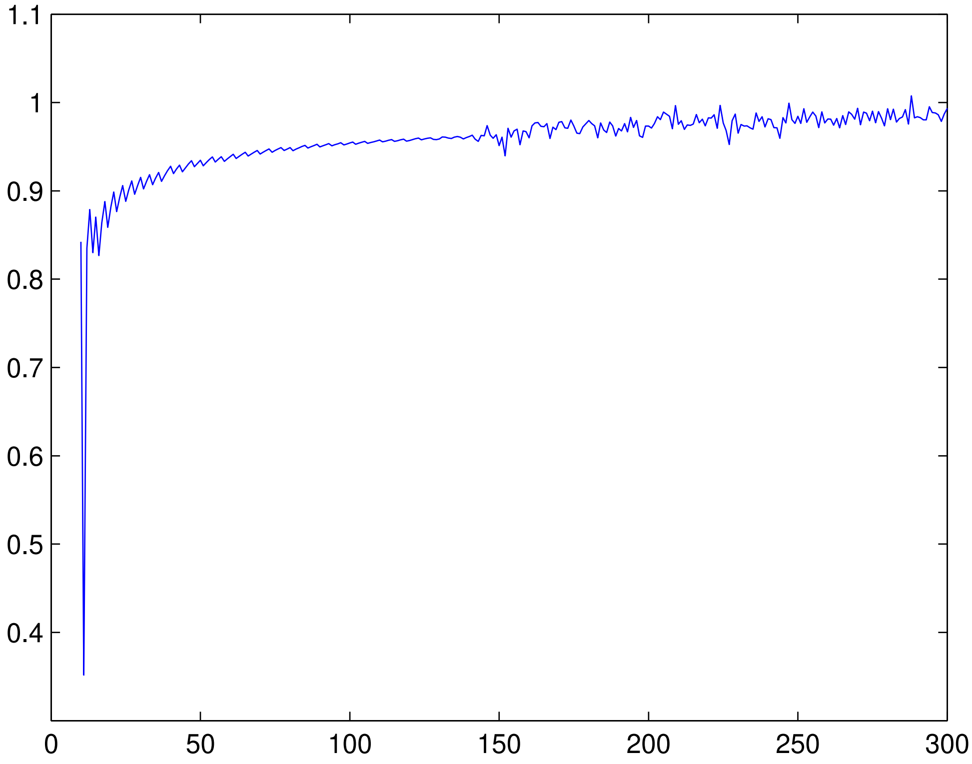

Figure 1: $Q(n)$ for $\varepsilon = 0.75$.

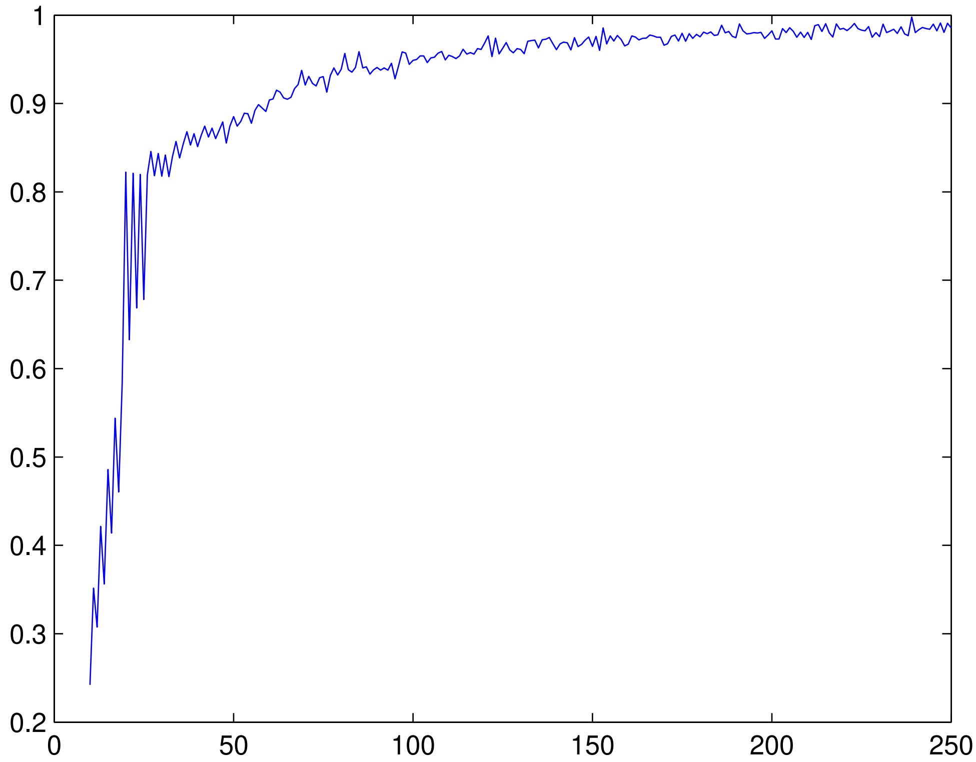

Figure 2: $Q(n)$ for $\varepsilon = 0.8$.

Appendix B

We refer to [Hun80] for the group theoretical notions used in the following.

In this Appendix we prove Theorem 5.3.1, which stated that for any n, the retrievable permutations form a group, and are precisely those permutation $\pi : \mathbf{Z}_n \to \mathbf{Z}_n$ for which there exists $b, g \in \mathbf{Z}_n$ with g relatively prime to n such that for each $i \in \mathbf{Z}_n$,

$$\pi(i) = b + gi.$$

Note that numbers g that are relatively prime to n form precisely all generators of the additive group $\mathbf{Z}_n$.

Proof. (of Theorem 5.3.1). In the following all indices of variables are considered elements of $\mathbf{Z}_n$, and all arithmetic with indices takes place within this additive group. Wlog. we do the proof with n-vectors of variables indexed as $x = (x_1, x_2, \ldots, x_{n-1}, x_0)$ and $y = (y_{n-1}, y_{n-2}, \ldots, y_0)^T$. In this case, $x\mathrm{Circ}(y)$ has variables lined up nicely so the kth entry $(x\mathrm{Circ}(y))_k$ has for each term the x-index and y-index summing to k. Namely, for each $k = 0, 1, \ldots, n-1$ we have that

$$(x\mathrm{Circ}(y))_k = \sum_{\substack{i,j \in \mathbf{Z}_n \\ i+j=k}} x_i y_j.$$

We first show any permutation $\pi : \mathbf{Z}_n \to \mathbf{Z}_n$ that is of the form

$$\pi(i) = b + gi,$$

for some $b \in \mathbf{Z}_n$ and generator g of the additive group $\mathbf{Z}_n$ is retrievable. Wlog. we can assume that $b = 0$, since b only produces a cyclic shift by b places. It is clear that a permutation π is retrievables iff π composed with a cyclic shift is retrievable. Define permutation π_1 by

$$\pi_1(i) = \pi(i - n) = g(i - n),$$

for each $i \in \mathbf{Z}_n$. Then we get that the jth entry of $\pi(x)\mathrm{Circ}(\pi_1(y))$ equals

$$x_{\pi(1)} y_{\pi_1(n-1+j)} + x_{\pi(2)} y_{\pi_1(n-2+j)} + \ldots x_{\pi(n-1)} y_{\pi_1(1+j)} + x_{\pi(0)} y_{\pi_1(j)}.$$

Consider an arbitrary term $x_{\pi(k)} y_{\pi_1(n-k+j)}$ of the above expression. It has indices summing as follows:

$$\begin{aligned} \pi(k) + \pi_1(n-k+j) &= gk + g(n-k+j-n) \\ &= gj. \end{aligned}$$

So all terms have indices summing to the same value gj. Since g is a generator of the additive group $\mathbf{Z}_n$, we see that all the n sum-values are presents at the n entries of $\pi(x)\mathrm{Circ}(\pi_1(y))$. In other words, $\pi(x)\mathrm{Circ}(\pi_1(y))$ is a permutation of $x\mathrm{Circ}(y)$.

Let us now do the converse directions. Suppose π is a retrievable permutation, and let π_1 be a permutation of the y variables such that $\pi(x)\mathrm{Circ}(\pi_1(y))$ is a permutation of $x\mathrm{Circ}(y)$. Since the x-indices and y-indices of each term of the j entry of $x\mathrm{Circ}(y)$ have to sum to the same number j, there must exists $b_0, b_1, \ldots, b_{n-1}$ so that the indices of each term of the jth entry of $\pi(x)\mathrm{Circ}(\pi_1(y))$ sum to b_j, for each $j = 0, 1, \ldots, n-1$. Observe that $\{b_0, b_1, \ldots, b_{n-1}\} = \{0, 1, \ldots, n-1\}$. The jth entry of $\pi(x)\mathrm{Circ}(\pi_1(y))$ equals

$$x_{\pi(1)}y_{\pi_1(n-1+j)} + x_{\pi(2)}y_{\pi_1(n-2+j)} + \ldots x_{\pi(n-1)}y_{\pi_1(1+j)} + x_{\pi(0)}y_{\pi_1(j)},$$

which we can rewrite as

$$\sum_{i=1}^{n} x_{\pi(n-i)}y_{\pi_1(i+j)}.$$

So we have the following condition satisfied:

$$(\forall j, i \in \mathbf{Z}_n), \quad \pi_1(i+j) + \pi(n-i) = b_j. \tag{1}$$

This implies that for any $s, t \in \mathbf{Z}_n$, we have

$$(\forall i \in \mathbf{Z}_n), \quad \pi_1(i+s) = \pi_1(i+t) + (b_s - b_t).$$

In particular,

$$(\forall i \in \mathbf{Z}_n), \quad \pi_1(i) = \pi_1(i+1) + (b_0 - b_1),$$

and

$$(\forall i \in \mathbf{Z}_n), \quad \pi_1(i+1) = \pi_1(i+2) + (b_1 - b_2),$$

which is equivalent to saying

$$(\forall i \in \mathbf{Z}_n), \quad \pi_1(i) = \pi_1(i+1) + (b_1 - b_2).$$

Repeating this for all s and t with $t = s+1$, we get there exists some number $g \in \mathbf{Z}_n$ so that $g = b_0 - b_1 = b_1 - b_2 = \ldots = b_{n-2} - b_{n-1} = b_{n-1} - b_0$. The number g must be a generator of $\mathbf{Z}_n$, since otherwise not every element of $\mathbf{Z}_n$ would be in the range of π_1. We can conclude that we can write

$$b_j = b_0 - gj,$$

for all $j = 0, 1, \ldots, n-1$. However, specifying condition (1) with $i = 0$, we have

$$(\forall j \in \mathbf{Z}_n), \quad \pi_1(j) + \pi(0) = b_j.$$

So π_1 is defined by

$$(\forall j \in \mathbf{Z}_n), \quad \pi_1(j) = (b_0 - \pi(0)) - gj.$$

Which implies by condition (1) that π is defined, for each $i = 0, 1 \ldots, n-1$, by

$$\begin{aligned}\pi(n-i) &= b_0 - \pi_1(i)\\ &= b_0 - (b_0 - \pi(0)) + gi\\ &= \pi(0) + gi.\end{aligned}$$

Hence we have that for each $i \in \mathbf{Z}_n$,

$$\begin{aligned}\pi(i) &= \pi(0) + g(n-i)\\ &= \pi(0) + (-g)i.\end{aligned}$$

Since $(-g)$ is also a generator of $\mathbf{Z}_n$, we conclude that π is of the form stated by the theorem.

By the above it can thus be seen that the retrievable permutations form a group R_n. Namely, composing $\pi_1(i) = b_1 + g_1 i$ with $\pi_2(i) = b_2 + g_2 i$ one gets

$$\begin{aligned}\pi_1(\pi_2(i)) &= b_1 + g_1\pi_2(i)\\ &= b_1 + g_1 b_2 + g_1 g_2 i.\end{aligned}$$

The generators of $\mathbf{Z}_n$ are precisely all integers (modulo n) that are relatively prime to n. So the product $g_1 g_2$ is again a generator, this showing the composition is of the required form. The inverse of a permutation $\pi(i) = b + gi$ is given by $\pi^{-1}(i) = c + hi$, where c is the unique number such that $gc = -b$, and h is the unique number so that $gh = 1$ in $\mathbf{Z}_n$. □

Each choice for b and g yield a distinct permutation π, so $|R_n| = n\phi(n)$, where ϕ is the *Euler totient function*, giving the number of natural numbers relatively prime to n. This is maximized for prime n, in which case $|R_n| = n^2 - n$. Modulo cyclic shifts, R_n is isomorphic to the *character group* $Z_n^\times$ (integers from $\{1, 2, \ldots, n-1\}$ relatively prime to n under multiplication) through *regular representation* $g \mapsto \pi(i) = gi$. Namely, letting H_n be the sugroup of all cyclic shifts, i.e., permutations of the form $\pi(i) = b + i$, then $R_n/H_n \simeq Z_n^\times$.