Lower Bounds for Small Depth Arithmetic and Boolean Circuits

Thesis submitted for the degree
Doctor of Philosophy

by

Amir Shpilka

Submitted to the Senate of the Hebrew University
August 2001

Lower Bounds for Small Depth Arithmetic and Boolean Circuits

Thesis submitted for the degree
Doctor of Philosophy

by

Amir Shpilka

Submitted to the Senate of the Hebrew University
August 2001

This work was carried out under the supervision of

Professor Avi Wigderson and Professor Noam Nisan

Acknowledgments

First of all I would like to thank my advisor Avi Wigderson. I benefited from knowing Avi in so many ways - Avi taught me a lot about complexity theory, introducing me to the most recent results in the field. Avi is a great lecturer and listening to his courses as well as being his teaching assistant was a great experience, and of course I enjoyed doing research with him. But most of all I would like to thank Avi for his kindness, his generosity, his infinite optimism and his encouragement. I believe that only the will to continue and work with Avi kept me in complexity for so (too ?) many years.

I would like to thank all the people with whom I shared ideas during various stages of my phd. Mostly I would like to thank Michael Ben Or, Nati Linial and Noam Nisan whose door was always open to me, who had to hear all the false/unpolished ideas that I had, and who suffered without complaining!

A special thank is due to Ran Raz with whom I had the pleasure of writing the paper presented in chapter 4. I was very fortunate to have the opportunity to work with Ran, because both of his professional skills and him being a great person to spend time with. I also thank him for inviting me to spend my post-doc years with him.

And of course how can one forget all the people of the theory lab at the Hebrew university. Thanks to them the lab was a place where you can always come to, meet friends, drink coffee and do nothing. A partial list of all the people who formed this theory pub is Ido Bergman, Eli Ben-Sasson, Yonatan Bilu, Tamir Chazan, Danny Gutfreund, Shlomo Hoory, Ron Lavi, Boaz Leslau, Avner Magen, Ahuva Mualem, Ofer Neiman, Elan Pavlov, Amir Ronen, Eyal Rozenman and Ronen Shaltiel.

I would like to thank Zehava Benbenishty, my mother in law, who spent many days babysitting Inbal. Last but not least, I would like to thank my wife, Carmit, for her unstoppable support and encouragement. But this is a family matter now.

Contents

Abstract

In this thesis we study the complexity of arithmetic circuits and Boolean circuits. We improve the state of the art results for several models of computing, mainly for bounded depth arithmetic and Boolean circuits. Our lower bounds are proved for some interesting functions including matrix product and the symmetric polynomials.

In particular we obtain the following results.

Depth 3 arithmetic circuits.

- We prove an $\Omega(n^2)$ lower bound on the size of depth 3 arithmetic circuits over fields of characteristic zero for the elementary symmetric polynomials of degree $\Omega(n)$. This is the first quadratic lower bound proved for a nontrivial model of an arithmetic circuit. This lower bound also shows that the construction in [4] is essentially optimal.
- We prove an $n^{2-\epsilon}$ lower bound on the size of depth 3 arithmetic circuits over fields of characteristic zero for polynomials of constant degree. No super-linear lower bound for polynomials of constant degree was known before.
- We show a depth 4 formula computing the d'th elementary symmetric polynomial in n variables of size $n2^{O(\sqrt{d})}$, and a depth 6 formula for it of size $O(nd^3 \log d)$. These new formulas together with the quadratic lower bound show a gap between depth 3 and depth 4 circuits. Both of these construction improve the construction of Ben-Or [4] for small values of d.

Bounded depth circuits with arbitrary gates.

- We show that the number of edges in any constant depth arithmetic circuit for **matrix product** (over any field) is super-linear in m^2 (where $m \times m$ is the size of each matrix). That is, the lower bound is super-linear in the number of input variables. Moreover, if the circuit is bilinear the result applies also for the case where the circuit gets for free any product of two linear functions.
- We show that the number of edges in any constant depth arithmetic circuit for the trace of the product of 3 matrices (over fields with characteristic 0) is super-linear in m^2. (Note that the trace is a **single-output** function).
- We give explicit examples for n Boolean functions $f_1, ..., f_n$, such that any constant depth **Boolean circuit with arbitrary gates** for $f_1, ..., f_n$ has a super-linear number

of edges. The lower bound is proved also for **circuits with arbitrary gates over any finite field**. The bound applies for matrix product over finite fields as well as for several other explicit functions.

We also consider the models of arithmetic bilinear circuits and arithmetic quadratic circuits. These models are the natural models to consider when dealing with the problem of matrix product. We prove the following results.

Bilinear circuits for matrix product.

- We prove that any *quadratic* circuit that computes the product of two $m \times m$ matrices over the field $GF(2)$ has at least $3m^2 - o(m^2)$ product gates. This result improves the former lower bound of Bshouty [8].
- We prove that any *bilinear* circuit for matrix product over the field $GF(p)$ must have at least $(2.5 + \frac{1.5}{p^3-1})m^2 - o(m^2)$ product gates. This result improves the former result of Bläser [5].

These are the best lower bounds known for matrix product in bilinear and quadratic circuits over finite fields.

All our results are proved using various techniques from algebra and combinatorics. In particular we consider dimension of spaces of polynomials, partial derivatives of polynomials, results from coding theory, analysis of the structure of bounded depth graphs and more.

When dealing with depth 3 circuits we consider (following [33]) the dimension of the space of all partial derivatives of the product gates (restricted to some affine subspace of the inputs). Comparing this measure to the dimension of the space of all partial derivatives that a certain function has (again, restricted to some affine subspace of the inputs) we get lower bounds.

In the case of bounded depth circuits with arbitrary gates we analyze the structure of the graph of the circuit, and show that if it does not contain too many edges then, in some sense, it is weakly connected. We then show that the graph of any bounded depth circuit computing matrix product is not weakly connected.

For bilinear circuits computing matrix product we show that any such circuit defines in a very natural way a linear code of matrices $\Gamma : \{0,1\}^{m^2} \mapsto \{0,1\}^s$. Where s is the number of product gates in the circuit. We then bound s from below using methods from coding theory.

Chapter 1

Introduction

Algebraic complexity theory investigates the complexity of computing algebraic functions using the arithmetic operations $\{+,-,\cdot\}$ or the Boolean operations $\{\vee,\ \wedge,\ \neg\}$ (over the Boolean domain). The most natural models for computing algebraic functions are the models of arithmetic circuits and Boolean circuits. The research of these two models has gone in two different directions. On the on hand researchers sought efficient algorithms (circuits) for interesting functions. Much efforts had been invested in this direction with noticeable success, and many clever algorithms were found. As examples we mention Strassen's algorithm for multiplying two $n \times n$ matrices over any field [51] and Csanky's algorithm for a parallel computation of the determinant of an $n \times n$ matrix [11]. On the other hand researchers tried, equally hard, to prove that some explicit function is hard to compute, that is to prove a lower bound on the amount of resources used by any circuit that computes the function. Unfortunately similar to most computational models almost no lower bounds were proved (although counting arguments show that almost all functions are hard to compute). The best known lower bound is the classical $\Omega(n \log d)$ (for a polynomial in n indeterminates of degree d over a large field) resulting from Strassens's degree bound [52] together with the Baur-Strassen theorem [3]. No other methods are known for proving lower bounds for general circuits. For Boolean circuits the situation is even worst - the best known lower bound is a lower bound of $4.5n$ of Lachish and Raz [24]. In addition there is no lower bound for depth. See [7, 54, 16] for a survey.

As it became clear that proving lower bounds for general circuits is a difficult task, researchers started to investigate restricted versions of arithmetic and Boolean circuits. This line of research proved to be fruitful when dealing with the monotone analogs, that is Boolean circuits without $\neg$ gates and arithmetic circuits that don't use subtractions. In 1985 Razborov [38] gave the first super polynomial lower bound for monotone Boolean circuits computing the clique function. This bound was later strengthened and today there is an exponential lower bound for the clique function due to Alon and Boppana [2]. Raz and Wigderson gave a linear lower bound for monotone Boolean circuits computing the matching function [37]. Monotone arithmetic circuits have also been investigated and exponential lower bounds are known for several functions [43, 45, 23, 59, 32]. Shamir and Snir [46] also proved a linear lower bound for the depth of monotone arithmetic circuits (see [55] as well).

Another restricted model that received a lot of attention is the *bounded depth* model. However for bounded depth circuits the known results heavily depend on the field over which the circuit is defined. For bounded depth Boolean circuits Ajtai [1] and Furst et al. [15] gave exponential lower bounds, which were later strengthened by Yao [61] and Håstad [21]. In 1987 Razborov [39] and Smolensky [50] proved exponential lower bounds for $AC(p)$ circuits (Boolean circuits with $(\mathrm{mod}\ p)$ gates for some constant prime p). No super-linear lower bounds are known for $AC(m)$ circuits, computing a single output function, where m is a composite number. Despite the success of proving exponential lower bounds for bounded depth Boolean circuits, almost no lower bounds are known for bounded depth arithmetic circuits. The only exponential lower bound is the lower bound of Grigoriev and Karpinski, Grigoriev and Razborov [18, 19] for depth 3 arithmetic circuits over finite fields. The only exponential lower bounds for bounded depth arithmetic circuits over fields of non-constant size, are the results of [33] for restricted depth 3 circuits. The best lower bound known for bounded depth arithmetic circuits over fields of characteristic zero is the slightly super-linear lower bound of Pudlak [34] and Lokam [28].

In this thesis we study bounded depth arithmetic and Boolean circuits. We prove the first quadratic lower bounds for depth 3 arithmetic circuits over fields of characteristic zero. We also prove the first super-linear lower bounds for explicit polynomial of constant degree for all kind of bounded depth circuits over any field. We investigate the complexity of computing the product of two $m \times m$ matrices in bounded depth circuits and in bilinear circuits, and improve the state of the art results.

1.1 Boolean circuits

Definition 1.1.1 *A Boolean circuit is a directed acyclic graph as follows. Nodes of in-degree 0 are called inputs and are labeled with input variables. Nodes of out-degree 0 are called outputs. All nodes other than the inputs are labeled with an OR, or AND or NOT ($\{\vee,\ \wedge,\ \neg\}$). The inputs to the circuit are Boolean variables and each node in the circuit computes in a natural way a Boolean function in the original input variables. The size of a circuit C is denoted by $size(C)$ and is defined to be the number of edges in it. The depth of a circuit is defined to be the length of the longest directed path from an input to an output in the circuit. For a polynomial f we denote by $s(f)$ the size of a smallest circuit that computes f.*

Exponential lower bounds are known for bounded depth Boolean circuits [1, 15, 61, 21]. Razborov [39] and Smolensky [50] showed that even if we allow the circuit to use $\pmod p$ gates (for some fixed prime p), then it is still exponentially hard to compute the majority function.

Another model of Boolean circuits that we consider in this work is the model of *Boolean circuits with arbitrary gates.*

Definition 1.1.2 *A Boolean circuit with arbitrary gates, is a boolean circuit in which we allow each gate (of fan-in k) to compute an arbitrary function* $g : \{0,1\}^k \to \{0,1\}$[1].

One important special case of Boolean circuits with arbitrary gates is *constant depth threshold circuits*, where we allow MAJORITY gates (i.e., the output is 1 iff more than $k/2$ inputs are 1) and negations. Constant depth threshold circuits are quite powerful. They can efficiently compute many arithmetic functions (see [29] for a survey) and they give a natural generalization of interesting circuit complexity classes. For example, the class ACC of constant depth and polynomial size circuits over the base {AND,OR,NOT,MOD M (for every constant M)} is contained in the class of quasi-polynomial size threshold circuits of depth 3 [62]. Another reason that constant depth threshold circuits are extremely interesting is that they are equivalent to neural networks of constant depth. Exponential lower bounds are known for the size of threshold circuits of depth 2 [20]. For constant depth threshold circuits only super-linear lower bounds are known [22].

1.2 Arithmetic circuits

Definition 1.2.1 *An arithmetic circuit is a directed acyclic graph as follows. Nodes of in-degree 0 are called inputs and are labeled with input variables. Nodes of out-degree 0 are called outputs. Each edge is labeled with a constant from the field and each node other than an input is labeled with one of the following operations* $\{ + , \cdot \}$ *(in the first case the node is a plus gate and in the second case a product gate). The computation is done in the following way. An input just computes the value of the variable that labels it. Then, if* $v_1, ..., v_k$ *are the vertices that fan into* v *then we multiply the result of each* v_i *with the value of the edge that connects it to* v. *If* v *is a plus gate we sum all the results, otherwise* v *is a product gate and we multiply all the results. Obviously, each node in the circuit computes a polynomial in the input variables. The size of a circuit* C *is denoted by* $size(C)$ *and is defined to be the number of edges in it. The depth of a circuit is defined to be the length of the longest directed path from an input to an output in the circuit. For a polynomial* f *we denote by* $s(f)$ *the size of a smallest circuit that computes* f.

[1]Since g_v is not necessarily a symmetric function, we need to order the inputs to v, so we know which input is the first variable of g_v etc.

Arithmetic circuits compute polynomials in the ring $F[x_1, ..., x_n]$ (where $x_1, ..., x_n$ are the input variables for the circuit), and we would like to give explicit examples for polynomials that are hard to compute. Note that for finite fields the representation of a function $f : F^n \to F$ as a polynomial is not unique (since for every i we have the equation $x_i^p = x_i$, where $p \neq 0$ is the characteristic of the field). Usually, it is only required that the circuit computes the given polynomials as functions, that is, the circuit may compute other polynomials that represent the same functions.

Lower bounds for the size of arithmetic circuits for explicit polynomials are known only if we allow polynomials with large degree or large coefficients (e.g., [52, 3]). However, if we limit the degree and the coefficients to be of size $O(1)$ then no non-trivial lower bound is known. For constant depth arithmetic circuits, exponential lower bounds are known for fields F with characteristic $p = 2$ [39, 50]. For other finite characteristics, exponential lower bounds are known only for depth 3 [18, 19]. In spite of these results for depth 3 circuits over fields of characteristic > 0, it seems that the situation in characteristic 0 is completely different. [32, 33] proved exponential lower bounds for restricted depth 3 circuits, but for general depth 3 circuits, there are no such strong results. It seems that although these circuits look very restricted they are actually quite powerful. For example, Ben-Or [4] showed quadratic depth 3 formulas that compute the elementary symmetric polynomials. This construction is in sharp contrast to the exponential lower bounds known for Boolean circuits computing the threshold functions [1, 15, 61, 21]. Thus bounded depth arithmetic circuits are more powerful than bounded depth Boolean circuits.

1.3 Matrix product

One of the functions that we study in this thesis is *matrix product*. We prove lower bounds for this function and a variant of it in several models of computation.

The problem of computing the product of two matrices is one of the most studied computational problems: We are given two $m \times m$ matrices $x = (x_{i,j})$, $y = (y_{i,j})$, and we wish to compute their product, i.e. there are m^2 outputs where the (i, j)'th output is

$$(x \cdot y)_{i,j} = \sum_{k=1}^{m} x_{i,k} \cdot y_{k,j} \ .$$

In 69' Strassen surprised the world by showing an upper bound of $O(m^{\log_2 7})$ [51]. This bound was later improved and the best upper bound today is $O(m^{2.376})$ [10] (see also [16] for a survey). The best lower bound is a lower bounds of $2.5m^2 - o(m^2)$, on the number of products needed to compute the function [8, 5]. Thus the following problem is still open: Can matrix product be computed by a circuit of size $O(m^2)$?

The standard computational model for computing polynomials is the model of arithmetic circuits. This is indeed the most general model, but for matrix product two other models are usually considered, *quadratic circuits* and *bilinear circuits*. A bilinear arithmetic circuit is an arithmetic circuit with the additional restriction that a product gate is only allowed to compute the product of two linear functions, one in the variables of X and the other in the variables of Y. Thus, bilinear circuits have the following structure. First, there are many plus gates computing linear forms in X and linear forms in Y. Then there is one level of product gates that compute bilinear forms, and finally there are many plus gates that eventually compute the outputs. The size and the depth of a bilinear circuit are defined similarly to the size and depth of a general arithmetic circuit.

The requirement that the circuit is bilinear seems restrictive. It is easy to show, however, that over fields of characteristic zero, the bilinearity assumption does not change (up to a constant factor) the size and the depth of the circuit. More accurately, by paying a constant factor in the size and in the depth, we can transform any arithmetic circuit computing a bilinear form into an equivalent bilinear circuit. Roughly speaking, this is done in the following way: Since the circuit computes polynomials of degree two, it doesn't really need to keep track of any monomial of higher degree. Therefore, we only need to keep track of monomials of degree one or two, and we can do that by replacing each gate by a constant number of new gates (at most 5 new gates) that satisfy the bilinearity assumption. Thus we have:

Proposition 1.3.1 *If a set of bilinear forms is computed by an arithmetic circuit of depth d and size s over a field of characteristic zero, then there is a bilinear circuit of depth $3d$ and size $5s$ over the same field computing the same set of bilinear forms.*

In addition we note that the best circuits that we have today for matrix product (over any field) are bilinear circuits.

The second model that we consider is the model of quadratic circuits. A quadratic circuit is an arithmetic circuit with the additional restriction that product gates are applied only on two linear functions. Notice that the only difference between quadratic circuits and bilinear circuits is that in the quadratic model the product gates compute *quadratic forms* in x, y, whereas in the bilinear model the product gates compute *bilinear forms* in x, y. This model is more general than the model of bilinear circuits, but it is still more restricted than the general model. However it is interesting to note that over infinite fields we can assume w.l.o.g. that any arithmetic circuit that computes the product of two $m \times m$ matrices is a quadratic circuit [53].

1.4 Overview of the results and methods

The results in the thesis are based on the following papers:

- Depth 3 arithmetic circuits over fields of characteristic zero, a joint work with Professor Avi Wigderson, appeared in 14th Annual IEEE conference on Computational Complexity 1999 [49], to appear in Journal of Computational Complexity.
- Affine projections of symmetric polynomials, appeared in 16th annual IEEE conference on Computational Complexity 2001 [47].
- Lower bounds for matrix product, in bounded depth circuits with arbitrary gates, a joint work with Professor Ran Raz, appeared in Thirty-Third Annual ACM Symposium on Theory of Computing (STOC) 2001 [36].
- Lower bounds for matrix product, to appear in 42nd Annual Symposium on Foundations of Computer Science (FOCS) 2001 [48] .

We now describe the main results and the techniques of these papers.

1.4.1 Depth 3 arithmetic circuits over fields of characteristic zero

We prove quadratic lower bounds for depth-3 arithmetic circuits over fields of characteristic zero. Such bounds are obtained for the elementary symmetric functions, the (trace of) iterated matrix multiplication, and the determinant. As corollaries we get the first nontrivial lower bounds for computing polynomials of constant degree, and a gap between the power of depth-3 arithmetic circuits and depth-4 arithmetic circuits. We also give new shorter formulae of constant depth for the elementary symmetric functions

The main technical contribution relates the complexity of computing a polynomial in this model to the wealth of partial derivatives it has on every affine subspace of small co-dimension. More accurately, we compare the dimension of the space of partial derivatives of the polynomial (after we restrict the polynomial to a certain subspace of the inputs) to the space of partial derivatives of the product gates of the circuit (again after the restriction). From this comparison we get that the if the circuit computes the polynomial then it has a linear number of product gates each of linear degree.

Lower bounds for related models utilize an algebraic analog of Nečhiporuk lower bound on Boolean formulae.

1.4.2 Affine projections of symmetric polynomials

We introduce a new model for computing polynomials - a depth 2 circuit with a symmetric gate at the top and plus gates at the bottom, i.e the circuit computes a symmetric function in linear functions - $S_m^d(\ell_1, \ell_2, ..., \ell_m)$ (S_m^d is the d'th elementary symmetric polynomial in m variables, and the ℓ_i's are linear functions). We refer to this model as the symmetric model. This new model is related to standard models of arithmetic circuits, especially to depth 3 circuits. In particular we show that in order to improve the results of [49], i.e to prove super-quadratic lower bounds for depth 3 circuits, one must first prove a super-linear lower bound for the symmetric model.

We prove two nontrivial linear lower bounds for our model. The first lower bound is for computing the determinant, and the second is for computing the sum of two monomials. The main technical contribution relates the maximal dimension of linear subspaces on which S_m^d vanishes, and lower bounds to the symmetric model. In particular we show that an answer of the following problem (which is very natural, and of independent interest) will imply lower bounds on symmetric circuits for many polynomials:

"What is the maximal dimension of a linear subspace of $\mathcal{C}^m$, on which S_m^d vanishes ?"

We give two partial solutions to the problem above, each enables us to prove a different lower bound. The methods that we use to answer the problem involve taking partial derivatives of polynomials, symmetrizing polynomials and the use of the basic theorem of the symmetric functions. Using our techniques we also prove quadratic lower bounds for depth 3 circuits computing the elementary symmetric polynomials of degree αn (where $0 < \alpha < 1$ is a constant), thus extending the result of [49].

1.4.3 Lower bounds for matrix product, in bounded depth circuits with arbitrary gates

We prove super-linear lower bounds for the number of edges in constant depth circuits with n inputs and up to n outputs. Our lower bounds are proved for all types of constant depth circuits, e.g., constant depth arithmetic circuits and constant depth Boolean circuits with arbitrary gates. The bounds apply for several explicit functions, and, most importantly, for matrix product. In particular, we obtain the following results:

1. We show that the number of edges in any constant depth arithmetic circuit for **matrix product** (over any field) is super-linear in m^2 (where $m \times m$ is the size of each matrix). That is, the lower bound is super-linear in the number of input variables. Moreover, if the circuit

is bilinear the result applies also for the case where the circuit gets for free any product of two linear functions.

2. We show that the number of edges in any constant depth arithmetic circuit for the trace of the product of 3 matrices (over fields with characteristic 0) is super-linear in m^2. (Note that the trace is a **single-output** function).

3. We give explicit examples for n Boolean functions $f_1, ..., f_n$, such that any constant depth **Boolean circuit with arbitrary gates** for $f_1, ..., f_n$ has a super-linear number of edges. The lower bound is proved also for **circuits with arbitrary gates over any finite field**. The bound applies for matrix product over finite fields as well as for several other explicit functions.

The idea behind these results is the following. We show that any bounded depth graph with a "small" number of edges is in some sense weakly connected. We then consider bounded depth circuits for matrix product, and we show that if we look at the graph of the circuit then this graph is not weakly connected. Therefor the circuit for matrix product contains a large number of edges.

1.4.4 Lower bounds for matrix product, in bilinear and quadratic circuits

We prove lower bounds on the number of product gates in bilinear and quadratic circuits that compute the product of two $m \times m$ matrices over finite fields. In particular we obtain the following results:

1. We show that the number of product gates in any *bilinear* (or *quadratic*) circuit that computes the product of two $m \times m$ matrices over $GF(2)$ is at least $3m^2 - o(m^2)$.

2. We show that the number of product gates in any *bilinear* circuit that computes the product of two $m \times m$ matrices over $GF(p)$ is at least $(2.5 + \frac{1.5}{p^3-1})m^2 - o(m^2)$.

These results improve the former results of Bshouty and Bläser [8, 5] who proved lower bounds of $2.5m^2 - o(m^2)$.

The technique in which we prove these results is the following. We show that any bilinear circuit that computes matrix product defines in a very natural way a *linear code of matrices* $\Gamma : \{0,1\}^{m^2} \mapsto \{0,1\}^s$, where s is the number of product gates in the circuit. Then using methods from coding theory we show that s must be large.

Chapter 2

Depth-3 Arithmetic Circuits over Fields of Characteristic Zero

2.1 Introduction

In this chapter we deal mainly with depth-3 arithmetic circuits, better known as $\Sigma\Pi\Sigma$ circuits. A $\Sigma\Pi\Sigma$ circuit can be viewed as a sum of products of linear functions of the variables. This is a very restricted model, but it can clearly compute any multivariate polynomial, some of which surprisingly cheaply. Being the "simplest" nontrivial model, it has received significant attention as detailed below.

Despite its innocence, no super-linear lower bounds are known for this model when the field is large, except of the degree lower bound of [3]. This state of affairs is in contrast with what is known for Boolean circuits with $mod\, q$ gates, and arithmetic circuits over finite fields. In the first model [39] and [50] proved exponential lower bounds e.g. for the majority function for any constant depth. In the second model (for depth-3 only) [18] and [19] proved exponential lower bounds for some symmetric functions.

These techniques cannot be extended to give analogous results for large fields, as by the above mentioned result of Ben-Or [4], they are simply false. One path which was taken to handle large fields was to further restrict the model [32, 33]. They consider homogeneous circuits, which for depth-3 circuits amounts to requiring a homogeneous linear function at every gate in the bottom level. In that model Nisan and Wigderson [33] were able to prove exponential lower bounds, even for the majority polynomial. Thus, the general model is sometimes exponentially more powerful than its homogeneous variant.

We are interested in the general model of $\Sigma\Pi\Sigma$ circuits. We prove near quadratic lower bounds for a number of natural functions, such as the elementary symmetric functions, the (trace of) iterated matrix product, and the determinant. We show in particular the following lower bounds (where n denotes the total number of variables):

- $\Omega(n^2)$ lower bound for the elementary symmetric functions of degree $\Omega(n)$ (Thus proving that the construction in [4] is essentially optimal).
- $n^{2-\epsilon}$ lower bound for polynomials of constant degree (No super-linear bound was known before).
- $\tilde{\Omega}(n^2)$ lower bound for the determinant (But we believe that the true complexity is exponential).

The proof combines the idea of partial derivatives of [33] and the idea of approximating "high rank" multiplication gates of [18]. Some of these lower bounds, (together with upper bounds) provide near quadratic separation between depth-3 and depth-4 circuits. Of special interest is a new (depth-6) formula of size $O(nd^3 \log d)$ for the elementary symmetric polynomials of degree d (which beats the Ben-Or construction for small degrees).

We prove two general theorems that express a lower bound for the depth-3 complexity of a polynomial f, in terms of parameters k and D, where D measures the wealth of partial derivatives f has on every affine subspace of co-dimension k. These are stated and proved in section 2.3. Specific lower bounds follow from proving lower bounds on D for specific polynomials and well chosen values of k. These are stated and proved in section 2.4.

In section 2.5 we prove new upper bounds for the elementary symmetric polynomials.

In section 2.6 we look at different (but related) models of computation. The first model is formulae without any depth restriction, whose inputs are allowed to be not only the variables, but also arbitrary univariate polynomials. For this model we prove a quadratic lower bound for the Discriminant function (which is the determinant of a Vandermonde matrix). The proof uses an algebraic analog of a Nečhiporuk-like argument (see [31]).

Finally, we study the computation of polynomials of the form $t^m f$, where f is our target (homogeneous) polynomial, and t is a new indeterminate. We show general computation of such products of f is no stronger than *homogeneous* computation of f. This result has the same flavor of results on the monotone vs general computation of "slice functions". It yields an exponential gap between the depth-3 homogeneous complexity of a polynomial and the homogeneous complexity of its derivative with respect to a single variable.

2.2 Definitions and Tools

2.2.1 $\Sigma\Pi\Sigma$ Circuits

The main model we shall deal with in this chapter is the following.

Definition 2.2.1 *A $\Sigma\Pi\Sigma$ circuit is a leveled depth-3 circuit with a plus gate at the top, multiplication gates at the middle level, and plus gates at the bottom. A homogeneous $\Sigma\Pi\Sigma$ circuit is a $\Sigma\Pi\Sigma$ circuit that is allowed to compute only homogeneous linear functions in the bottom level.*
For a polynomial f, we denote by $s_{\Sigma\Pi\Sigma}(f)$ (respectively $L_{\Sigma\Pi\Sigma}(f)$) the size of the smallest $\Sigma\Pi\Sigma$ circuit (respectively formula) computing f, and by $s^H_{\Sigma\Pi\Sigma}(f)$ (respectively $L^H_{\Sigma\Pi\Sigma}(f)$) the size of the smallest homogeneous $\Sigma\Pi\Sigma$ circuit (respectively formula) computing f.

Comment: Clearly, for all variants, $s \leq L$. All our lower bounds will be on s, and all our upper bounds on L.

From the definition it is clear that if f is computed by such circuit then f has a representation of the form: $f = \sum_{j=1}^{s} M_j$ where $M_j = \prod_{i=1}^{deg(M_j)} \ell_{i,j}$, each $\ell_{i,j}$ is a linear function in the variables, and $deg(M_j)$ is the fan-in of the jth multiplication gate. We stress that this linear function may involve a constant term (and indeed without this ability the model is homogeneous), and that different multiplication gates may use the same linear function. It will be useful for us to separate out the homogeneous part of such functions.

Definition 2.2.2 *For a linear function ℓ we let ℓ^h be the homogeneous part of ℓ, and $\ell^0 = \ell - \ell^h$ is the constant term.*
Let $M = \prod_{i=1}^{deg(M)} \ell_i$. We denote $M^h = \{\ell_1^h, ..., \ell_{deg(M)}^h\}$ and let $\dim(M^h)$ be the dimension of the linear span of the set M^h.

All our lower bounds for this model will trade off the number of edges that fan into gates at the middle level (multiplication gates) with the number of gates in that level. We will use the obvious

Proposition 2.2.3 *Every $\Sigma\Pi\Sigma$ circuit with multiplication gates $\{M_1, ..., M_s\}$ has size*

$$\Omega\left(\sum_{j=1}^{s} deg(M_j)\right).$$

2.2.2 Partial Derivatives

We now recall basic definitions and results from [33] on partial derivatives, which will be key for our lower bounds.

Let F be a field of characteristic zero. We will consider polynomials over a set of variables X.

Definition 2.2.4 *For any set of polynomials $V \subseteq F[X]$ we use $dim(V)$ for the dimension of the linear span of V (in other words the maximum number of linearly independent polynomials in V over F).*

Remark: Observe that $dim(V)$ is invariant under any full rank linear transformation on the set of variables X.

Definition 2.2.5 *[33] Let f be a polynomial and d an integer. We let $\partial_d(f)$ denote the set of partial derivatives of order d of f.*

Example 1

$$\partial_2(x^2y) = \{0, 2x, 2y\} \ .$$

A fundamental observation of [33] is that this dimension commutes with the arithmetic operations, i.e it is invariant under scalar multiplication and is sub-additive. Here we'll use it only for addition gates.

Proposition 2.2.6 *[33] For every $f_1, f_2, \cdots, f_r \in F[X]$ and $\alpha \in F$, $\alpha \neq 0$ we have:*

- $dim(\partial_d(\alpha f_1)) = dim(\partial_d(f_1))$.
- $dim(\partial_d(\sum_i f_i)) \leq \sum_i dim(\partial_d(f_i))$.

For multiplication gates, we'll use stronger bounds than those in [33], which apply only to product of linear functions.

Proposition 2.2.7 *For a multiplication gate M with $\dim(M^h) = m$, and for every d, $dim(\partial_d(M)) \leq \binom{m+d}{d}$.*

Proof: From the definition it is clear that M is a function of m linearly independent linear functions (w.l.o.g. - the first m), i.e $M = M(\ell_1, ..., \ell_m)$. Therefore $\frac{\partial M}{\partial x} = \sum_{i=1}^{m} \frac{\partial M}{\partial \ell_i} \frac{\partial \ell_i}{\partial x}$. Since ℓ_i is a linear function, $\frac{\partial \ell_i}{\partial x}$ is a scalar, so $\frac{\partial M}{\partial x}$ is a linear combination of the $\frac{\partial M}{\partial \ell_i}$'s. A similar thing happens when we look at derivatives of order d, each derivatives lies in the linear span of order d derivatives of M with respect to the ℓ_i-s. Therefore to bound $dim(\partial_d(M))$ it is sufficient to bound the number of such derivatives. Since M is a polynomial the order in which we take derivatives doesn't matter, i.e $\frac{\partial^2 M}{\partial \ell_1 \partial \ell_2} = \frac{\partial^2 M}{\partial \ell_2 \partial \ell_1}$. Therefore the number of order d partial derivatives is at most the number of ways to write d as a sum of m integers (the value of the i-th integer corresponds to the order we take derivatives of M w.r.t. ℓ_i). $\square$

Proposition 2.2.8 *For a multiplication gate M with $deg(M) = m$, and for every d, $dim(\partial_d(M)) \leq \binom{m}{d}$.*

Proof: Write $M = \prod_{i=1}^{m} \ell_i$. Denote $[m] = \{1, ..., m\}$. Since each ℓ_i is a linear function we get,

$$\partial_d(M) \subset \ span\{\prod_{i \in T} \ell_i | T \subset [m], |T| = m - d\}.$$

The result follows since $\dim(span\ \{\ \prod_{i \in T} \ell_i \mid T \subset [m], |T| = m - d\ \}) \leq \binom{m}{d}$. $\square$

2.2.3 Restrictions to affine subspaces

A key ingredient of our lower bound technique will be to study the dimension of a set of partial derivatives not over the whole vector space F^n, but over affine subspaces of it.

Definition 2.2.9 *Let A be an affine subspace of F^n with the set of coordinates $(x_1, ..., x_n)$. Call $B \subset X$ a base for A if A can be represented by the set of equations:*

$$\{x_b = \ell_b \mid b \in B\}$$

where ℓ_b is a linear function on the set of variables $X - B$. Note that every A has such a base B, so lets fix one such base.
Define $\phi_B : F[X] \mapsto F[X \setminus B]$ to be the homomorphism which assigns to every variable x_b with $b \in B$ the linear function ℓ_b (and leaves the other variables untouched). This map is extended by multiplicativity to monomials and then by additivity to polynomials.

Definition 2.2.10 *Let A be an affine subspace of F^n, for a polynomial f we denote by $f|_A^{(B)}$ the restriction $\phi_B(f)$ o f f to A. For a set of polynomials V, $V|_A^{(B)} = \{f|_A^{(B)} : f \in V\}$.*

Definition 2.2.11 *Let C be an arithmetic circuit, and ϕ_B as above. Then $C|_A^{(B)}$ is the circuit obtained by applying ϕ_B to the inputs, and removing sub-circuits whose output becomes identically zero.*

Clearly, restrictions commute with arithmetic operations. Thus

Proposition 2.2.12 *Assume C computes f. Then for every affine subspace A (chosen together with a base B), $C|_A^{(B)}$ computes $f|_A^{(B)}$.*

The complexity of computing a polynomial f will be related in the next section to the dimension of one of two different sets of polynomials derived from f.

1. $(\partial_d(f))|_A^{(B)}$ is the set of restrictions of all order d partial derivatives on f to A.
2. $\partial_d(f|_A^{(B)})$ is the set defined by first restricting f itself to A, and then taking all order d partial derivatives. To formally define it, we need an "inverse" to ϕ_B below.

Definition 2.2.13 *Let's assume that $X - B = \{x_{i_1}, ..., x_{i_{n-k}}\}$. Define $\ell_B : F^{n-k} \mapsto F^n$ by*

$$\ell_{B(i)} = \begin{cases} l_b & i = b \in B \\ x_i & i \notin B \end{cases}$$

where $\ell_{B(i)}$ is the i-th coordinate of ℓ_B.

Definition 2.2.14 *Let $\partial_d(f|_A^{(B)}) = \partial_d(f \circ \ell_B)$.*

We conclude this subsection with a remark on the arbitrariness of the choice of the base B of the affine subspace A. Clearly, this choice affect the sets defined above, as well as the restricted circuit, at least in the obvious sense that the restricted polynomials will be over different sets of indeterminates.

Still, the choice of B has no affect on the proofs, so from this point of view it can certainly be arbitrary. Moreover, it turns out that while the sets of polynomials defined by (1) and (2) above change with different choices of B, the dimension of each is invariant under this choice.

Proposition 2.2.15 *For every B, B' there's a full rank linear transformation $P : F^{n-k} \mapsto F^{n-k}$ such that $\ell_B = \ell_{B'} \circ P$.*

Claim 2.2.16 *For every B, B' bases for A, $dim(\partial_d(f|_A^{(B)})) = dim(\partial_d(f|_A^{(B')}))$.*

Proof: We show the proof for the case $d = 1$ but for larger d-s it's the same. According to Proposition 2.2.15 there's a full rank linear transformation P such that $\ell_{B'} = \ell_B \circ P$. For convenience let $P(X_1, ..., X_{n-k}) = (Y_1, ..., Y_{n-k})$, we get:

$$\frac{\partial}{\partial X_i}(f \circ \ell_{B'}) = \frac{\partial}{\partial X_i}(f \circ \ell_B \circ P) = \sum_{j=1}^{n-k}((\frac{\partial}{\partial Y_j}(f \circ \ell_B)) \circ P)\frac{\partial Y_j}{\partial X_i} \ .$$

Since $\frac{\partial Y_j}{\partial X_i}$ is a scalar in F we get that

$$\frac{\partial}{\partial X_i}(f \circ \ell_{B'}) \in span((\frac{\partial}{\partial Y_j}(f \circ \ell_B)) \circ P) \ .$$

Therefore $dim(\partial_1^{B'}(f)) \leq dim(\partial_1^{B}(f))$. From symmetry (since P has full rank) the result follows. $\square$

From now on we shall assume that every affine subspace A comes together with a base B. For short we define:

Definition 2.2.17 $\partial_d(f|_A) = \partial_d(f|_A^{(B)})$.

2.3 Main Theorems

The following theorems will be our main tool in deriving the lower bounds. Observe that in both theorems, the lower bound is the minimum of two functions. It will be clear from their proofs that they both actually give a trade-off: either the first function bounds from below the number of edges that fan into the middle level (multiplication gates), or the second lower bounds the number of gates in that level.

Theorem 2.3.1 *Let f be a polynomial. Assume that for some integers d, k, D, for every affine subspace A of co-dimension k, $dim((\partial_d(f))|_A) > D$. Then*

$$s_{\Sigma\Pi\Sigma}(f) \geq \min(\frac{k^2}{d}, \frac{D}{\binom{k+d}{d}}) \ .$$

Theorem 2.3.2 *Let f be a polynomial. Assume that for some integers d, k, D, for every affine subspace A of co-dimension k, $dim(\partial_d(f|_A)) > D$. Then for every m*

$$s_{\Sigma\Pi\Sigma}(f) \geq \min(km, \frac{D}{\binom{m}{d}}) \ .$$

The way to apply these theorems is to prove, that for certain polynomials, D can be chosen large, for appropriate values of d, k. We note that the theorems are incomparable – each can give a better lower bound than the other for some polynomials.

To prove these theorems we will first need a technical lemma. It basically shows how to define an affine subspace that nullifies high rank multiplication gates.

Lemma 2.3.3 *Fix an integer z. Let $\mathbb{C}$ be a $\Sigma\Pi\Sigma$ circuit, and let $S = \{j \mid \dim(M_j^h) \geq k\}$. If $|S| < \frac{k}{z}$ then there is an affine subspace A of co-dimension k such that in $\mathbb{C}|_A$ all the multiplication gates are of dimension less then k. Moreover, in A all the multiplication gates in S have a zero of order $\geq z$.*

Proof: Let's assume w.l.o.g. that $S = \{1, 2, ..., r\}$ for some $0 \leq r < \frac{k}{z}$. Now lets take for each $1 \leq j \leq r$, z linear functions, $\ell_{j,1}, ..., \ell_{j,z}$, from M_j, such that all the $\ell_{j,i}^h$'s are linearly independent. It's possible to make such a choice since $r < \frac{k}{z}$ and $\dim(M_j^h) \geq k$. Because of the independence there's an affine subspace A of co-dimension at most rz such that

$$\forall \vec{x} \in A, \quad \ell_{j,i}^h(\vec{x}) = -\ell_{j,i}^0 \quad \text{(i.e } \ell_{j,i}(\vec{x}) = 0 \text{ for the relevant } (j,i)\text{-s)}.$$

So now in $\mathbb{C}|_A$ all the multiplication gates are of dimension less than k, and $M_1, ..., M_r$ each have a zero of order $\geq z$ on A. □

We can now turn to the proofs of the main theorems.

Proof of Theorem 2.3.1: Fix any k and d, and set $z = d + 1$. Assume that $\mathcal{C}$ is a depth-3 circuit computing f. If the assumption of Lemma 2.3.3 does not hold, then by Proposition 2.2.3 we have a $\frac{k^2}{d}$ lower bound on the number of edges that fan into the multiplication gates. Otherwise the assumption holds, and let A be the subspace guaranteed by the lemma.

Now we know that all the gates in $\mathcal{C}$ with dimension $\geq k$ have a zero of degree $d + 1$ when restricted to A. Thus for each such multiplication gate M, $dim(\partial_d(M)|_A) = 0$. By Proposition 2.2.7 all the other multiplication gates have $dim(\partial_d(M)|_A) \leq \binom{k+d}{d}$ so by Proposition 2.2.6 there must be at least $\frac{D}{\binom{k+d}{d}}$ multiplication gates in $\mathcal{F}$. □

Proof of Theorem 2.3.2: Assume that we are given d, k, m. We now look at an optimal $\Sigma\Pi\Sigma$ circuit for f. If there are more than k multiplication gates of degree greater than m then we are done. Otherwise we can find an affine subspace A of co-dimension k so if we restrict the circuit to A all the remaining gates will be of degree $\leq m$. Using Propositions 2.2.6, 2.2.8 we see that there are at least $\frac{D}{\binom{m}{d}}$ multiplication gates in the restricted circuit. □

It is clear from our theorems that the lower bounds for a polynomial f are expressed in terms of the dimensions of the sets $\partial_d(f|_A)$ and $(\partial_d f)|_A$, for some arbitrary affine subspace A. These sets can be quite complicated even for very simple polynomials. To bound the dimensions, we establish two simple lemmas relating these to each other and to the more easily understood set, $\partial_d(f)$.

Lemma 2.3.4 *For every polynomial f and affine subspace A with a base B we have*

$$dim((\partial_d f)|_A) \geq dim(\partial_d(f) \cap F[X \setminus B]) \ .$$

Proof: The polynomials in $\partial_d(f)$ which do not depend on variable from B are not effected by the restriction to A. $\square$

Definition 2.3.5 *For a set of variables B, let $\partial_d^B f$ denote all order d derivatives of f, where at least one of the d derivations are with respect to a variable in B.*

Lemma 2.3.6 *For every polynomial f and affine subspace A with a base B we have*

$$dim(\partial_d(f|_A)) \geq dim((\partial_d f)|_A) - dim(\partial_d^B f) \ .$$

The proof actually shows that

$$(\partial_d f)|_A \subset \mathrm{span}\left(\partial_\mathrm{d}(\mathrm{f}|_\mathrm{A}) \cup (\partial_\mathrm{d}^\mathrm{B}(\mathrm{f}))|_\mathrm{A}\right) \ .$$

Proof: We will demonstrate the proof for $|B| = 1$ and $d = 1$. w.l.o.g. assume that $B = \{x_1\}$ and A is given by $x_1 = \ell(x_2, ..., x_n)$. According to the chain rule we get:

$$\partial_1(f|_A) = \{\frac{\partial f(\ell, x_2, ..., x_n)}{\partial x_i}|1 < i\} = \{\frac{\partial f}{\partial x_1}(\ell, x_2, ..., x_n) \cdot \frac{\partial \ell}{\partial x_i} + \frac{\partial f}{\partial x_i}(\ell, x_2, ..., x_n) \mid 1 < i\} \ .$$

Since $\partial_1^B f = \frac{\partial f}{\partial x_1}$, the set $(\partial_1 f)|_A$ (which is actually the set $\{\frac{\partial f}{\partial x_i}(\ell, x_2, ..., x_n)|1 \leq i\}$) is spanned by the set $\partial_1(f|_A) \cup \{\frac{\partial f}{\partial x_1}(\ell, x_2, ..., x_n)\}$.

The proof for larger values of $|B|, d$ is achieved using similar arguments. View the restriction to A as a composition of a polynomial with a set of linear functions. Then from the chain rule of partial derivatives, and from the fact that all derivatives of linear functions are constants, it follows that the set $(\partial_d f)|_A$ is spanned by the two sets $\partial_d(f|_A)$ and $\partial_d^B f$. $\square$

2.4 Lower Bounds

2.4.1 Elementary symmetric functions

Definition 2.4.1

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

X is the set of variables $\{x_1, ..., x_n\}$. This is the d'th elementary symmetric function.

Theorem 2.4.2 *For every* $\log n \le d \le 2n/3$ *we have* $s_{\Sigma\Pi\Sigma}(S_n^d) \ge \max(\Omega(\frac{n^2}{d}), \Omega(nd))$.

This lower bound is tight (for large values of d) in view of the following theorem (which will be proved in section 5):

Theorem 2.4.3 (Ben-Or) *For every* d, $s_{\Sigma\Pi\Sigma}(S_n^d) \le O(n^2)$.

To prove Theorem 2.4.2 we first need to prove the following theorem:

Theorem 2.4.4 *For every* $\log n \le d \le n/10$, $s_{\Sigma\Pi\Sigma}(S_n^{2d}) \ge \Omega(\frac{n^2}{d})$.

The proof is based on a Lemma showing that S_n^{2d} has lots of partial derivatives even when restricted to affine subspaces of small co-dimension.

Lemma 2.4.5 *For every* n, k, d *and for every affine subspace* A *of co-dimension* k,

$$dim((\partial_d(S_n^{2d}))|_A) \ge \binom{n-k}{d} .$$

Proof of Lemma 2.4.5: w.l.o.g. assume that A is defined by the following equations:

$$\forall 1 \le j \le k \ \ x_j = \ell_j \ .$$

The set $\partial_d(S_n^{2d})$ is actually the set $\{S_{n-d}^d(X - \{x_i | i \in T\})\}$, which for convenience we denote by $\{S_{n-d}^d(X - T)\}$ (recall Definition 2.4.1) where T ranges over the $\binom{n}{d}$ subsets of $[n]$ of size d. By [17] (as used in [33]), this set is spanned by all degree d multilinear monomials. There are $\binom{n-k}{d}$ linear independent degree d multilinear monomials in the variables of the set $X - B$, using Lemma 2.3.4 the result follows. □

Proof of Theorem 2.4.4: If we plug Lemma 2.4.5 into Theorem 2.3.1, we get a lower bound of

$$s_{\Sigma\Pi\Sigma}(S_n^{2d}) \ge \min(\frac{k^2}{d}, \frac{\binom{n-k}{d}}{\binom{k+d}{d}}) \ \ \forall k, d \ .$$

Let's estimate $\frac{\binom{n-k}{d}}{\binom{k+d}{d}}$ by $(\frac{n-k}{k+d})^d$ and take $k = n/9$, $n/10 \ge d \ge \log n$. We now get $(\frac{n-k}{k+d})^d \ge (\frac{8n/9}{2n/9})^{\log n} = n^2$. Therefore $s_{\Sigma\Pi\Sigma}(S_n^{2d}) \ge \min(\Omega(\frac{n^2}{d}), n^2) = \Omega(\frac{n^2}{d})$ and the lower bound follows. □

This theorem already gives us the desired result (of Theorem 2.4.2) for $\log n \le d \le \sqrt{n}$. The next lemma together with Theorem 2.3.2 gives the rest of the Theorem, for the right choice of parameters.

Lemma 2.4.6 *For every n, k, d, and for every subspace A of co-dimension k with $k \le \min(d/2, n-d)$*

$$dim(\partial_{\frac{d-2k}{2}}(S_n^d|_A)) \ge \binom{n-2k}{\frac{d-2k}{2}}.$$

We will derive this lemma from the following one:

Lemma 2.4.7 *For every affine subspace A of co-dimension $\le \min(d/2, n-d)$ we have $(S_n^d)|_A \ne 0$.*

To prove this we need some new notations:

Definition 2.4.8 *For a linear function $\ell = c_1x_1 + c_2x_2 + ... + c_nx_n + c_0,\ \ \ell \ne 0$ denote:*

$$LM(\ell) = \begin{cases} c_jx_j & \text{for the first } 0 < \mathrm{j} \text{ with } \mathrm{c_j} \ne 0 \\ c_0 & \text{no such j exists} \end{cases}$$

$$LT(\ell) = \begin{cases} x_j & \text{for the first } 0 < \mathrm{j} \text{ with } \mathrm{c_j} \ne 0 \\ 1 & \text{no such j exists} \end{cases}$$

Proof of Lemma 2.4.7: Let co-dimension $(A) = k$. w.l.o.g. assume that A is defined by the following linear equations:

$$x_i = \begin{cases} 0 & 1 \le i \le r \\ \ell_i(x_{d+1}, ..., x_n) & r+1 \le i \le k \end{cases}$$

where the ℓ_i are nonzero linear functions. The first r equations reduce $S_n^d(X)$ to the function $S_{n-r}^d(X \setminus [r])$. So assume w.l.o.g. that $r = 0$. Denote $Y = \{LT(\ell_i) \mid LT(\ell_i) \ne 1\ ,\ 1 \le i \le k\}$, $|Y| = t$. We get that $(S_n^d)|_A$ contains all terms of the form $(\prod_{i=1}^k (\ell_i))S_{n-k}^{d-k}(X \setminus [k])$. In particular it contains all the monomials of the polynomial

$$(\prod_{i=1}^k LM(\ell_i))(\prod_{x_j \in Y} x_j)S_{n-(k+t)}^{d-(k+t)}(X \setminus ([k] \cup Y)).$$

But it is also easy to see that every monomial $\mathcal{M}$ of degree $d - (k - t)$ appearing in $(S_n^d)|_A$, such that $(\prod_{i=1}^k LT(\ell_i))(\prod_{x_j \in Y} x_j)$ divides $\mathcal{M}$, comes from $(\prod_{i=1}^k (\ell_i))S_{n-k}^{d-k}(X \setminus [k])$. Therefore nothing cancels

$$(\prod_{i=1}^k LM(\ell_i))(\prod_{x_j \in Y} x_j)S_{n-(k+t)}^{d-(k+t)}(X \setminus ([k] \cup Y))\ .$$

and we get $(S_n^d)|_A \ne 0$. $\square$

Proof of Lemma 2.4.6: Take all derivatives of order $(d-(k+t))/2$ (same t as in the above proof) with respect to the variables from $X \setminus ([k] \cup Y)$. The set of all derivatives contains polynomials of the form

$$\left\{ (\prod_{i=1}^{k} LT(\ell_i))(\prod_{x_j \in Y} x_j) S_{n-(k+d+t)/2}^{(d-(k+t))/2}(X \setminus ([k] \cup Y \cup R)) + Q_R \right\} \textit{ such that}$$

$$R \subset X \setminus ([k] \cup Y), \mid R \mid = (d-(k+t))/2 ,$$

where Q_R is a polynomial in which no monomial is divisible by

$$(\prod_{i=1}^{k} LT(\ell_i))(\prod_{x_j \in Y} x_j) .$$

These derivatives are linearly independent, so the dimension of their span is $\binom{n-(k+t)}{(d-(k+t))/2}$ and it is minimized when $t = k$. $\square$

Proof of Theorem 2.4.2: For $\log n \leq d \leq \sqrt{n}$ we can use Theorem 2.4.4. So let's assume that $\sqrt{n} \leq d \leq 2n/3$. Now plug Lemma 2.4.6 into Theorem 2.3.2 with $k = d/4$, $m = n/100$ to get:

$$s_{\Sigma\Pi\Sigma}(S_n^d) \geq \min(dn/100, \frac{\binom{n-d/2}{\frac{d}{4}}}{\binom{n/100}{\frac{d}{4}}}) \geq \min(dn/100, \frac{\binom{2n/3}{\frac{d}{4}}}{\binom{n/100}{\frac{d}{4}}}) \geq \min(dn/100, 2^d) \geq \Omega(nd). \quad \square$$

2.4.2 Product of inner products

Definition 2.4.9

$$PIP_n^d(X,Y) = \prod_{i=1}^{d} (\sum_{j=1}^{n} x_j^{(i)} y_j^{(i)}) .$$

This is the product of d inner products. $X = \cup_{i=1}^{d} X^{(i)}$ and $Y = \cup_{i=1}^{d} Y^{(i)}$ with each $X^{(i)}$ and $Y^{(i)}$ containing n variables.

We have $2nd$ variables, and the lower bound that we show is strongest when $d \leq \sqrt{\log n}$.

Theorem 2.4.10 *For $d \leq \frac{1}{2}\sqrt{\log n}$, $s_{\Sigma\Pi\Sigma}(PIP_n^d) \geq \Omega(n^{2\frac{d}{d+2}})$.*
For $d > \frac{1}{2}\sqrt{\log n}$, $s_{\Sigma\Pi\Sigma}(PIP_n^d) \geq \Omega(n^{2-\frac{4}{\sqrt{\log n}}})$.

The proof, again, is based on a lemma that gives a lower bound for the wealth of partial derivatives that PIP has on affine subspaces.

Lemma 2.4.11 *For any affine subspace A of co-dimension k, $dim(\partial_d(PIP_n^d|_A)) \geq n^d - 2^{2d+1}kn^{d-1}$.*

Proof: Let B be any basis for A, $|B| \leq k$. Consider only order d derivatives of PIP_n^d with respect to x variables, one from each of the sets $X^{(i)}$. The set of the derivatives is the set of all monomials of degree d containing exactly one variable from each of the sets $Y^{(i)}$.

Since every y-variable appears in exactly n^{d-1} of these monomials, we have $dim(\partial_d PIP_n^d) \cap F[X \setminus B]) \geq n^d - kn^{d-1}$.

Similarly, since every monomial has at most 2^{2d} partial derivatives of order d, we also get $dim(\partial_d^B PIP_n^d) \leq 2^{2d} n^{d-1}$.

Combining these with Lemma 2.3.4 and Lemma 2.3.6 we get the bound. □

Proof of Theorem 2.4.10: It suffices to prove the bound for $d \leq \frac{1}{2}\sqrt{\log n}$, since for higher d the bound follows by a simple restriction to $PIP^n_{\frac{1}{2}\sqrt{\log n}}$. Use Lemmas 2.4.11 and Theorem 2.3.2 to get

$$s_{\Sigma\Pi\Sigma}(PIP_n^d) \geq \min(k^2, \frac{n^d - 2^{2d+1}kn^{d-1}}{\binom{k}{d}}) \ .$$

If we take $k = n^{\frac{d}{d+2}}$ we get

$$s_{\Sigma\Pi\Sigma}(PIP_n^d) \geq \min(n^{2\frac{d}{d+2}}, \frac{n^d - 2^{2d+1}n^{d-\frac{2}{d+2}}}{\binom{n^{\frac{d}{d+2}}}{d}}) \ .$$

The way to evaluate the second term inside the min is:

$$\frac{n^d - 2^{2d+1}n^{d-\frac{2}{d+2}}}{\binom{n^{\frac{d}{d+2}}}{d}} \geq \frac{n^d - 2^{2d+1}n^{d-\frac{2}{d+2}}}{n^{\frac{d^2}{d+2}}} \geq \frac{1}{6}(n^{2\frac{d}{d+2}})$$

for $d < (1/2)\sqrt{\log n}$. □

Remark: We showed a lower bound for $\Sigma\Pi\Sigma$ circuits calculating $PIP_n^d(X,Y)$, but there is a trivial $\Pi\Sigma\Pi$ circuit for it of linear size. This is an unusual example where $\Pi\Sigma\Pi$ circuit is more efficient then $\Sigma\Pi\Sigma$ circuit, and provides a (near) quadratic gap between the two.

Another separation we can deduce is the following separation between depth-4 circuits and depth-3 circuits:

Theorem 2.4.12

$$\forall d, \ L_4(PIP_n^d(X,Y) + PIP_n^d(W,Z)) = O(nd) \ ,$$

where $L_4(f)$ is the size of the smallest depth 4 formulae ($\Sigma\Pi\Sigma\Pi$ or $\Pi\Sigma\Pi\Sigma$) for f.

$$for \ \ d \leq \frac{1}{2}\sqrt{\log n}, \ \ s_{\Sigma\Pi\Sigma}(PIP_n^d(X,Y) + PIP_n^d(W,Z)) \geq \Omega(n^{2\frac{d}{d+2}}) \ .$$

2.4.3 (Trace of) Iterated matrix multiplication

Another function which has a similar lower bound, with a similar proof technique is the trace of product of matrices.

Definition 2.4.13

$$TR_n^{2d}(X^{(1)}, ..., X^{(2d)}) = \sum_{i_1,...,i_{2d}} x^{(1)}_{i_1,i_2} \cdot x^{(2)}_{i_2,i_3} \cdots x^{(2d)}_{i_{2d},i_1} .$$

This is the trace of the product of $2d$ $n \times n$ *matrices (each* $X^{(i)}$ *is an* $n \times n$ *matrix).*

Here we have $2dn^2$ variables and the lower bound depends on d and n, as in the previous example.

Theorem 2.4.14 *For* $d \leq \frac{1}{2}\sqrt{\log n}$, $s_{\Sigma\Pi\Sigma}(TR_n^{2d}) \geq \Omega(n^{4\frac{d}{d+2}})$.
For $d > \frac{1}{2}\sqrt{\log n}$, $s_{\Sigma\Pi\Sigma}(TR_n^{2d}) \geq \Omega(n^{4-\frac{8}{\sqrt{\log n}}})$.

The proof will go in the usual manner:

Lemma 2.4.15 *For any affine subspace* A *of co-dimension* k, $dim(\partial_d(TR_n^{2d}|_A)) \geq n^{2d} - 2^{2d+1}kn^{2d-2}$.

Proof: The proof is very similar to that of Lemma 2.4.11. Let B be any basis for A, $|B| \leq k$. Consider only order d derivatives of of TR_n^{2d} with respect to "even" sets $X^{(2i)}$ (one variable from each). These derivatives form the set of all monomials in the "odd" sets $X^{(2i-1)}$ (one variable from each).

Since every variable appears in exactly n^{2d-2} of these monomials, we have

$$dim(\partial_d TR_n^{2d}) \cap F[X \setminus B]) \geq n^{2d} - kn^{2d-2} .$$

Again every monomial has at most 2^{2d} partial derivatives of order d thus $dim(\partial_d^B TR_n^{2d}) \leq 2^{2d}n^{2d-2}$.

Combining these with Lemma 2.3.4 and Lemma 2.3.6 we get the bound

$$dim(\partial_d(TR_n^{2d}|_A)) \geq (n^{2d} - kn^{2d-2}) - n^{2d-2}2^{2d} \geq n^{2d} - kn^{2d-2}2^{2d+1} .$$

$\square$

Proof of Theorem 2.4.14: Again, it suffices to prove the bound for $d \leq \frac{1}{2}\sqrt{\log n}$. We use Lemmas 2.4.15 and Theorem 2.3.2 to get

$$s_{\Sigma\Pi\Sigma}(TR_n^{2d}) \geq \min(k^2, \frac{n^{2d} - 2^{2d+1}kn^{2d-2}}{\binom{k}{d}}) \ .$$

Using similar estimates (and replacing n with n^2) in the proof of Theorem 2.4.10 we get

$$s_{\Sigma\Pi\Sigma}(TR_n^d) \geq \min(n^{4\frac{d}{d+2}}, \Omega(n^{4\frac{d}{d+2}})) \ ,$$

where k equals $n^{2\frac{d}{d+2}}$. □

Although we can only prove near quadratic lower bounds on depth 3 circuits computing TR_n^{2d} we suspect that any $\Sigma\Pi\Sigma$ circuit for it has size n^{2d} (and there is a trivial $\Sigma\Pi\Sigma$ circuit of size n^{2d} computing TR_n^{2d}).

2.4.4 Determinant

The last lower bound that we show is for the determinant function. We show an almost n^4 lower bound, where the number of variables is n^2. Note that it is a very weak lower bound compared to what is known over finite fields. Getting a super-polynomial lower bound for this function seems to be the main next challenge

Definition 2.4.16

$$DET_n(X) = \sum_{\sigma \in S_n} sign(\sigma) \prod_{i=1}^{n} x_{i,\sigma(i)} \ ,$$

where S_n is the symmetric group on n variables.

Here X is $n \times n$ matrix.

Theorem 2.4.17 *$s_{\Sigma\Pi\Sigma}(DET_n(X)) \geq \Omega(n^4/\log n)$.*

Denote by $s^m_{\Sigma\Pi\Sigma}(f)$ the size of the smallest depth-3 circuit computing f in which the degree of every multiplication gate is at most m. We have a stronger lower bound for the case where $m = o(\frac{n^2}{\log n})$.

Theorem 2.4.18 *$s^m_{\Sigma\Pi\Sigma}(DET_n(X)) \geq \Omega(e^{n^2/me})$.*

Proof of Theorem 2.4.18: This is an immediate corollary of Proposition 2.2.8 and the fact that $\partial_d(DET_n) = \binom{n}{d}^2$. These facts give:

$$s^m_{\Sigma\Pi\Sigma}(Det_n) \geq \frac{\binom{n}{d}^2}{\binom{m}{d}} \approx (\frac{n^2}{md})^d.$$

Maximizing d we get $d = \frac{n^2}{me}$ and the lower bound follows. $\square$

Using tools like Theorems 2.3.2, 2.3.1 yields a $\frac{n^3}{\log n}$ lower bound. But a straight forward induction together with Theorem 2.4.18 yields a $\frac{n^4}{\log n}$ lower bound. The following simple observation is needed for the induction:

Proposition 2.4.19 *If $\mathcal{C}$ is a $\Sigma\Pi\Sigma$ circuit that computes $f(X)$, where $f(X)$ is a linear function in $x \in X$, i.e $f(X) = xg(X \setminus \{x\}) + h(X \setminus \{x\})$, and x appears in N multiplication gates, each of degree at most m, then there is a $\Sigma\Pi\Sigma$ circuit that computes $g(X \setminus \{x\})$ with $2N$ multiplication gates each of degree at most m.*

Proof: Since $g(X \setminus \{x\}) = f(x+1, X \setminus \{x\}) - f(x, X \setminus \{x\})$, we can construct a new circuit from $\mathcal{C}$ in the obvious manner, and all the gates not including x will be canceled. $\square$

Proof of Theorem 2.4.17: The proof is by induction on n. We begin with $x_{1,n}$, if all the multiplication gates including it are of degree $< \frac{n^2}{4e\log n}$ then we make the following assignment:

$$x_{i,n} = 0 \ \ \forall i > 1.$$

Using Proposition 2.4.19 we get a new circuit of degree $< \frac{n^2}{4e\log n}$ for DET_{n-1} and according to Theorem 2.4.18 there are at least $\frac{1}{2}n^4$ multiplication gates in it. Otherwise we can find a multiplication gate of degree $\geq \frac{n^2}{4e\log n}$ containing $x_{1,n}$ and a linear function $x_{1,n} = \ell_n(X \setminus \{x_{1,n}\})$ that nullifies it. We do this for $x_{1,n}...x_{1,2}$. If at any stage there is no multiplication gate of degree $\geq \frac{n^2}{4e\log n}$ involving the variable we look at, then we make the appropriate assignment (at the $k'th$ stage we put $x_{i,n-k+1} = 0 \ \forall i > 1$). After doing so for $x_{1,2}$ we put $x_{i,1} = 0 \ \forall i > 1, \ \ x_{1,1} = 1$. Thus our restricted circuit computes DET_{n-1}. Therefore we get the recursion:

$$s_{\Sigma\Pi\Sigma}(DET_n) \geq min(s_{\Sigma\Pi\Sigma}(DET_{n-1}) + (n-1)\frac{n^2}{4e\log n}, \frac{n^4}{4})$$

Solving we get $s_{\Sigma\Pi\Sigma}(DET_n) \geq \Omega(\frac{n^4}{\log n})$. $\square$

Restricting the fan-in of the gates on level 1

Since it is hard to prove lower bounds even for such a restricted model, we will restrict the model even further, we will allow the linear functions that are computed at the first level to consist of only one variable, i.e every function will have the form:

$$\ell = \alpha_i x_i + \ell^0$$

The reason to consider this model is that it is strong enough for the construction of [4] for the elementary symmetric polynomials. For this model we can prove an exponential lower bound for DET_n.

Theorem 2.4.20 *Every restricted depth-3 circuit that computes DET_n must have at least $\Omega(\frac{2^n}{n})$ edges.*

Proof: Let's write each multiplication gate in the form:

$$M = \prod_{i=1}^{n} \prod_{j=1}^{n} (\alpha_{i,j} x_{i,j} + \alpha_{i,j}^0)$$

But, in DET_n no two variables from the same column appear in the same monomial, therefore from each of the gates we only need to collect the multilinear monomials in which there are no two variables from the same column. Therefore from each of the terms: $\prod_{j=1}^{n}(\alpha_{i,j}x_{i,j} + \alpha_{i,j}^0)$, only the part of degree 1 can be used to create a monomial that will appear in the result. So we replace each $\prod_{j=1}^{n}(\alpha_{i,j}x_{i,j} + \alpha_{i,j}^0)$ with its degree one monomial: $\sum_{j=1}^{n} \beta_{i,j} x_{i,j}$. Now each multiplication gate looks like

$$M = \prod_{i=1}^{n} (\sum_{j=1}^{n} \beta_{i,j} x_{i,j})$$

The rest of the proof is similar to the proof of Theorem 2.5 from [33]. By Lemma 2.2.6 for each multiplication gate M, $dim(\partial_{\frac{n}{2}}(M)) \leq 2^n$, and we know that $dim(\partial_{\frac{n}{2}}(DET_n)) \geq \binom{n}{\frac{n}{2}}^2$ therefore Lemma 2.2.6 gives us a lower bound on the number of multiplication gate of $\frac{\binom{n}{\frac{n}{2}}^2}{2^n} \approx \frac{2^n}{n}$. $\square$

Remark: Notice that replacing each $\prod_{j=1}^{n}(\alpha_{i,j}x_{i,j} + \alpha_{i,j}^0)$ with $\sum_{j=1}^{n} \beta_{i,j} x_{i,j}$ may change the number of addition gates in the bottom level, but it doesn't raise the degree of each multiplication gate! Since we count the number of multiplication gates and the sum of their degrees, this replacement doesn't change the result.

2.5 Upper bounds for the Elementary symmetric function

Ben-Or showed the following construction for Depth-3 circuits:

Theorem 2.5.1 (Ben-Or) *For every d, $L_{\Sigma\Pi\Sigma}(S_n^d) \leq O(n^2)$.*

Proof: S_n^d is the coefficient of t^{n-d} in the polynomial $\prod_{i=1}^n (t+x_i)$. Using interpolating at n+1 distinct points we can get this coefficient. Notice that evaluation at a point can be computed by a single product gate of degree n, and since we have n+1 points of interpolation the circuit is of size $O(n^2)$. □

In contrast, allowing depth-6 or even depth-4, the construction of Ben-Or can be greatly improved for small d. With our lower bound above for $d = \log n$, it provides a near quadratic separation between depth-3 and depth-4 circuits.

Theorem 2.5.2 *For every d, $L_4(S_n^d) \leq n2^{O(\sqrt{d})}$.*

Theorem 2.5.3 *For every d, $L_6(S_n^d) \leq O(nd^3 \log d)$.*

Proof of Theorem 2.5.3: Let $T_n^d(X) = \sum_{i=1}^n x_i^d$. The well known Newton Identities provide a polynomial relation expressing each S_n^d in terms of $\{T_n^k : k \leq d\}$. More precisely, let the degree d truncation of the series for e^y be the polynomial $E^d(y) = \sum_{j=0}^d y^j/j!$. Then $S_n^d(X)$ is the coefficient of z^d in the polynomial

$$\Pi_{k=1}^d E^{\lfloor d/k \rfloor}((-1)^{k-1} T_n^k z^k)$$

Clearly, this is a $\Pi\Sigma$ circuit of size $d \log d$ in the variables T_n^k, each of which is a $\Pi\Sigma\Pi$ circuit of size at most nd in the original X variables. Finally, interpolation over the variable z (as in Ben-Or's construction) requires $d+1$ values for this variable and gives a $\Sigma\Pi\Sigma\Pi\Sigma\Pi$ circuit of size at most $nd^3 \log d$ for S_n^d. □

Proof of Theorem 2.5.2: We proceed in the same manner to get:

$$\prod_{k=1}^{d} E^{\lfloor d/k \rfloor}((-1)^{k-1} T_n^k z^k) =$$

$$\prod_{k=1}^{d} \sum_{j=0}^{\lfloor d/k \rfloor} \frac{(-1)^{j(k-1)}}{j!} (T_n^k)^j z^{kj}.$$

To get the coefficient of z^d we sum over all the partitions of d:

$$\sum_{\sum_{k=1}^d k j_k = d} \prod_{k=1}^{d} \frac{(-1)^{j_k(k-1)}}{j!} (T_n^k)^{j_k}.$$

Since each $(T_n^k)^{j_k}$ has a $\Pi\Sigma\Pi$ formula of size at most nkj_k, and there are at most $2^{O(\sqrt{d})}$ partitions of d ([27] Theorem 15.7), we get a formula of size at most

$$\sum_{\sum_{k=1}^d k j_k = d} \sum_{k=1}^{d} nkj_k = 2^{O(\sqrt{d})} nd.$$

□

2.6 Strange models

In this section we assume that the field is either $\mathbb{R}$ or $\mathbb{C}$.

2.6.1 A Nečhiporuk-like lower bound

We will now define a different kind of formula, for which we will prove quadratic lower bounds. Our proof uses similar ideas to those of [31]. In his paper Nečhiporuk proves lower bounds on boolean formula by considering the number of subfunctions that a certain function has (a subfunction is the function resulting from restricting some of the variables). In our proof we consider the discriminant polynomial and it's subfunctions resulting from restricting all but one variable.

Definition 2.6.1 *A PolyFormula in the set of variables X over a field F is a binary tree whose internal nodes are labeled by addition or multiplication gate, its inputs are labeled by polynomials in one of the variables of X or by constants from the field. The size of this formula is the number of input nodes that are labeled with polynomials in the variables. We denote the smallest size of a PolyFormula that computes a polynomial f with $\hat{L}(f)$.*

Notice the difference between the definition of PolyFormula and the definition of a Formula where we allow edges to be labeled with constants.

Claim 2.6.2 *If we change the definition of a Formula to a definition where we do not allow constants on edges then this variant changes only affects the number of inputs labeled with constants.*

Proof: We can push constants down towards the inputs until we are left with constants only on edges leaving inputs. Any constant on such edge can be replaced by an input gate labeled with it which fans out to a multiplication gate that will multiply this constant with the appropriate input.
□

Since we count only the number of polynomials in x in the definition of PolyFormula, one may assume that there might be many constants too, but the following lemma bounds the number of constants in terms of $\hat{L}$.

Lemma 2.6.3 *We can always assume that the number of constants in a PolyFormula is at most 3 times the number of inputs labeled with polynomials in X.*

Proof: We can change the PolyFormula (without changing its size or output) such that in the new PolyFormula each constant only enters a plus or a multiplication gate whose other input is an X-input, or it enters a plus gate that adds it to a multiplication gate of two X-polynomials. Otherwise we could delete this constant and change the other constants (we are just throwing away constant sub-formulas). After making this change we can easily prove the claim by an induction on the formula size. □

Definition 2.6.4

$$Disc(X) = \prod_{i<j}(x_j - x_i)$$

X is the set of variables $x_1, ..., x_n$. This is the discriminant polynomial and it is the determinant of the Vandermonde matrix ($a_{i,j} = {x_i}^{j-1}$).

Theorem 2.6.5 $\hat{L}(Disc_n) \geq (1/3)n(n-1)$.

Definition 2.6.6 *A $d-$ scheme is a PolyFormula in one variable, x, with a set of constants $C = \{c_1, ..., c_k\}$, so if we run over all the values for C in F^k we get all deg d monic[1] polynomials in x.*

We denote by $s(d)$ the minimal size of a PolyFormula that is a $d-$ scheme.

A trivial upper bound for $s(d)$ is d.

Theorem 2.6.7 *If F is algebraically closed then $s(3d) \leq 2d$.*

Proof: Let's look at the following PolyFormula:

$$(x^2 + b)(x + a) + (c - ab) = x^3 + ax^2 + bx + c.$$

Thus using only two polynomials in x we can generate every monic deg 3 polynomial. Since F is algebraically closed every monic polynomial of degree $3d$ is the multiplication of d deg 3 monic polynomials whose coefficients are in F. Therefore we can take d distinct copies of this formula, multiply all of them, to get a general deg $3d$ monic polynomial. □

Definition 2.6.8 *We say that a PolyFormula in one variable x, with a set of constants $C = \{c_1, ..., c_k\}$ is of dimension d if running over all possible assignments to C yields a manifold of polynomials in x of dimension d.*

[1] A monic polynomial is a polynomial whose leading coefficient equals 1.

An easy observation is:

Claim 2.6.9 *A PolyFormula of dimension d is a polynomial mapping from F^k to a manifold of dimension d, where k is the number of constants in the PolyFormula.*

Since every polynomial mapping is a C^∞ mapping and F^k is a manifold of dimension k, we get by Theorem 2.6.10 (see [30]) that the dimension of a PolyFormula with k constants is at most k.

Theorem 2.6.10 *If M,N are manifolds of dimensions m, n respectively, where $m < n$ and $\Phi : M \to N$ is a C^1 mapping, then $\Phi(M)$ has measure 0 (Provided M has only countably many components).*

Together with Lemma 2.6.3 we get the following corollary.

Corollary 2.6.11 $\hat{L} \geq k/3$ *for every PolyFormula of dimension k.*

We can now deduce the following lower bound.

Theorem 2.6.12 $s(d) \geq \frac{d}{3}$.

Proof: Follows from Definition 2.6.1, Corollary 2.6.11 and the fact that every d-scheme is a PolyFormula of dimension d. $\square$

We now prove our main theorem.

Proof of Theorem 2.6.5: Let's view a PolyFormula for $Disc(x_1, ..., x_n)$ as a PolyFormula in x_n over the field $F(x_1, ..., x_{n-1})$ and remove all unnecessary gates. Clearly it is a PolyFormula of dimension $n-1$ (when we run over all substitutions of elements from F to $x_1, ..., x_{n-1}$). According to Corollary 2.6.11 the number of input gates labeled with polynomials in x_n is at least $\frac{n-1}{3}$. Since we could do it for every x_i we get $\hat{L}(Disc_n) \geq (1/3)n(n-1)$. $\square$

2.6.2 The derived homogeneous formula

Theorem 2.6.13 (Observation) *If $f(X)$ is an homogeneous polynomial computed by a $\Sigma\Pi\Sigma$ circuit of size s then there's an m such that $g(X,t) = t^m f(X)$ can be computed by an* **homogeneous** *$\Sigma\Pi\Sigma$ circuit of size $\leq s^2$.*

Proof: Let d be the highest degree of all multiplication gates in the $\Sigma\Pi\Sigma$ circuit computing f. We will now change the circuit in the following way:

- Replace each linear function $\ell = \sum_{i=1}^{n} \alpha_i x_i + \ell^h$ with $\hat{\ell} = \sum_{i=1}^{n} \alpha_i x_i + \ell^h t$.

- Multiply each multiplication gate of degree $d-k$ with t^k.

Each monomial of degree k in the former circuit is now multiplied by t^{d-k}, since f is homogeneous the result will be $t^{d-deg(f)} f$. □

Definition 2.6.14 *Let f be an homogeneous polynomial, the derived homogeneous circuit of f is a homogeneous circuit of smallest size that computes $t^m f$ for some m. We call it $\mathcal{F}_D(f)$ and we denote by $s^D_{\Sigma\Pi\Sigma}(f)$ the size of a smallest $\mathcal{F}_D(f)$ for f.*

Corollary 2.6.15 *For S_n^{2d}*
1. $s^D_{\Sigma\Pi\Sigma}(S_n^{2d}) \leq n^2$
2. $s^H_{\Sigma\Pi\Sigma}(S_n^{2d}) \geq \binom{n}{d}^2 / 2^{2d}$

Proof: According to Theorem 2.5.1 there is a $\Sigma\Pi\Sigma$ circuit of size n^2 that computes S_n^{2d}, therefore according to Theorem 2.6.13 there is an homogeneous $\Sigma\Pi\Sigma$ circuit of size n^2 (the size remains the same under this transformation) that computes $t^{n-2d} S_n^{2d}$.

In [33] it was proved that each homogeneous $\Sigma\Pi\Sigma$ circuit that computes S_n^{2d} requires size $\geq \binom{n}{d}^2 / 2^{2d}$. □

Corollary 2.6.16 *There is an exponential gap between computing homogeneously an homogeneous polynomial f, and computing homogeneously $t^m f$ for some m.*

Corollary 2.6.17 *Taking derivatives of an homogeneous polynomial f with respect to a single variable or assigning a fixed value (say 1) to a variable may exponentiate the size of the depth-3 homogeneous circuit computing it.*

The corollary above is an unconditional result (for a weaker model) of a phenomenon which was observed before. In [54] it was shown that assuming PERM is hard, taking multiple derivatives w.r.t. many variables may exponentiate size of general circuits. One way of seeing it is by looking at the following restriction of $PIP(X, Y)$:

Definition 2.6.18 $PIP(X, \vec{y}) = \prod_{i=1}^{n} \sum_{j=1}^{n} x_{i,j} y_j$.

Now we get $PERM(X) = \frac{\partial^n}{\partial y_1 \cdots \partial y_n} PIP(X, \vec{y})$.

2.7 Conclusion

To conclude we will represent some related open problems.

2.7.1 Depth-3 reductions

While quite basic and powerful, depth-3 circuits seem too weak to perform reductions between interesting functions. Most known reductions when depth is not restricted do not seem to carry through when depth is restricted.

The first problem concerns self reductions of elementary symmetric functions. We have proved a lower bound for the $\log n$ elementary symmetric function. The proof doesn't work for S_n^d where $d \geq \log n$, and so we have no lower bound for higher d. In the Boolean setting there is a trivial reduction from smaller to higher values of d (setting variables to 1). Is there an algebraic analog?

Problem 2.7.1 *Does $s_{\Sigma\Pi\Sigma}(S_n^d) \leq s_{\Sigma\Pi\Sigma}(S_n^r)$ for $d < r$?*

For polynomially related functions we define below a natural notion of reduction, and list the most frustrating problems of this type.

Definition 2.7.1 *For two polynomials f, g we say that $f \leq_3 g$ if $s_{\Sigma\Pi\Sigma}(f) \leq s_{\Sigma\Pi\Sigma}(g)^{O(1)}$.*

A very interesting and nontrivial reduction for circuits without depth restriction is between the Determinant and Iterated Matrix Multiplication (see [16]). Again, we do not know if it extends to the bounded depth case.

Problem 2.7.2 *Does $DET \leq_3 IMM$?*

Another curiosity is with the Permanent function.

Definition 2.7.2

$$PERM_n(X) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} x_{i,\sigma(i)}$$

where X is an $n \times n$ matrix and S_n is the symmetric group on n variables.

PERM is known to be $\#P$ complete, and so we expect it to be harder than DET in any model. Surprisingly we have a more efficient depth-3 circuit for PERM than for DET.
The following theorem is due to it Ryser, see e.g [27].

Theorem 2.7.3 (Ryser) *There is an homogeneous $\Sigma\Pi\Sigma$ circuit for $PERM(X)$ of size $O(n2^n)$.*

Proof: $PERM(X) = \sum_{T\subset[n]}(-1)^{|T|}\prod_{i=1}^{n}\sum_{j\in T}x_{i,j}$. □

Problem 2.7.3 *Does $DET \leq_3 PERM$, or maybe $PERM \leq_3 DET$?*

Problem 2.7.4 *Find a simply exponential ($2^{O(n)}$ size) $\Sigma\Pi\Sigma$ circuit for DET.*

2.7.2 Strange models

In Theorem 2.4.20 we showed a lower bound for depth-3 circuits which use linear functions of only one variable. What happens if we allow linear functions of two variables ?

Problem 2.7.5 *Show an exponential lower bound for $\Sigma\Pi\Sigma$ circuits of the form:*

$$\sum_{j=1}^{m}\prod_{i=1}^{deg(M_j)}(\alpha_{i,j,1}x_{k_1}+\alpha_{i,j,2}x_{k_2}+c_{i,j})$$

.

Chapter 3

Affine Projections of Symmetric Polynomials

3.1 Introduction

We introduce a new model for computing polynomials. The new model is a depth two circuit with unbounded fan-in plus gates at the bottom and a gate computing an elementary symmetric polynomial of the gates of the first level, at the top. Clearly each plus gate computes some linear function ℓ_i, and the top gate computes

$$S_m^d(\ell_1, ..., \ell_m)$$

for some d (m is the number of gates in the first level). We call m the size of the circuit, and d its degree. We call such a circuit a symmetric circuit. We show the following:

- Universality of the model: We prove that every polynomial can be computed in this model. More specifically, we show that for every polynomial f we have

  $$s_{\text{sym}}(f) \leq 2^d \cdot d \cdot \text{mon}(f) \ ,$$

  where $d = \deg(f)$, $s_{\text{sym}}(f) =$ the size of a smallest symmetric circuit for f, and $\text{mon}(f) =$ the number of monomials of f.

- Relation to $\Sigma\Pi\Sigma$ circuits: We show that this model is weaker than $\Sigma\Pi\Sigma$ circuits (see definition 2.2.1). In particular we prove the following theorem

  $$s_{\text{sym}}(f) > \sqrt{s_{\Sigma\Pi\Sigma}(f)} \ ,$$

  where $s_{\Sigma\Pi\Sigma}(f)$ is the size of a smallest $\Sigma\Pi\Sigma$ circuit computing f. As a corollary we get that super-quadratic lower bounds for depth 3 circuits imply super-linear lower bounds for the symmetric model. On the other hand if f has a $\Sigma\Pi\Sigma$ circuit with m multiplication gates, each of degree at most d then

$$s_{\text{sym}}(f) \leq d \cdot 2^d \cdot m \ .$$

- Lower bounds for the symmetric model: We prove the following lower bounds

$$s_{\text{sym}}(DET_{\sqrt{n}}) \geq 2n - 3\sqrt{n} \ ,$$

$$s_{\text{sym}}(\prod_{i=1}^{\frac{n}{2}} x_i + \prod_{i=\frac{n}{2}+1}^{n} x_i) \geq \frac{3}{2}n - 2 \ .$$

Using our techniques we are able to extend the results of [49] (as presented in chapter 2) and prove the following lower bound

$$s_{\Sigma\Pi\Sigma}(S_n^d) = \Omega(d(n-d)) \ ,$$

for every $d > 1$. Thus for $d = \alpha n$ for some constant $0 < \alpha < 1$ we have

$$s_{\Sigma\Pi\Sigma}(S_n^{\alpha n}) = \Omega(n^2) \ .$$

Although this model seems a bit obscure, it is interesting to note that similar models are already known. Our model can be viewed as $\text{SYM}(\ell_1, \ldots, \ell_m)$, the similar models $\text{DET}(\ell_{1,1}, \ldots, \ell_{m,m})$ and $\text{PERM}(\ell_{1,1}, \ldots, \ell_{m,m})$ have already been studied. In [57] Valiant shows that the formula size of a polynomial f is at least (up to a small constant) the minimal rank of a matrix A, whose entries are linear forms, such that $det(A) = f$. In [57, 58, 60] Valiant studied p-computable families of polynomials. In this model Valiant shows that the permanent is p-complete (for p-definable polynomials under p-projections), i.e every p-definable polynomial can be represented as the permanent of a matrix with linear functions as entries, such that the rank of the matrix is the p-complexity of the polynomial (actually the linear functions in this model are not general, i.e $\ell \in \{x_i, \ -x_i, \ 0, \ 1\}$). Therefore, our model of symmetric polynomial in linear forms, is an interesting variation of the better known models of permanent and determinant in linear forms.

Another analog of our symmetric model is the model of bounded depth boolean threshold circuits (TC^0 circuits), i.e bounded depth boolean circuits with majority gates. TC^0 circuits received much attention [20, 62, 22], but no exponential lower bounds are known for them. Since the majority function is a symmetric function, we see that over the boolean domain circuits with symmetric gates have also been studied.

The chapter is organized as follows. In section 3.2 we define the model and prove the theorems that show that it is a universal model (i.e. that it can compute all the polynomials). We also prove an upper bound for every polynomial.

In section 3.3 we show the connection to depth 2 and depth 3 arithmetic circuits, and prove that super-quadratic lower bounds for depth 3 circuits imply super-linear lower bounds to the sym model.

In section 3.4 we state our main theorems, and deduce lower bounds from them. In section 3.5 we prove the theorems, and discuss their tightness.

3.1.1 Techniques

The proof of the universality of the model is not very difficult, but it is very interesting to note that there is no such universal property over the real field, $\mathbb{R}$. For example, one can prove that the polynomial $xy + wz$ cannot be represented as an elementary symmetric polynomial in linear forms with real coefficients. The major difference between the real and complex fields which makes the proof work, is that the complex field contains roots of unity of large order.

The proofs of the lower bounds are more involved, they are based on a lemma relating the complexity of computing a polynomial, f, in the symmetric model, and the maximal dimension of a subspace $A \subset \mathbb{C}^n$ such that $f|_A = 0$. In order to use this lemma (lemma 3.4.2 of the paper) it is necessary to find an answer to the following problem

Problem 3.1.1 *What is the maximal dimension of a linear subspace $A \subset \mathbb{C}^m$ such that $S^d_m|_A = 0$?*

Solving this problem is an open problem of independent interest. We are able to give two partial answers, each is tight for a different range of degrees.

Theorem 3.1.1 *For every affine subspace, A, such that $\deg(S^d_m|_A) < d$ we have that*

$$dim(A) \leq \max(m - d, d - 1) \ ,$$

$$dim(A) < \frac{m + d}{2} \ .$$

The proof of this theorem uses some algebraic tools such as partial derivatives, symmetrization of polynomials and the algebraic independence of the elementary symmetric polynomials. This theorem also improves a theorem of a similar nature from [49] (Theorem 2.4.2 of chapter 2).

3.2 Definition of the model

From now on we only consider computations over the complex field, $\mathbb{C}$.

We define another set of symmetric polynomials beside $\{S^d_n\}$ (recall definition 2.4.1).

Definition 3.2.1 *We denote by $T_n^d(x_1, ..., x_n)$ the d'th power sum,*

$$T_n^d(x_1, ..., x_n) = \sum_{i=1}^{n} {x_i}^d .$$

We are now ready to define our model.

Definition 3.2.2 *A depth 2 symmetric circuit is a layered graph with two levels. Each node at the first level computes a linear function in the set of variables X. Denote these linear functions with $L_1, ..., L_m$. At the second level there is only one node, which computes $S_m^d(L_1, ..., L_m)$. In other words it computes the d'th elementary symmetric polynomial in the linear functions computed in the first level of the circuit. The polynomial computed at the node of the second level is the output of the circuit. The size of this circuit is m, i.e. it is the number of nodes in the first level. The degree of the circuit is d.*

We say that a function f is computed by a symmetric circuit of size m and degree d if there is such a circuit that computes f as its output. For a function f that can be computed by a symmetric circuit we denote by $s_{\mathrm{sym}}(f)$ the size of a smallest symmetric circuit, of degree $\deg(f)$, that computes f.

Comment: Although our definition allows the circuit to be of arbitrarily high degree, we will only be interested in circuits which have the same degree as the degree of the polynomial they compute. So from now on we will only be interested in computations of the form

$$f = S_m^{\mathrm{deg(f)}}(\ell_1, ..., \ell_m) .$$

It is easy to see that a symmetric circuit computes a polynomial of degree at most d in X. It is not clear however, that it can compute all the polynomials of degree d. The following theorem assures us that it can, and gives a generic upper bound for every polynomial.

Theorem 3.2.3 *Every polynomial $f \in \mathbb{C}[x_1, ..., x_n]$ of degree d can be computed by a symmetric circuit. Moreover, $s_{\mathrm{sym}}(f) \leq d \cdot 2^d \cdot \mathrm{mon}(f)$, where $\mathrm{mon}(f) =$ the number of monomials of f.*

The proof of Theorem 3.2.3 is in two stages. First we show that every polynomial can be represented as a sum of powers of linear functions, then we show that every sum of powers of linear functions can be computed by a symmetric circuit.

In order to show that every polynomial can be represented as a sum of powers of linear functions, we first show how to represent one monomial as a sum of powers of linear functions.

Lemma 3.2.4 *There exist* 2^d *linear function* $L_1, ..., L_{2^d}$ *such that*

$$\prod_{i=1}^{d} x_i = \sum_{i=1}^{2^d} (L_i)^d .$$

The proof of the lemma is just an arithmetic analog of the well known inclusion-exclusion principle.

Proof: We first prove the following claim:

Claim 3.2.5

$$\prod_{i=1}^{d} x_i = (-1)^d \frac{1}{d!} \sum_{T \subset [d]} (-1)^{|T|} (\sum_{i \in T} x_i)^d .$$

Proof: The coefficient of $\prod_{i=1}^{d} x_i$ in the right hand-side is 1, therefore it is sufficient to prove that the coefficient of any other monomial is 0 there. Since the polynomial we compute is homogeneous, any monomial other then $\prod_{i=1}^{d} x_i$ does not contain all the variables, so let's look at the monomial $\prod_{i=1}^{k} {x_i}^{\alpha_i}$ (where $k < d$). This monomial is computed by the sub-sum

$$(-1)^d \frac{1}{d!} \sum_{[k] \subset T \subset [d]} (-1)^{|T|} (\sum_{i \in T} x_i)^d.$$

The coefficient of $\prod_{i=1}^{k} {x_i}^{\alpha_i}$ in $(\sum_{i \in T} x_i)^d$ is $\binom{d}{\alpha_1 \alpha_2 \ldots \alpha_k}$ (for every $[k] \subset T \subset [d]$). Therefore the coefficient of $\prod_{i=1}^{k} {x_i}^{\alpha_i}$ in the right hand-side is

$$(-1)^d \frac{1}{d!} \sum_{[k] \subset T \subset [d]} (-1)^{|T|} \binom{d}{\alpha_1 \alpha_2 \ldots \alpha_k} =$$

$$(-1)^{d+k} \frac{1}{d!} \binom{d}{\alpha_1 \alpha_2 \ldots \alpha_k} \sum_{r=0}^{d-k} \binom{d-k}{r} (-1)^r = 0.$$

□

Now we can write

$$\left((-1)^d \frac{1}{d!} (-1)^{|T|} \right)^{\frac{1}{d}} = c_T$$

for some $c_T \in \mathbb{C}$. So we get:

$$\prod_{i=1}^{d} x_i = \sum_{T \subset [d]} (c_T (\sum_{i \in T} x_i))^d = \sum_{i=1}^{2^d} {L_i}^d.$$

for some linear forms $L_1, ..., L_{2^d}$. □

Since every polynomial is a sum of monomials we get that every polynomial can be represented as a sum of powers of linear functions.

Lemma 3.2.6 *Every polynomial $f \in \mathbb{C}[x_1, ..., x_n]$ of degree d can be written as sum of at most $2^d \cdot \operatorname{mon}(f)$ powers of linear functions.*

Proof: Let $f \in \mathbb{C}[x_1, ..., x_n]$ be a polynomial of degree d. Let z be a new variable. Multiply every monomial of f of degree k, such that $k < d$, with z^{d-k}. After doing so we get a new function $\tilde{f} \in \mathbb{C}[x_1, ..., x_n, z]$ such that

$$\tilde{f}(x_1, ..., x_n, 1) = f(x_1, ..., x_n).$$

According to Lemma 3.2.4 for every monomial M of $\tilde{f}$ there exists 2^d linear functions $\tilde{L}_{M,1}, .., \tilde{L}_{M,2^d}$ such that $M = \sum_{i=1}^{2^d} (\tilde{L}_{M,i})^d$. Since

$$\tilde{f} = \sum_{monomial} M$$

we get that

$$\tilde{f} = \sum_{monomial} \sum_{i=1}^{2^d} (\tilde{L}_{M,i})^d \ .$$

Hence there exists $r = 2^d \cdot \operatorname{mon}(\tilde{f})$ linear functions, $\{\tilde{L}_i\}_{i=1}^r$, satisfying $\tilde{f} = \sum_{i=1}^r (\tilde{L}_i)^d$. Let $L_i = \tilde{L}_i|_{z=1}$, we get $f = \sum_{i=1}^r (L_i)^d$.

Notice that if f is homogeneous then $f = \tilde{f}$ and the L_i's are linear forms. □

Next we show that sum of powers of linear functions can be represented as an elementary symmetric function in linear functions.

Lemma 3.2.7 *Let w be a primitive root of unity of order d. We have:*

$$\sum_{i=1}^{k} {y_i}^d = -S_{kd}^d(-y_1, -wy_1, -w^2y_1, \ldots$$

$$\ldots, -w^{d-1}y_1, -y_2, -wy_2, \ldots, -w^{d-1}y_k) \ .$$

Proof: $S_m^r(z_1, ..., z_m)$ is the coefficient of t^{m-r} in the polynomial $\prod_{i=1}^m (t + z_i)$. So the coefficient of t^{kd-d} in

$$\prod_{\substack{i = 1, ..., k \\ j = 0, ..., d-1}} (t - w^j y_i)$$

is

$$S_{kd}^d(-y_1, -wy_1, -w^2y_1, \ldots$$

$$\ldots, -w^{d-1}y_1, -y_2, -wy_2, ..., -w^{d-1}y_k) \ .$$

Now

$$\prod_{\substack{i=1,\ldots,k \\ j=0,\ldots,d-1}} (t - w^j y_i) = \prod_{i=1}^{k} (t^d - {y_i}^d).$$

Therefore the coefficient of t^{kd-d} is $-\sum_{i=1}^{k} {y_i}^d$. And the result follows. $\square$

Let ρ be a $2d$ primitive root of unity. We have

$$\sum_{i=1}^{k} {y_i}^d = \rho^d S^d_{kd}(-y_1, -\rho^2 y_1, -\rho^4 y_1, \ldots$$
$$\ldots, -\rho^{2d-2} y_1, -y_2, -\rho^2 y_2, \ldots, -\rho^{2d-2} y_k) =$$
$$S^d_{kd}(-\rho y_1, -\rho^3 y_1, -\rho^5 y_1, \ldots$$
$$\ldots, -\rho^{2d-1} y_1, -\rho y_2, -\rho^3 y_2, \ldots, -\rho^{2d-1} y_k).$$

Now we can deduce our theorem.

Proof of Theorem 3.2.3: Put $r = 2^d \cdot \mathrm{mon}(f)$, and write $f = \sum_{i=1}^{r} (L_i)^d$ (Lemma 3.2.6). We have

$$f = \sum_{i=1}^{r} (L_i)^d = S^d_{rd}(-wL_1, -w^3 L_1, -w^5 L_1, \ldots$$
$$\ldots, -w^{2d-1} L_1, -wL_2, -w^3 L_2, \ldots, -w^{2d-1} L_r)$$

($w^{2d} = 1$ a primitive root of unity). $\square$

NOTE: The proof of theorem 3.2.3 is valid over any fields that has a $2d$-primitive root of unity, an element c such that $c^d = d!$, and the existence of $\frac{1}{d!}$.

3.3 Symmetric circuits vs. depth-3 circuits

In this section we show the relation between the symmetric model and depth 2 and 3 circuits. This relation actually motivated the study of symmetric circuits. In [49] quadratic lower bounds were proved for depth 3 circuits computing some explicit polynomials. However, proving super-quadratic lower bounds for $\Sigma\Pi\Sigma$ circuits, over fields of characteristic 0, remains an open problem. We prove that super-quadratic lower bounds for $\Sigma\Pi\Sigma$ circuits imply super-linear lower bounds for symmetric circuits. The difficulty to prove super-quadratic lower bounds for depth 3 circuits motivates the attempts to prove super-linear lower bounds for symmetric circuits, which we hope will be an easier task.

Definition 3.3.1 *A $\Sigma\Pi$ circuit is a layered depth 2 circuit with a plus gate at the top and product gates at the bottom. The size of a $\Sigma\Pi$ circuit is the number of edges in it. For a polynomial f we denote by $s_{\Sigma\Pi}(f)$ the size of a smallest $\Sigma\Pi$ circuit that computes f.*

This model of computation is not very interesting since the number of gates at the bottom level is clearly the number of monomials of f. Thus, for a homogeneous polynomial f we have

$$s_{\Sigma\Pi}(f) = \deg(f) \cdot \mathrm{mon}(f) \; .$$

Together with theorem 3.2.3 we get the following corollary.

Corollary 3.3.2 *For a homogeneous polynomial f we have: $s_{\mathrm{sym}}(f) \leq 2^{\deg(f)} \cdot s_{\Sigma\Pi}(f)$.*

Recall definition 2.2.1 of $\Sigma\Pi\Sigma$ circuits. From the definition it is clear that if f is computed by a $\Sigma\Pi\Sigma$ circuit with s multiplication gates, then f has a representation of the form

$$f = \sum_{j=1}^{s} M_j$$

where

$$M_j = \prod_{i=1}^{\deg(M_j)} \ell_{i,j} \; ,$$

each $\ell_{i,j}$ is a linear function in the input variables, and $\deg(M_j)$ is the fan-in of the j'th multiplication gate. We stress that the $\ell_{i,j}$'s are linear functions and may involve a constant term (and indeed without this ability the model is homogeneous, and very restricted as shown in [32, 33]), and that different multiplication gates may use the same linear function.

A slight modification of theorem 2.5.1 relates the sizes of a depth 3 formulae for f and a symmetric circuit for it (since it is a straight forward generalization we omit the proof).

Theorem 3.3.3 ([4]) *For every d, m and any m linear functions $L_1, ..., L_m$, there is a $\Sigma\Pi\Sigma$ formulae of size $m(m+1)$ computing $S^d_m(L_1, \ldots, L_m)$.*

As a corollary we get that if f can be computed by a symmetric circuit of size m then there is a depth 3 formulae of size $m(m+1)$ computing f. Combining this with corollary 3.3.2 we get:

Theorem 3.3.4 *For a homogeneous polynomial f we have:*

$$\sqrt{s_{\Sigma\Pi\Sigma}(f)} - 1 \leq s_{\mathrm{sym}}(f) \leq 2^{\deg(f)} \cdot s_{\Sigma\Pi}(f) \; .$$

As a corollary we see that every super-quadratic lower bound for depth 3 circuits will imply super-linear lower bound for the symmetric model. Hence it seems an easier task to prove super-linear lower bounds for the symmetric model. We prove two nontrivial linear lower bounds, but although our lower bounds are linear it is important to notice that they do not follow from the results on $\Sigma\Pi\Sigma$ circuits, and their proofs require new techniques.

Despite the fact that we cannot prove any strong separation between the symmetric model and depth 3 circuits we conjecture the following.

Conjecture 3.3.1 *There is a polynomial, f, in n variables such that $s_{\Sigma\Pi\Sigma}(f)$ is polynomial in n and $s_{\mathrm{sym}}(f)$ is exponential in n. In particular, we suspect that*

$$f(x_1, ..., x_n) = \prod_{i=1}^{\frac{n}{2}} x_i + \prod_{i=\frac{n}{2}+1}^{n} x_i$$

is such a polynomial.

3.4 Results

In this section we prove lower bounds for symmetric circuits and depth 3 circuits. Our bounds for the symmetric model uses the following lemma that shows that good upper bounds on the dimension of every affine subspace $A \in \mathbb{C}^m$ such that $S_m^d|_A = 0$ imply lower bounds on many functions. Before stating the lemma we need the following definition.

Definition 3.4.1 *We say that a polynomial f in n variables is non-trivial, if it cannot be represented as a polynomial of less than n independent linear functions.*

Lemma 3.4.2 *Assume that there is a function g such that whenever $S_m^d|_A = 0$, where A is an affine subspace, the dimension of A is at most $g(m, d)$. Then for every non-trivial polynomial f in n variables, that has a subspace $B \subset \mathbb{C}^n$ of dimension D satisfying $f|_B = 0$, we must have*

$$g(s_{\mathrm{sym}}(f), \deg(f)) \geq D \; .$$

Proof: Assume that there is a symmetric circuit for f of size m and degree $d = \deg(f)$. Then we can write

$$f(x_1, ..., x_n) = S_m^d(L_1, ..., L_m)$$

for some linear functions $L_1, ..., L_m$. Let $B \subset \mathbb{C}^n$ be a linear subspace of dimension D such that $f|_B = 0$. Notice that since $f|_B = 0$, we have

$$0 = S_m^d(L_1, ..., L_m)|_B = S_m^d(L_1|_B, ..., L_m|_B).$$

Since $dim(B) = D$, the non-triviality of f implies that there are D independent linear functions among $\{L_i|_B\}$. Hence there is a subspace of dimension D such that S_m^d vanishes on it (this subspace is the image of B under the linear transformation $(x_1, ..., x_n) \to (L_1, ..., L_m)$). According to our assumption we get

$$g(m, d) \geq D \ .$$

As a result we have

$$g(s_{\mathrm{sym}}(f), \deg(f)) \geq D \ .$$

□

Thus an upper bound on the dimension of a linear space A, such that $S_m^d|_A = 0$, will give lower bounds for symmetric circuits. The following theorems give such upper bounds.

Theorem 3.4.3 *For every affine subspace, A, such that $\deg(S_m^d|_A) < d$ we have that*

$$dim(A) \leq \max(m - d, d - 1) \ .$$

This theorem is tight for $d > \frac{m}{2}$, for smaller values of d we have the following theorem.

Theorem 3.4.4 *For every $d \geq 2$ and every affine subspace, A, such that $\deg(S_m^d|_A) < d$ we have that*

$$dim(A) < \frac{m+d}{2} \ .$$

Combining lemma 3.4.2 with theorems 3.4.3, 3.4.4 yields the following corollary.

Corollary 3.4.5 *For every non-trivial polynomial f in n variables, that has a subspace $B \subset \mathbb{C}^n$ of dimension D satisfying $f|_B = 0$, we have:*

$$\max(S_{\mathrm{sym}}(f) - \deg(\mathrm{f}), \deg(\mathrm{f}) - 1) \geq D \ ,$$

and

$$\frac{S_{\mathrm{sym}}(f) + \deg(\mathrm{f})}{2} > D \ .$$

Therefore

$$S_{\mathrm{sym}}(f) > 2D - \deg(\mathrm{f}) \ ,$$

and if $\deg(\mathrm{f}) - 1 < S_{\mathrm{sym}}(f) - \deg(\mathrm{f})$ *then we also have*

$$S_{\mathrm{sym}}(f) \geq D + \deg(\mathrm{f}) \ .$$

Before proving the theorems we show how to use them to prove lower bounds.

3.4.1 Lower bounds

In this section we use lemma 3.4.2 to prove lower bounds for explicit functions such as the determinant, and the sum of two monomials. We also use theorem 3.4.4 to strengthen theorem 2.4.2 of chapter 2.

Lower Bounds for the Symmetric Model

Theorem 3.4.6

$$s_{\mathrm{sym}}(DET_{\sqrt{n}}) \geq 2n - 3\sqrt{n} \ . \tag{3.1}$$

$$s_{\mathrm{sym}}(\prod_{i=1}^{\frac{n}{2}} x_i + \prod_{i=\frac{n}{2}+1}^{n} x_i) \geq \frac{3}{2}n - 2 \ . \tag{3.2}$$

Proof: We first prove the theorem for $DET_{\sqrt{n}}$. Let B be the subspace of all $\sqrt{n} \times \sqrt{n}$ matrices whose first row is all zeros. Clearly $DET_{\sqrt{n}}|_B = 0$, $\dim(B) = n - \sqrt{n}$. According to corollary 3.4.5 we get that

$$S_{\mathrm{sym}}(DET_{\sqrt{n}}) \geq 2(n - \sqrt{n}) - \sqrt{n} = 2n - 3\sqrt{n} \ .$$

We now prove the lower bound for $\prod_{i=1}^{\frac{n}{2}} x_i + \prod_{i=\frac{n}{2}+1}^{n} x_i$. Let B be the subspace defined by $x_1 = x_n = 0$. Clearly $f|_B = 0$, and $\dim(B) = n - 2$. Plugging it into corollary 3.4.5 we get

$$\max(S_{\mathrm{sym}}(\prod_{i=1}^{\frac{n}{2}} x_i + \prod_{i=\frac{n}{2}+1}^{n} x_i) - \frac{n}{2}, \frac{n}{2} - 1) \geq n - 2 \ ,$$

hence $S_{\mathrm{sym}}(\prod_{i=1}^{\frac{n}{2}} x_i + \prod_{i=\frac{n}{2}+1}^{n} x_i) \geq \frac{3}{2}n - 2$. □

Comment: In order to apply lemma 3.4.2 to our functions, we have to show that they are non-trivial, since it is quite obvious we omit the proof.

Although we can only prove linear lower bounds, the following argument shows, as usual, that almost all polynomials are hard for the symmetric model.

Comment: The set $\{ f \mid \deg(f) = d \ , \ s_{\text{sym}}(f) \leq m \}$ is a variety of dimension $\leq mn$ in the space of all degree d polynomials in n variables. Since this space is of dimension $\binom{n+d-1}{d}$ we have a lower bound of

$$s_{\text{sym}}(f) \geq \frac{\binom{n+d-1}{d}}{n}$$

for a generic polynomial f in n variables of degree d.

Lower Bounds for $\Sigma\Pi\Sigma$ circuits

The next lower bound extends theorem 2.4.2 of chapter 2 to a wider range of degrees.

Theorem 3.4.7 *$s_{\Sigma\Pi\Sigma}(S_n^d) = \Omega(d(n-d))$ for every $d \geq 2$.*

Proof: Assume that C is a depth 3 circuit that computes S_n^d. We want to show that there are at least $\frac{n-d}{2}$ multiplication gates of degree $\geq d$ in C. This will certainly imply the theorem. So, suppose that there are less then $\frac{n-d}{2}$ such gates. We denote these gates with $M_1, ..., M_r$ for some $r < \frac{n-d}{2}$. The following claim, which is similar to lemma 2.3.3 of chapter 2, enables us to use theorem 3.4.4 in the scenery of depth 3 circuits.

Claim 3.4.8 *There is an affine subspace $A \subset \mathbb{C}^n$, such that $\forall i \leq r \quad M_i|_A =$ constant, and $dim(A) > \frac{n+d}{2}$.*

Proof: We will find linear functions $\ell_1, ..., \ell_k$ (for some $k \leq r$) such that each ℓ_i divides a different gate M_i, (i.e. there is a gate in the first level that computes ℓ_i that fans into M_i), and such that all the gates $M_1, ..., M_r$ vanish or set to constants when restricted to the set of common zeros of all the ℓ_i's.

So let $\ell_1 = \ell_{i_1}$ be some linear function that divides M_1. Let ℓ_2 be a linear function that divides M_2 such that ℓ_{i_1} and ℓ_2 have a common zero (i.e. there is $x \in \mathbb{C}^n$ such that $\ell_{i_1}(x) = \ell_2(x) = 0$). If no such ℓ_2 exists then we move to M_3. Suppose we reached M_k and we already picked $\ell_{i_1}, \ell_{i_2}, ..., \ell_{i_t}$ ($t < k$ of course). We pick ℓ_k to be a linear function that divides M_k such that $\ell_{i_1}, \ell_{i_2}, ..., \ell_{i_t}, \ell_k$ have a common zero. If there is no such ℓ_k then we proceed to M_{k+1}. After r steps we stop. Thus we have a set $\ell_{i_1}, \ell_{i_2}, ..., \ell_{i_t}$ for some $t \leq r$ such that all the ℓ's have a common zero. Let A be their set of common zeros:

$$A = \{ x \in \mathbb{C}^n \mid \ell_{i_1}(x) = \ell_{i_2}(x) = \ldots = \ell_{i_t}(x) = 0 \} .$$

Since A is defined by at most $r < \frac{n-d}{2}$ linear equations we have that $dim(A) \geq n - r > \frac{n+d}{2}$. Clearly $M_{i_j}|_A = 0$ for every $1 \leq j \leq t$. In addition we have that any other multiplication gate M_i

for $1 \leq i \leq r$ is set to constant when restricted to A. This is because either one of the ℓ_{i_j} divides M_i, or all the linear functions that divide M_i are set to constants when restricted to A (this follows immediately from the fact that we couldn't find a linear function that divides M_i that is linear independent from the linear functions we already chose). $\square$

Let A be the subspace guaranteed by the claim, we consider $S_n^d|_A$. Notice that $C|_A$ computes $S_n^d|_A$. In $C|_A$ all the gates are either constants or of degree less than d. Therefore $\deg(S_n^d|_A) < d$. So according to theorem 3.4.4 we must have $dim(A) < \frac{n+d}{2}$, in contradiction. $\square$

Corollary 3.4.9 *For every constant* $0 < \alpha < 1$, $s_{\Sigma\Pi\Sigma}(S_n^{\alpha n}) = \theta(n^2)$.

Proof: Immediate from theorems 3.3.3, 3.4.7. $\square$

3.5 Proof of theorems

We first introduce the algebraic tools that we will need for the proofs.

3.5.1 Algebraic tools

The fact that we will use in the proofs of both theorem 3.4.3 and theorem 3.4.4 is the basic theorem of the symmetric functions. This theorem says that every symmetric polynomial can be represented as a polynomial in some elementary symmetric polynomials (see [41]).

Theorem 3.5.1 (basic theorem of symmetric functions) *Every symmetric polynomial*

$$f \in \mathbb{C}[x_1, ..., x_m]$$

can be written in a unique way as a polynomial in $T_m^1, ..., T_m^m$, *or as a polynomial in* $S_m^1, ..., S_m^m$.

This uniqueness of representation shows that $T_m^1, ..., T_m^m$ *are algebraically independent, and so are* $S_m^1, ..., S_m^m$.

It is interesting to note that although our theorems speak about one polynomial we need to use the algebraic independence of all the elementary symmetric polynomials.

The main technical tool in our proof is the symmetrizing of polynomials. We now define a linear operator that symmetrize polynomials.

Definition 3.5.2 *For a polynomial $f \in \mathbb{C}[X_1, ..., X_k]$ define*

$$\Psi(f) = \frac{1}{k!} \sum_{\sigma \in S_k} f(x_{\sigma(1)}, ..., x_{\sigma(k)}).$$

Proposition 3.5.3 *We have some easy observations about Ψ:*

- *$\Psi(f)$ is a symmetric polynomial of degree $\leq \deg(f)$.*
- *If g is a symmetric polynomial then $\Psi(gf) = g\Psi(f)$.*
- *Ψ is a linear operator, i.e. $\Psi(\alpha f + \beta g) = \alpha\Psi(f) + \beta\Psi(g)$ (α, β are constants from the field).*

Now that we have the basic algebraic tools we can proof our theorems.

3.5.2 Zeros of symmetric polynomials

In order to prove theorem 3.4.3 we need the following lemma that reduces the case of an affine subspace to the case of a linear subspace.

Lemma 3.5.4 *Let $A = V + u$ be an affine subspace of $\mathbb{C}^m$, where $V \subset \mathbb{C}^m$ is a linear space and $u \in \mathbb{C}^m$, such that $\deg(S_m^d|_A) < d$. Then $S_m^d|_V = 0$.*

Proof: Let A be defined by

$$\begin{aligned} x_{k+1} &= \ell_1(x_1, ..., x_k) \ , \\ x_{k+2} &= \ell_2(x_1, ..., x_k) \ , \\ &\cdots \\ x_m &= \ell_{m-k}(x_1, ..., x_k) \ . \end{aligned}$$

Write $\ell_i = L_i + \alpha_i$ where L_i is a linear form, and $\alpha_i \in \mathbb{C}$. Since $\deg(S_m^d|_A) < d$ we get that the homogeneous part of degree d of $S_m^d|_A$ is zero. Since

$$S_m^d|_A = S_m^d(x_1, ..., x_k, \ell_1, ..., \ell_{m-k}) \ ,$$

the homogeneous part of degree d is

$$S_m^d(x_1, ..., x_k, L_1, ..., L_{m-k}) \ ,$$

which equals to zero according to our assumption. Since V is the linear space defined by the equations

$$\begin{aligned} x_{k+1} &= L_1(x_1, ..., x_k) \ , \\ x_{k+2} &= L_2(x_1, ..., x_k) \ , \\ &\cdots \\ x_m &= L_{m-k}(x_1, ..., x_k) \ , \end{aligned}$$

we have that $S_m^d|_V = 0$. Hence, the result follows. $\square$

We can now turn to the proof of the Theorems.

3.5.3 Proof of theorem 3.4.3

We are going to assume by contradiction that there exist a subspace A of dimension k such that $k > \max(m-d, d-1)$ and such that $\deg(S_m^d|_A) < d$. We will show that as a consequence of our assumption we can write S_k^d as a polynomial in $S_k^1, S_k^2, ..., S_k^{d-1}$. This is a contradiction because of the basic theorem of the symmetric functions.

Proof: Lemma 3.5.4 allows us to assume w.l.o.g. that A is a linear subspace and that $S_m^d|_A = 0$. So assume that $dim(A) = k$, and let A be defined by

$$\begin{aligned} x_{k+1} &= L_1(x_1, ..., x_k), \\ x_{k+2} &= L_2(x_1, ..., x_k), \\ ... &= ... \\ x_m &= L_{m-k}(x_1, ..., x_k). \end{aligned}$$

In other words,

$$A = \{\ (x_1, ..., x_m) \mid x_{k+i} = L_i,\ \ \forall 1 \leq i \leq m-k\ \}.$$

We have

$$0 = S_m^d|_A = S_m^d(x_1, ..., x_k, L_1, ..., L_{m-k}) =$$
$$S_k^d(x_1, ..., x_k) + S_{m-k}^d(L_1, ..., L_{m-k}) +$$
$$+ \sum_{i=1}^{d-1} S_k^{d-i}(x_1, ..., x_k) \cdot S_{m-k}^i(L_1, ..., L_{m-k})\ .$$

Assume by contradiction that $k > \max(m-d, d-1)$. Since $k > m-d$, we have that $d > m-k$ and therefore

$$S_{m-k}^d(L_1, ..., L_{m-k}) = 0.$$

So we have

$$0 = S_k^d(x_1, ..., x_k) +$$
$$+ \sum_{i=1}^{d-1} S_k^{d-i}(x_1, ..., x_k) \cdot S_{m-k}^i(L_1, ..., L_{m-k})\ . \tag{3.3}$$

We now apply Ψ (as defined in definition 3.5.2) to both sides of the equation and get:

$$0 = \Psi(S_k^d(x_1, ..., x_k) +$$
$$+ \sum_{i=1}^{d-1} S_k^{d-i}(x_1, ..., x_k) \cdot S_{m-k}^i(L_1, ..., L_{m-k}))\ .$$

From the linearity of Ψ and proposition 3.5.3 we get that

$$0 = \Psi\left(S_k^d(x_1, ..., x_k)\right) +$$

$$+\sum_{i=1}^{d-1} \Psi\left(S_k^{d-i}(x_1, ..., x_k) \cdot S_{m-k}^i(L_1, ..., L_{m-k})\right) =$$

$$= S_k^d(x_1, ..., x_k) +$$

$$+\sum_{i=1}^{d-1} S_k^{d-i}(x_1, ..., x_k) \cdot \Psi\left(S_{m-k}^i(L_1, ..., L_{m-k})\right) .$$

Each $\Psi\left(S_{m-k}^l(L_1, ..., L_{m-k})\right)$ (where $l < d$) is a symmetric polynomial of degree at most $d-1$ in the k variables $x_1, ..., x_k$. Theorem 3.5.1 shows that it is a polynomial in

$$S_k^1(x_1, ..., x_k), S_k^2(x_1, ..., x_k), ..., S_k^{d-1}(x_1, ..., x_k).$$

So after applying Ψ to both sides of equation 3.3 we can write the result as

$$0 = S_k^d(x_1, ..., x_k) + Q(S_k^1, S_k^2, ..., S_k^{d-1}),$$

(for some polynomial Q in $d-1$ variables) i.e.

$$S_k^1(x_1, ..., x_k), S_k^2(x_1, ..., x_k), ..., S_k^d(x_1, ..., x_k)$$

are algebraic dependent. But we assumed that $k > d-1$, so by theorem 3.5.1 we have a contradiction. □

3.5.4 Proof of theorem 3.4.4

As in the proof of theorem 3.4.3 we will use the linear operator Ψ that symmetrizes polynomials. In addition we will also have to consider partial derivatives of polynomials.

Partial derivatives with respect to vectors

Definition 3.5.5 *For $v \in \mathbb{C}^m$, and a polynomial $f \in \mathbb{C}[X_1, ..., X_m]$ define $\partial_v(f) = \sum_{i=1}^m v_i \frac{\partial f}{\partial x_i}$ (where $v = (v_1, ..., v_m)$).*

The following proposition shows some easy properties of ∂_v. It is similar in spirit to some propositions from [14, 33].

Proposition 3.5.6 • *$\partial_v(L) = L(v) = L(v_1, ..., v_m)$ for every linear function L.*

- $\partial_v(fg) = g\partial_v(f) + f\partial_v(g)$, *for every pair of functions* f, g.
- $\partial_v(\prod_{i=1}^t L_i) = \sum_{i=1}^t (L_i(v) \prod_{j\neq i} L_j)$, *for every set of linear functions* $L_1, ..., L_t \in \mathbb{C}[X_1, ..., X_m]$ *(follows immediately from the previous properties).*

The following lemma shows that for a vector v of a special structure, $\partial_v(S_k^r(x_1, ..., x_k))$ is equal to $S_{d-1}^{r-1}(x_1, ..., x_{d-1})$, when restricted to a certain subspace of the inputs.

Lemma 3.5.7 *Let* U *be the following subspace*

$$U = \{\ (x_1, ..., x_k) \in \mathbb{C}^k \mid x_d = x_{d+1} = ... = x_k = 0\}.$$

Let $v \in \mathbb{C}^k$ *be a vector satisfying:*

1. $v_1 = v_2 = ... = v_{d-1} = 1$
2. $\sum_{i=1}^k v_i = 0$

Then for every $r \leq d$ *we have*

$$\partial_v\ (S_k^r(x_1, ..., x_k))\ |_U = (1-r)S_{d-1}^{r-1}(x_1, ..., x_{d-1})\ .$$

Proof: From the linearity of the derivative we have

$$\partial_v(S_k^r(x_1, ..., x_k))|_U = \sum_{\substack{T \subset [k] \\ |T| = r}} \partial_v(\prod_{i\in T} x_i)|_U\ .$$

According to proposition 3.5.6 the right hand side is equal to

$$\sum_{\substack{T \subset [k] \\ |T| = r}} (\sum_{i\in T} v_i \prod_{\substack{j \in T \\ j \neq i}} x_j)|_U =$$

$$= \sum_{\substack{T \subset [k] \\ |T| = r-1}} (\sum_{i\notin T} v_i)(\prod_{i\in T} x_i)|_U\ .$$

From the definition of U, we have that $x_d, ..., x_k = 0$. Therefore the right hand side is equal to

$$\sum_{\substack{T \subset [d-1] \\ |T| = r-1}} (\sum_{i\notin T} v_i)(\prod_{i\in T} x_i)\ .$$

Since v satisfies that

$$v_1 = v_2 = ... = v_{d-1} = 1\ ,\ \textit{and}$$

$$\sum_{i=1}^{k} v_i = 0 \ ,$$

we have that for $r \leq d-1$,

$$\sum_{i=r}^{k} v_i = -(r-1) = 1-r \ .$$

So we get

$$\begin{aligned} &\partial_v(S_k^r(x_1,...,x_k))|_U = \\ &\sum_{\substack{T \subset [d-1] \\ |T| = r-1}} (1-r)(\prod_{i \in T} x_i) = \\ &(1-r)S_{d-1}^{r-1}(x_1,...,x_{d-1}) \ . \end{aligned} \tag{3.4}$$

□

The following basic lemma shows the conditions under which we can find a vector satisfying the properties of lemma 3.5.7.

Lemma 3.5.8 *In every vector space of dimension r there is a vector with at least r coordinates that are all 1.*

The proof

The proof of Theorem 3.4.4 has the same spirit as the proof of theorem 3.4.3, but in order to write S_k^d as a polynomial in $S_k^1, S_k^2, ..., S_k^{d-1}$ we need to use partial derivatives as well.

Proof of Theorem 3.4.4 As before, lemma 3.5.4 allows us to assume w.l.o.g. that A is a linear subspace and that $S_m^d|_A = 0$. So assume that $dim(A) = k$, and let A be defined by

$$\begin{aligned} x_{k+1} &= L_1(x_1,...,x_k), \\ x_{k+2} &= L_2(x_1,...,x_k), \\ ... &= \ ... \\ x_m &= L_{m-k}(x_1,...,x_k). \end{aligned}$$

We have:

$$0 = S_m^d|_A = S_m^d(x_1,...,x_k,L_1,...,L_{m-k}) =$$

$$S_k^d(x_1,...,x_k) + S_{m-k}^d(L_1,...,L_{m-k}) +$$

$$\sum_{i=1}^{d-1} S_k^i(x_1,...,x_k)\cdot S_{m-k}^{d-i}(L_1,...,L_{m-k})\ . \tag{3.5}$$

Assume by contradiction that

$$k \geq \frac{m+d}{2}$$

Define

$$L_0(x_1,...,x_k) = x_1 + ... + x_k \ (= S_k^1(x_1,...,x_k))\ .$$

Let V be the vector space:

$$V = \{\ v \in \mathbb{C}^k \mid L_i(v) = 0 \ \ \forall\, 0 \leq i \leq m-k\ \}.$$

According to our assumption, the dimension of V is at least

$$dim(V) \geq k - (m-k+1) \geq d-1.$$

According to Lemma 3.5.8 there is a vector $v \in V$ s.t, w.l.o.g. , its first $d-1$ coordinates are 1. Since $L_i(v) = 0\ \ \forall 0 \leq i \leq m-k$, proposition 3.5.6 gives

$$\partial_v\left(S_k^1(x_1,...,x_k)) = \partial_v(L_0(x_1,...,x_k)\right) = L_0(v) = 0\ ,$$

$$\text{and}\ \ \partial_v\left(S_{m-k}^l(L_1,...,L_{m-k})\right) = 0 \ \ \forall 1 \leq l \leq d\ .$$

Therefore by applying ∂_v to both sides of Equation 3.5 we get by proposition 3.5.6 that

$$0 = \partial_v\left(S_k^d(x_1,...,x_k)\right) +$$

$$+\sum_{i=2}^{d-1} \partial_v\left(S_k^i(x_1,...,x_k)\right)\cdot S_{m-k}^{d-i}(L_1,...,L_{m-k})\ . \tag{3.6}$$

Let U be the subspace

$$U = \{\ (x_1,...,x_k) \in V \mid x_d = x_{d+1} = ... = x_k = 0\}.$$

According to Lemma 3.5.7 and proposition 3.5.3 we have that for every $r \leq d$

$$\partial_v\left(S_k^r(x_1,...,x_k)\right)|_U = (1-r)S_{d-1}^{r-1}(x_1,...,x_{d-1})\ . \tag{3.7}$$

Therefore by restricting both sides of Equation 3.6 to U we get:

$$0 = \partial_v\left(S_k^d(x_1,...,x_k)\right)|_U +$$

$$+\sum_{i=2}^{d-1}\left(\partial_v\left(S_k^i(x_1,...,x_k)\right)\cdot S_{m-k}^{d-i}(L_1,...,L_{m-k})\right)|_U\ ,$$

and according to equation 3.7 this is equal to

$$0 = (1-d)S_{d-1}^{d-1}(x_1, ..., x_{d-1}) +$$

$$+ \sum_{i=2}^{d-1} (1-i) S_{d-1}^{i-1}(x_1, ..., x_{d-1}) \cdot (S_{m-k}^{d-i}(L_1, ..., L_{m-k})|_U) \; .$$

We now apply Ψ to both sides of the equation

$$0 = \Psi((1-d)S_{d-1}^{d-1}(x_1, ..., x_{d-1}) +$$

$$+ \sum_{i=2}^{d-1} (1-i) S_{d-1}^{i-1}(x_1, ..., x_{d-1}) \cdot (S_{m-k}^{d-i}(L_1, ..., L_{m-k})|_U))$$

$$= (1-d)S_{d-1}^{d-1}(x_1, ..., x_{d-1}) +$$

$$\sum_{i=2}^{d-1} (1-i) S_{d-1}^{i-1}(x_1, ..., x_{d-1}) \cdot \Psi\left(S_{m-k}^{d-i}(L_1, ..., L_{m-k})|_U\right) .$$

For $1 \leq i \leq d-2$, $\Psi\left(S_{m-k}^{i}(L_1, ..., L_{m-k})|_U\right)$ is a symmetric polynomial of degree at most $d-2$ in $x_1, ..., x_{d-1}$ and therefore can be written as a polynomial in $S_{d-1}^{1}(x_1, ..., x_{d-1}), ..., S_{d-1}^{d-2}(x_1, ..., x_{d-1})$. So we have

$$0 = (1-d)S_{d-1}^{d-1} + Q(S_{d-1}^{1}, ..., S_{d-1}^{d-2}),$$

for some polynomial Q. But, according to theorem 3.5.1, $S_{d-1}^{1}, ..., S_{d-1}^{d-2}, S_{d-1}^{d-1}$ are algebraically independent, and $(1-d)$, which is the coefficient of S_{d-1}^{d-1}, is not 0, so we have a contradiction. Thus $k < \frac{m+d}{2}$. □

3.5.5 Tightness of results

In this subsection we show that theorems 3.4.3, 3.4.4 are tight in some cases.

Claim 3.5.9 *For $d \geq \frac{m+1}{2}$ the result of Theorem 3.4.3 is tight.*

Proof: According to theorem 3.4.3 the smallest dimension of an affine subspace A such that $S_m^d|_A = 0$ is at most $\max(m-d, d-1) = d-1$ (since $d \geq \frac{m+1}{2}$). Let A be the subspace of $\mathbb{C}^m$ defined by

$$A = \{ \ (x_1, \ldots, x_m) \ \mid \ x_d = x_{d+1} = \ldots = x_m \ \} \; .$$

Since every monomial of S_m^d contains at least one variable of $x_d, ..., x_m$ we clearly have $S_m^d|_A = 0$. Since $dim(A) = d-1$ we are done. □

Claim 3.5.10 *For $d = 2$ the result of theorem 3.4.4 is tight (in [35] the exact behavior of S_m^2 is determined).*

Proof: Since $d = 2$, Theorem 3.4.4 implies that the co-dimension of every subspace B such that $S_m^2|_B = 0$, is at least $\frac{m}{2}$. To prove a lower bound on the co-dimension write S_m^2 as

$$S_m^2 = x_m(\sum_{i=1}^{m-1} x_i) + \frac{1}{2}((\sum_{i=1}^{m-1} x_i)^2 - \sum_{i=1}^{m-1} {x_i}^2).$$

When we restrict our polynomial to

$$A = \{\ (x_1, ..., x_m) \mid \sum_{i=1}^{m-1} x_i = 0\ \}$$

we get:

$$S_m^2|_A = \frac{1}{2}(-\sum_{i=1}^{m-2} {x_i}^2 - (\sum_{i=1}^{m-2} x_i)^2)$$

We now have two cases, m odd and m even.

- m even. We restrict our polynomial to:

  $$\tilde{A} = \{\ (x_1, ..., x_m) \in A \mid \sum_{l=1}^{m-2} x_l = 0, \text{ and}$$

  $$x_{2j-1} = i \cdot x_{2j},\ \ j = 1, ..., \frac{m-2}{2}\ \}$$

  $(i^2 = -1)$, and we get

  $$S_m^2|_{\tilde{A}} = 0.$$

  Since we had $2 + \frac{m-2}{2}$ restrictions, we see that the dimension of $\tilde{A}$ is $\frac{m}{2} - 1$. Hence theorem 3.4.4 is tight for $d = 2$ and m even.

- m odd. In a similar way we get that the dimension of A is at least $\lceil \frac{m}{2} \rceil + 1$.

□

As one can see, combining theorems 3.4.3, 3.4.4 with lemma 3.4.2 we can only get lower bounds of the form $s_{\text{sym}}(f) \geq 2n$ for polynomials in n variables. Proving a better lower bounds using lemma 3.4.2 seems a difficult task because of the next theorem.

Theorem 3.5.11 *[42] There exists a linear subspace A of dimension $\lfloor m/l \rfloor$ s.t $S_m^d|_A = 0$, where l is the smallest integer such that $l \not| d$.*

Therefore, for odd values of d we have a space A of dimension $\lfloor m/2 \rfloor$ s.t $S_m^d|_A = 0$. So lemma 3.4.2 cannot give a lower bound better than $2n$ for polynomials of odd degree in n variables.

The subspace A, of theorem 3.5.11, is constructed in such a way that in the representation of S_m^d as a polynomial in $\{T_m^i\}$, every monomial of this representation, i.e. every product of the form

$$\prod_i T_m^{d_i} \ \ where \ \ \sum d_i = d \ ,$$

is equal to zero when restricted to A.

Proof of theorem 3.5.11 Let w be a l-primitive root of unity. Take

$$A = \{(x_1, wx_1, w^2x_1, ..., w^{l-1}x_1, x_2, wx_2, \dots$$

$$\dots, w^{l-1}x_{\lfloor m/l \rfloor}, 0, ..., 0) \mid x_i \in \mathbb{C}\}.$$

Notice that if $l \not| \, k$ then $T_m^k|_A = 0$. Writing S_m^d using the Newton identities, we see that in every product $\prod_{k=1}^d (T_m^k)^{m_k}$ (s.t $\sum_{k=1}^d km_k = d$) there is some k with $l \not| \, k$ and $m_k \neq 0$. Therefore every such product is zero when restricted to A. Hence $S_m^d|_A = 0$. $\square$

Thus for all odd d the function $g(m, d)$ that appears in lemma 3.4.2 is at most $\frac{m}{2}$. Therefore one must consider the arithmetic properties of the degree, d, when trying to answer problem 3.1.1.

Chapter 4

Lower Bounds for Matrix Product, in Bounded Depth Circuits with Arbitrary Gates

4.1 Introduction

Exponential lower bounds are well known for constant depth Boolean circuits over the base $\{\text{AND}, \text{OR}, \text{NOT}\}$ [1, 15, 61, 21]. However, for many other types of constant depth circuits almost nothing is known. We prove super-linear lower bounds for the number of edges in constant depth circuits with n inputs and up to n outputs. Our lower bounds are proved for all models of Boolean and arithmetic circuits and, in particular, for Boolean circuits with arbitrary gates. The bounds apply for several explicit functions and, in particular, for matrix product.

In general, our lower bound for circuits of depth $d \geq 2$ is $\Omega(n \cdot \lambda_d(n))$, where $\lambda_d(n)$ is a slowly growing function. A description of the functions $\lambda_d(n)$ is given below in Subsection 4.1.5. Our main method is a graph theoretic argument that analyzes certain superconcentration properties of the circuit as a graph. Hence, the same lower bounds are obtained for all types of circuits. Our results and proof methods are related to the works of [13, 34], where lower bounds of $\Omega(n \cdot \lambda_d(n))$ were proved for the size of superconcentrators. Pudlak used similar methods to prove lower bounds of $\Omega(n \cdot \lambda_d(n))$ for the number of edges in constant depth arithmetic circuits with n inputs and up to n outputs over fields with characteristic 0 [34]. Pudlak's results hold for the parallel prefix problem as well as for other explicit functions.

In all that comes below, the size of a circuit means the number of edges in it.

4.1.1 Matrix Product

Matrix product is among the most studied computational problems. Surprising upper bounds of $O(m^{2+\epsilon})$ (where $\epsilon < 1$, and $m \times m$ is the size of each matrix) were obtained by Strassen in [51], and improved in many other works (see [16] for a survey). The best known lower bound, however, is a lower bound of $3 \cdot m^2$ for the number of products needed [48] (which we show in chapter 5). In particular, the following problem is still open: Can matrix product be computed by circuits of size $O(m^2)$? Non-trivial size-depth tradeoffs for matrix product are also not known. In particular, the following problem is still open: Can matrix product be computed by circuits of size $O(m^2)$ and logarithmic depth ? We prove that matrix product cannot be computed by circuits of size $O(m^2)$ and constant depth.

In the case of bilinear circuit our lower bound proof works also if the circuit gets for free any product of two linear functions. That is, the lower bound is proven for the number of edges above the product gates. We prove that if the circuit is of depth 1 above these products (i.e., total depth 3) it is of size $\Omega(m^3)$. For $d \geq 2$, we prove that if the circuit is of depth d above the products (i.e., total depth $d+2$) it is of size $\Omega(m^2 \cdot \lambda_d(m))$. In the general (non-bilinear) case for fields of characteristic different than 0, our lower bound is $\Omega(m^3)$ for circuits of depth 1 and $\Omega(m^2 \cdot \lambda_d(m))$ for circuits of depth $d \geq 2$. The last lower bound is a special case of a more general lower bound for circuits with arbitrary gates over finite fields. That lower bound is discussed in Subsection 4.1.2.

4.1.2 Circuits with Arbitrary Gates

In a *Boolean circuit with arbitrary gates*, we allow each gate (of fan-in k) to compute an arbitrary function $g : \{0,1\}^k \to \{0,1\}$. We give explicit examples for (up to) n Boolean functions $f_1, ..., f_n$, such that any constant depth Boolean circuit with arbitrary gates for $f_1, ..., f_n$ is of super-linear size (as before, the bound for depth $d \geq 2$ is $\Omega(n \cdot \lambda_d(n))$). The bound holds for matrix product over $GF[2]$ (where the dimension of each matrix is $m = \sqrt{n/2}$), as well as for matrix product over other finite fields (where, say, each field element is represented by its bits). The bound also holds for the parallel prefix problem and for other problems from [34].

One important special case of this model is *constant depth threshold circuits*. Note, however, that for constant depth threshold circuits super-linear lower bounds are already known [22].

The above results for Boolean circuits with arbitrary gates can be generalized to circuits over larger domains. Let F be some fixed finite set (e.g., some fixed finite field), and assume that the input variables range over F (or over a subset of F). A *circuit with arbitrary gates over F* allows each gate (of fan-in k) to compute an arbitrary function $g : F^k \to F$. By a reduction to the Boolean case, we get explicit examples for (up to) n functions $f_1, ..., f_n$, such that any constant depth circuit with arbitrary gates over F for $f_1, ..., f_n$ is of super-linear size (as before, the bound for depth $d \geq 2$ is $\Omega(n \cdot \lambda_d(n))$). In particular, this gives lower bounds for circuits with arbitrary gates over any finite field F. The bound holds for matrix product over F as well as for many other functions.

4.1.3 Arithmetic Circuits

Lower bounds for the size of arithmetic circuits for explicit polynomials are known only if we allow polynomials with large degree or large coefficients (e.g., [52, 3]). However, if we limit the degree and the coefficients to be of size $O(1)$ then no non-trivial lower bound is known. For constant depth arithmetic circuits, exponential lower bounds are known for fields F with characteristic $p = 2$ [39, 50]. For other finite characteristics, exponential lower bounds are known only for depth 3 [18, 19] (and for depth larger than 3 no non-trivial lower bound was known). For characteristic 0, the best lower bounds for depth 3 are the quadratic bounds of $\Omega(n^2)$ [49, 47].

We get (for any field F) explicit examples for (up to) n polynomials $f_1, ..., f_n$ such that any constant depth arithmetic circuit (over F) for $f_1, ..., f_n$ is of super-linear size. One such example is matrix product (over F). For finite fields (and hence also for any field with characteristic different than 0), this follows by the general lower bound for circuits with arbitrary gates over F, as discussed in Subsection 4.1.2. For fields with characteristic 0, this follows from the bilinear lower bound for matrix product, as discussed in Subsection 4.1.1. Similar bounds for fields with characteristic 0 were previously proved by Pudlak [34]. Pudlak gives explicit examples for n linear functions $f_1, ..., f_n$ such that any constant depth arithmetic circuit with linear gates (i.e., products are not allowed) for $f_1, ..., f_n$ is of super-linear size. (Over fields with characteristic 0, the assumption that all the gates in the circuit are linear can be made w.l.o.g.)

For fields with characteristic 0, our results (as well as Pudlak's results) also give explicit examples for **one** polynomial $h = f_1 \cdot y_1 + \cdots + f_n \cdot y_n$ (in the input variables $x_1, ..., x_n, y_1, ..., y_n$) such that any constant depth arithmetic circuit for h is of super-linear size. This follows easily by the result of [3] and was noted to us by Toni Pitassi and Avi Wigderson.

4.1.4 Methods and Related Work

Our main lemma gives an analysis of the structure of a constant depth circuit as a graph. Let G be a directed acyclic graph. Denote by V_G the set of all nodes of G. Denote by I_G the set of all nodes of indegree 0 (inputs), and by O_G the set of all nodes of outdegree 0 (outputs). The depth of G is the length of the longest directed path in G. Roughly speaking, the main lemma shows that if G is of depth d and has less than $n \cdot \lambda_d(n)$ edges then one can remove from G a set of $\epsilon \cdot n$ inputs and $\epsilon \cdot n$ outputs (for some small constant ϵ) and a small number of intermediate nodes, such that in the new graph the total number of directed paths from I_G to O_G is small.

Lemma 4.1.1 *For any $0 < \epsilon < 1/400$ and any directed acyclic graph G of depth d, with more than n vertices and less than $\epsilon \cdot n \cdot \lambda_d(n)$ edges, the following is satisfied:*

For some k, s.t., $\sqrt{n} \leq k = o(n)$, there exist subsets $I \subset I_G$, $O \subset O_G$ and $V \subset V_G$, s.t., $|I|, |O| \leq 5\epsilon \cdot d \cdot n$, and $|V| = k$, and such that the total number of directed paths from $I_G \setminus I$ to $O_G \setminus O$, that do not pass through nodes in V, is at most $\epsilon \cdot n^2/k$.

Lemma 4.1.1 is restated (in a slightly more general form) as Corollary 4.3.12.

The main lemma is used to transform any circuit of depth d and size less than $\epsilon n \lambda_d(n)$ into a new circuit of depth 1 (and relatively small size). This is done by removing from the original circuit $5\epsilon dn$ inputs, $5\epsilon dn$ outputs, and a small number of intermediate nodes. The lower bounds then follow by a *rigidity argument*, in the spirit of Valiant's approach [56].

As mentioned before, similar methods were previously used to prove lower bounds for superconcentrators [13, 34] and for constant depth arithmetic circuits over fields with characteristic 0 [34]. In particular, methods similar to our main lemma are implicit in [34] (although the presentation there is different). Versions of these methods appeared already in [13]. One can think of Lemma 4.1.1 also as a generalization of the lower bounds for superconcentrators given in [13, 34]. In fact, all these lower bounds follow easily by a reduction to Lemma 4.1.1. Our proof for Lemma 4.1.1 relies on [34].

4.1.5 The Functions $\lambda_d(n)$

We will now give a description of the functions $\lambda_d(n)$. The exact definition is given below in Section 4.2. Let $\theta(n)$ be any function on the natural numbers, such that for any $k > 0$, $\theta(k) < k$. We denote by $\theta^*(n)$ the following function: $\theta^*(n) = \min\{i \mid \theta^{(i)}(n) \leq 1\}$, that is, the smallest integer i such that i-times iteration of θ gives a value smaller than 1. The functions $\lambda_d(n)$ will satisfy:

1. $\lambda_2(n) = \Theta(\log n)$,
2. $\lambda_3(n) = \Theta(\log \log n)$,
3. for any even $d > 2$, $\lambda_{d+1}(n), \lambda_d(n) = \Theta(\lambda^*_{d-2}(n))$
   (e.g., $\lambda_4(n), \lambda_5(n) = \Theta(\log^*(n))$).

4.1.6 Organization

In Section 4.2 we give the definition of the functions $\lambda_d(n)$ and prove some simple properties of these functions. In Section 4.3 we give the proof of Lemma 4.1.1. In Section 4.4 we prove our results for bilinear arithmetic circuits. In Section 4.5 we prove our results for Boolean circuits with arbitrary gates and for circuits with arbitrary gates over finite fields.

4.2 Slowly Growing Functions

In this section, we define the functions $\lambda_d(n)$ and we prove some easy properties of them. We start with a definition of the "star" operator.

Definition 4.2.1 *For a function f, define $f^{(i)}$ to be the composition of f with itself i times, i.e., $f^{(i)} = f \circ f \circ ... \circ f$ i times, $f^{(1)} = f$.*
For a function $f : N \to N$ such that $f(n) < n$ for $n > 0$, define:

$$f^*(n) = \min\{\, i \mid \, s.t., \;\; f^{(i)}(n) \leq 1\} \ .$$

We will need the following properties of f^* (taken from [34]).

Claim 4.2.2 *[34] Suppose $f(n) \leq \lfloor \sqrt{n} \rfloor$. For every $n \geq 0$ we have:*

1. *$\frac{f^{(i)}(n)}{f^{(i+1)}(n)} \geq f^{(i+1)}(n)$, for every $i > 0$, (provided that the denominator is not 0).*
2. *$f^{(i)}(n) \geq \frac{f^*(n)}{2}$, for every $i \leq \frac{f^*(n)}{2}$.*

The proof is taken from [34] as well.

Proof:

1. $$\frac{f^{(i)}(n)}{f^{(i+1)}(n)} \geq \frac{f^{(i)}(n)}{\lfloor \sqrt{f^{(i)}(n)} \rfloor} \geq \sqrt{f^{(i)}(n)} \geq f^{(i+1)} \ .$$

2. From (1) it follows that if $f^{(i+1)}(n) > 1$ then $f^{(i)}(n) > f^{(i+1)}(n)$. Therefore

$$f(n) > f^{(2)}(n) > ... > f^{(f^*(n))}(n) \ .$$

Since the values of f are integers, the result follows.

□

Our lower bounds will be expressed in terms of the following set of slowly growing functions.

Definition 4.2.3 *Let*

$$\lambda_1(n) = \lfloor \sqrt{n} \rfloor \ ,$$

$$\lambda_2(n) = \lceil \log n \rceil \ ,$$

$$\lambda_d(n) = \lambda^*_{d-2}(n) \ .$$

Some easy to verify properties of these functions are:

Claim 4.2.4 *1. Each $\lambda_i(n)$ is a monotone increasing function tending to infinity with n.*

2. for $i \geq 2$, $\lambda_{2i}(n) = \theta(\lambda_{2i+1}(n))$.

3. for $i \geq 2$ and n large enough, $\lambda_i(n) \leq \lfloor\sqrt{\frac{n}{2}}\rfloor$.

Proof:

1. The fact that λ_i is increasing is immediate from the fact that λ_1 and λ_2 are.

2. Notice that $\lambda_3(n) = \theta(\log\log n)$. Since $\log\log n = \log^{(2)}(n)$, we have $\lambda_4(n) = \theta(\lambda_5(n))$. Using induction we get the desired result.

3. Clearly $\lambda_2(n), \lambda_3(n) \leq \sqrt{\frac{n}{2}}$, for n large enough. Assume that $\lambda_j(n) \leq \sqrt{\frac{n}{2}}$. We have
$$\sqrt{\frac{n}{2}} \geq \lambda_j(n) \geq (\lambda_j^{(2)}(n))^2 \geq \frac{1}{4}(\lambda_{j+2}(n))^2 ,$$
hence
$$\lambda_{j+2}(n) \leq (8n)^{\frac{1}{4}} \leq \sqrt{\frac{n}{2}}$$
for n large enough.

□

4.3 Superconcentration Properties of Graphs

In this section we prove our main lemma on graphs, and several stronger versions of it. The lemma will be used to analyze the structure of a constant depth circuit as a graph. For simplicity, we prove here the lemma for leveled graphs. The general case follows easily by a reduction to the leveled case. Let $G = (V_G, E_G)$ be a leveled graph of depth d. The number of levels in G is hence $d+1$ and all edges in the graph are between vertices of adjacent levels. In all the following we allow all graphs to be multi-graphs.

We will use the following notations: We denote by $L_0, ..., L_d$ the levels of G, that is, L_i is the set of vertices at level i. The set of vertices L_0 is also denoted by I_G (and we call these vertices *inputs*). The set of vertices L_d is also denoted by O_G (and we call these vertices *outputs*).

Let $U \subset V_G$ be a set of vertices. We denote by $E(U)$ the set of edges that touch vertices in U. We denote by $\Gamma(U)$ the set of neighbors of U. We denote by $maxdeg(U)$ the maximal degree of a vertex in U. For subsets $I \subset I_G$, $O \subset O_G$, $V \subset V_G$ we denote by $P_G[I, O, V]$ the total number of paths of length d between I and O that do not pass through vertices in V.

4.3.1 Depth 2

We will first prove the main lemma for graphs of depth 2. We prove first a stronger lemma (Lemma 4.3.1). The main lemma (for depth 2) will then follow as Corollary 4.3.3. These lemmas are not needed for the proofs for higher depth. Nevertheless, the methods used here give some hints for the proofs needed for the general case.

Lemma 4.3.1 *Let G be a leveled graph of depth 2 with at most $\epsilon n \lambda_2(\frac{n}{k})$ edges, for some $1 \leq k = o(n)$, and some $\epsilon > 0$. Assume that $|L_1| \geq k$. Then, there exists a set $V \subset L_1$ of size $k \leq |V| = o(n)$ such that*

$$P_G[I_G, O_G, V] \leq \frac{100\epsilon^2 n^2}{|V|} \quad .$$

For the proof of Lemma 4.3.1 we will need the following Lemma 4.3.2. For the proof of Lemma 4.3.2, the reader is referred to [34] (Lemma 4).

Lemma 4.3.2 *[34] Let $c_1 \geq c_2 \geq ... \geq c_t \geq 0$ be a sequence of real numbers, and let p, q be two integers such that $1 \leq p \leq q \leq t$. If for every $p \leq l \leq q$,*

$$\sum_{i=l}^{t} {c_i}^2 \geq \frac{1}{l} \quad ,$$

then

$$\sum_{i=1}^{t} c_i \geq \frac{1}{2} \log \left(\frac{q}{p} \right) \quad .$$

Proof of Lemma 4.3.1:

Denote $m = |L_1|$. Let $v_1, v_2, ..., v_m$ be the vertices of L_1, ordered according to their degree, from highest to lowest. That is, for every i, $deg(v_i) \geq deg(v_{i+1})$. For every $k \leq l \leq \sqrt{nk}$, denote $V_l = \{v_1, ..., v_l\}$. Then, for every such l,

$$P_G[I_G, O_G, V_l] \leq \sum_{i=l+1}^{m} (deg(v_i))^2 \quad .$$

Let $c_i = \frac{deg(v_i)}{10\epsilon n}$. Then,

$$\sum_{i=1}^{m} c_i = \frac{1}{10\epsilon n} \sum_{i=1}^{m} deg(v_i) = \frac{1}{5\epsilon n}|E_G| \leq \frac{1}{5} \left\lceil \log \left(\frac{n}{k} \right) \right\rceil < \frac{1}{2} \log \left(\frac{\sqrt{n}}{\sqrt{k}} \right) = \frac{1}{2} \log \left(\frac{\sqrt{nk}}{k} \right) \quad .$$

Therefore, by Lemma 4.3.2, for some $k \leq l \leq \sqrt{nk} = o(n)$,

$$\sum_{i=l}^{m} {c_i}^2 < \frac{1}{l} \quad .$$

Hence,

$$P_G[I_G, O_G, V_l] \leq \sum_{i=l}^{m} (deg(v_i))^2 = (100\epsilon^2 n^2) \sum_{i=l}^{m} c_i^2 \leq \frac{100\epsilon^2 n^2}{l} = \frac{100\epsilon^2 n^2}{|V_l|} \ .$$

□

As a corollary, we obtain our main lemma for $d = 2$.

Corollary 4.3.3 *Let G be a leveled graph of depth 2 with at most $\epsilon n \lambda_2(n)$ edges, for some $0 < \epsilon < 1/400$. Assume that $|L_1| \geq \sqrt{n}$. Then there exists a set $V \subset L_1$ of size $\sqrt{n} \leq |V| = o(n)$ such that*

$$P_G[I_G, O_G, V] \leq \frac{\epsilon n^2}{|V|}.$$

Proof: Note that

$$\epsilon n \lambda_2(n) \leq 2\epsilon n \lambda_2\left(\frac{n}{\sqrt{n}}\right) \ .$$

By Lemma 4.3.1, there is a set $V \subset L_1$, of size $\sqrt{n} \leq |V| = o(n)$ such that,

$$P_G[I_G, O_G, V] \leq \frac{100(2\epsilon)^2 n^2}{|V|} \leq \frac{\epsilon n^2}{|V|} \ .$$

□

4.3.2 Reducing the depth by 2

Roughly speaking, the main lemma shows that if G has a small number of edges, then one can remove from G a small set J of inputs and outputs, and a small set V of other vertices, such that the total number of paths of length d from $I_G \setminus J$ to $O_G \setminus J$, that do not pass through vertices in V, is small (i.e., $P_G[\ I_G \setminus J\ ,\ O_G \setminus J\ ,\ V\ ]$ is small).

The idea of the proof for depth larger than 2 is as follows. We start with a graph G of depth d and we reduce it's depth by steps. In each step we will reduce the depth of G from d to $d-2$ by eliminating the levels L_1 and L_{d-1}. This is done in the following way. We partition $L_1 \cup L_{d-1}$ into 3 sets (A, B, C). We will then eliminate each one of these sets of vertices. However, we will have a different method to eliminate each one of these sets. The vertices in A are just removed from the graph. More formally, we just add the vertices of A to the set of vertices V (that are later on ignored). The set B is eliminated by removing all the inputs and outputs that are connected to it. More formally, we just add all the inputs and outputs that are connected to vertices in B to the set J of inputs and outputs (that are later on ignored). The set C is eliminated by connecting directly every pair of vertices that are connected by a path of length 2 through C. More formally, we replace each path of length 2 (v, c, w), such that $v \in I_G$, $c \in C$ and $w \in L_2$, by the one edge

(v, w) (the same with L_{d-2} and O_G). The set C can hence be ignored. We continue to reduce the depth of G until we have a graph of depth 2 or 3 (w.l.o.g. 3). We then finish by proving the main lemma for depth 3.

To summarize, the proof is by induction on the depth. The proof uses a specific way to reduce the depth by 2. Given a partition (A, B, C) of $L_1 \cup L_{d-1}$, we will eliminate $L_1 \cup L_{d-1}$ by:

1. Removing A.

2. Removing all neighbors of B among $I_G \cup O_G$.

3. Adding an edge between vertices from I_G and L_2 that are connected through C (and the same with L_{d-2} and O_G) and then removing C.

This leads us to the following definition.

Definition 4.3.4 *Let G be a leveled graph of depth $d \geq 3$. Let (A, B, C) be a partition of $L_1 \cup L_{d-1}$. The graph $\hat{G}$ of depth $d-2$ is defined in the following way:*
The inputs of $\hat{G}$ are

$$I_{\hat{G}} = I_G \setminus \Gamma(B).$$

The outputs of $\hat{G}$ are

$$O_{\hat{G}} = O_G \setminus \Gamma(B).$$

The $d-1$ levels of $\hat{G}$ are

$$I_{\hat{G}}, L_2, ..., L_{d-2}, O_{\hat{G}}$$

(note that for $d = 3$ the levels of $\hat{G}$ are just $I_{\hat{G}}, O_{\hat{G}}$). The edges between the levels $L_2, ..., L_{d-2}$ are the same as they are in G. The other edges are defined in the following way:

- *For $d > 3$, we have to define the edges between $I_{\hat{G}}$ and L_2 and between L_{d-2} and $O_{\hat{G}}$. For every v, c, w such that $v \in I_{\hat{G}}$, $c \in C$, $w \in L_2$, and there are edges $(v, c), (c, w)$ in G, we add the edge (v, w) to $\hat{G}$ (i.e., we replace every such path of length 2 with an edge). In the same way, for every v, c, w such that $v \in L_{d-2}$, $c \in C$, $w \in O_{\hat{G}}$, and there are edges $(v, c), (c, w)$ in G, we add the edge (v, w) to $\hat{G}$.*

- *For $d = 3$, we have to define the edges between $I_{\hat{G}}$ and $O_{\hat{G}}$. This case is slightly different because we don't have a level between L_1 and L_2 to absorb the new edges. So instead, for every path of length 3 (v, c_1, c_2, w), such that, $v \in I_{\hat{G}}$, $w \in O_{\hat{G}}$ and $c_1, c_2 \in C$, we put an edge (v, w) in $\hat{G}$.*

Clearly $\hat{G}$ is a function of G, A, B, C. For convenience we will use the notation $\hat{G}$ instead of $\hat{G}(G, A, B, C)$.

Obviously, $\hat{G}$ is a multi-graph of depth $d-2$. In the construction of $\hat{G}$ we replaced each path of length 2 through C (or a path of length 3 in the case $d = 3$) with an edge. Therefore, we have the following easy corollary. The corollary shows that the number of relevant paths in the graph $\hat{G}$ is the same as it is in the graph G after removing the sets A and $\Gamma(B)$. Hence, in order to count paths in G it is enough to count the corresponding paths in $\hat{G}$. This easy observation makes the use of induction possible.

Proposition 4.3.5 *Let $G, A, B, C, \hat{G}$ be as in Definition 4.3.4. Then,*

1. $P_G[\, I_G \setminus \Gamma(B) \ ,\ O_G \setminus \Gamma(B) \ ,\ A \,] = P_{\hat{G}}[\, I_{\hat{G}} \ ,\ O_{\hat{G}} \ ,\ \emptyset \,]$.
2. *More generally: for any set of vertices $\hat{V} \subset L_2 \cup \cdots \cup L_{d-2}$, and any set of inputs and outputs $\hat{J} \subset I_{\hat{G}} \cup O_{\hat{G}}$,*

$$P_G[\, I_G \setminus (\hat{J} \cup \Gamma(B)) \ ,\ O_G \setminus (\hat{J} \cup \Gamma(B)) \ ,\ \hat{V} \cup A \,] = P_{\hat{G}}[\, I_{\hat{G}} \setminus \hat{J} \ ,\ O_{\hat{G}} \setminus \hat{J} \ ,\hat{V} \,] \ .$$

As mentioned above we will prove the main lemma by induction. In each step we reduce the depth by 2 by removing a set A of intermediate vertices and a set $\Gamma(B)$ of inputs and outputs. The final set V will be the union of the sets A from all steps of the induction, and the final set J will be the union of the sets $\Gamma(B)$ from all these steps.

In each step of the induction, we assume that the graph G has a relatively small number of edges. We would like to make sure that $\hat{G}$ also has a small number of edges. The next lemma shows that given certain bounds for $|A|$, $maxdeg(C)$, and for $|E_G|$, one can bound the number of edges in $\hat{G}$. The idea of the proof is that the only edges that we add to $\hat{G}$ are related to paths through C. Therefore (roughly speaking) we can bound $|E_{\hat{G}}|$ by $|E_G| \cdot maxdeg(C)$. Note that we assume here the bound

$$|E_G| \leq \epsilon n \lambda_d \left(\frac{n}{k}\right)$$

for some $k \geq 1$. This is more general than the original assumption $|E_G| \leq \epsilon n \lambda_d(n)$. The reason that we need a more general assumption is that the graph G may be a graph that was obtained after several steps of reduction (rather than the original graph). Roughly speaking, the parameter k corresponds to the number of intermediate vertices that were already removed from the original graph (the set V in the statement of the lemma). Since, we think of the set A as being removed from the graph as well (and being added to the set V) the bound that we want for $|E_{\hat{G}}|$ is

$$|E_{\hat{G}}| \leq \epsilon n \lambda_{d-2} \left(\frac{n}{k + |A|}\right).$$

Lemma 4.3.6 *Let $G, A, B, C, \hat{G}$ be as in Definition 4.3.4 and assume that for some $1 \le k = o(n)$ and for some $0 < \epsilon < 1/3$,*

$$|E_G| \le \epsilon n \lambda_d \left(\frac{n}{k}\right).$$

Assume that for some integer $1 \le i \le \frac{\lambda_d(\frac{n}{k})}{2} - 3$ we have:

1. $|A| \le \frac{2\epsilon n}{\lambda^{(i)}_{d-2}(\frac{n}{k})}$.
2. *$maxdeg(C) \le \lambda^{(i+3)}_{d-2}(\frac{n}{k})$.*

Then,

$$|E_{\hat{G}}| \le \epsilon n \lambda_{d-2}\left(\frac{n}{k+|A|}\right).$$

The proof is by a straight forward calculation using Claim 4.2.2, and Claim 4.2.4.

Proof: Since $k = o(n)$ we have:

$$k\lambda^{(i)}_{d-2}\left(\frac{n}{k}\right) \le k\lambda_1\left(\frac{n}{k}\right) \le k\sqrt{\frac{n}{k}} < \epsilon n \quad ,$$

therefore

$$k + |A| \le k + \frac{2\epsilon n}{\lambda^{(i)}_{d-2}(\frac{n}{k})} = \frac{k\lambda^{(i)}_{d-2}(\frac{n}{k}) + 2\epsilon n}{\lambda^{(i)}_{d-2}(\frac{n}{k})} < \frac{3\epsilon n}{\lambda^{(i)}_{d-2}(\frac{n}{k})} < \frac{n}{\lambda^{(i)}_{d-2}(\frac{n}{k})}.$$

We now have two different cases that follow from the construction in Definition 4.3.4:

1. $d > 3$: ¿From the construction of $\hat{G}$ we get that the degree of each vertex in $L_2 \cup L_{d-2}$ has increased by a factor of $maxdeg(C)$ at most. Therefore we have:

$$|E_{\hat{G}}| \le |E_G| \cdot maxdeg(C) \le \epsilon n \lambda_d\left(\frac{n}{k}\right)\lambda^{(i+3)}_{d-2}\left(\frac{n}{k}\right) \le$$

$$\le 2\epsilon n\left(\lambda^{(i+3)}_{d-2}\left(\frac{n}{k}\right)\right)^2 \le \epsilon n \lambda^{(i+2)}_{d-2}\left(\frac{n}{k}\right) \le \epsilon n \lambda^{(2)}_{d-2}\left(\frac{n}{k+|A|}\right) < \epsilon n \lambda_{d-2}\left(\frac{n}{k+|A|}\right)$$

(where all inequalities are due to Claim 4.2.2, Claim 4.2.4 and the bound that we proved on $k + |A|$).

2. $d = 3$: Since we dropped $L_1 \cup L_2$, the set of inputs and outputs now absorbs both the edges of $C \cap L_1$ and the edges of $C \cap L_2$, so we have:

$$|E_{\hat{G}}| \le |E_G| \cdot maxdeg(C)^2 \le \epsilon n \lambda_d\left(\frac{n}{k}\right)\left(\lambda^{(i+3)}_{d-2}\left(\frac{n}{k}\right)\right)^2 \le$$

$$\le \frac{1}{2}\epsilon n \lambda_d\left(\frac{n}{k}\right)\lambda^{(i+2)}_{d-2}\left(\frac{n}{k}\right) \le \frac{1}{2}\epsilon n \lambda^{(i+1)}_{d-2}\left(\frac{n}{k}\right) \le \frac{1}{2}\epsilon n \lambda_{d-2}\left(\frac{n}{k+|A|}\right)$$

(where, as before, all inequalities are due to Claim 4.2.2, Claim 4.2.4 and the bound that we proved on $k + |A|$).

□

So far, we have presented the construction of $\hat{G}$ from G, given an arbitrary partition (A, B, C). In order to maintain the bound on the number of edges of $\hat{G}$, we need A, C to satisfy the conditions of Lemma 4.3.6. Also, we need $\Gamma(B)$ to be not too large, in order to make sure that the total number of inputs and outputs removed in the process is small. The next lemma shows how to partition $L_1 \cup L_{d-1}$ into suitable sets (A, B, C). The way it is done is by first ordering the vertices in $L_1 \cup L_{d-1}$ according to their degrees (from highest to lowest), and then finding appropriate numbers $r_1 > r_2$, such that A will be the set of vertices with degree larger than r_1, B will be the set of vertices with degree larger than r_2 and at most r_1, and C will be the set of vertices with degree at most r_2.

Lemma 4.3.7 *Let G be a leveled graph of depth $d \geq 3$, such that, $|E_G| \leq \epsilon n \lambda_d(r)$ for large enough r (more accurately, we need $\lambda_d(r) > 72$). Then, there exists a partition (A, B, C) of $L_1 \cup L_{d-1}$, and $1 \leq i \leq \lambda_d(r)/2 - 3$ with the following properties:*

1. $|A| \leq \frac{2\epsilon n}{\lambda_{d-2}^{(i)}(r)}$.
2. $|\Gamma(B)| \leq 9\epsilon n$.
3. $maxdeg(C) \leq \lambda_{d-2}^{(i+3)}(r)$.

Proof: Denote,

$$W_0 = \{v \in L_1 \cup L_{d-1} \mid deg(v) > \lambda_{d-2}(r)\},$$

and for $i \geq 1$,

$$W_i = \{v \in L_1 \cup L_{d-1} \mid \lambda_{d-2}^{(i)}(r) \geq deg(v) > \lambda_{d-2}^{(i+1)}(r)\}.$$

Claim 4.3.8 *For every $1 \leq i \leq \lambda_d(r)/2 - 3$,*

$$|W_0 \cup W_1 \cup \cdots \cup W_{i-1}| \leq \frac{2\epsilon n}{\lambda_{d-2}^{(i+1)}(r)} .$$

Proof: The proof follows from the fact that the degree of each vertex in $W_0 \cup W_1 \cup \cdots \cup W_{i-1}$ is at least $\lambda_{d-2}^{(i)}(r)$. If the claim wasn't true we would have had

$$|E_G| \geq |E(W_0 \cup W_1 \cup \cdots \cup W_{i-1})| \lambda_{d-2}^{(i)}(r) \geq \frac{2\epsilon n}{\lambda_{d-2}^{(i+1)}(r)} \lambda_{d-2}^{(i)}(r) \geq$$

$$\geq 2\epsilon n \lambda_{d-2}^{(i+1)}(r) \geq \epsilon n \lambda_d(r) ,$$

in contradiction. □

Claim 4.3.9 *For some* $0 \leq i \leq \lambda_d(r)/2 - 4$,

$$|E(W_i \cup W_{i+1} \cup W_{i+2} \cup W_{i+3})| \leq 9\epsilon n.$$

Proof: The proof follows from the bound on $|E_G|$. Since $|E_G| \leq \epsilon n \lambda_d(r)$ we must have

$$\sum_{i=0}^{\frac{\lambda_d(r)/2-4}{4}} |E(W_{4i} \cup W_{4i+1} \cup W_{4i+2} \cup W_{4i+3})| \leq \epsilon n \lambda_d(r) \; .$$

If each of the sets $E(W_{4i} \cup W_{4i+1} \cup W_{4i+2} \cup W_{4i+3})$ was of size larger than $9\epsilon n$ then we would have had

$$|E_G| \geq 9\epsilon n \frac{\lambda_d(r)/2 - 4}{4} > \epsilon n \lambda_d(r)$$

(for large enough r), in contradiction. □

Fix i' to be such that Claim 4.3.9 is satisfied for i'. The proof of the lemma now follows for $i = i' + 1$, by taking

1. $A = W_0 \cup W_1 \cup \cdots \cup W_{i'-1}$.

2. $B = W_{i'} \cup W_{i'+1} \cup W_{i'+2} \cup W_{i'+3}$.

3. $C = (L_1 \cup L_{d-1}) \setminus (A \cup B)$.

□

4.3.3 Depth 3

The induction that we are about to perform on the depth will end with a graph of depth 2 or 3 (w.l.o.g., 3). We will now give the proof of the main lemma for the special case $d = 3$. As mentioned before, we need a more general lemma that assumes a more general bound for the number of edges in the graph.

The proof for $d = 3$ already gives the main idea of the proof for the general case. First we partition $L_1 \cup L_2$ into 3 sets (A, B, C) as described before. The partition (A, B, C) will satisfy the following: The first set A is small, the second set B is not connected to many inputs and outputs, and each vertex in the third set C has small degree. We then reduce the depth of the graph to 1, using Definition 4.3.4. Thus we get a graph, $\hat{G}$, with the same number of input-output paths, between $I_{\hat{G}}$ and $O_{\hat{G}}$. Since $\hat{G}$ is a graph of depth 1, the number of such paths is simply the number of edges. The last step would be to calculate the number of edges in $\hat{G}$, which is done using Lemma 4.3.6.

Lemma 4.3.10 *Let G be a leveled graph of depth 3 with at most $\epsilon n\lambda_3(\frac{n}{k})$ edges, for some $1 \le k = o(n)$ and for some $0 < \epsilon < 1/3$. Then, there exists a partition (A, B, C) of $L_1 \cup L_2$ such that:*

1. $|A| = o(n)$,
2. $|\Gamma(B)| \le 9\epsilon n$,
3. $P_G[\ I_G \setminus \Gamma(B)\ ,\ O_G \setminus \Gamma(B)\ ,\ A\ ] \le \epsilon n\lambda_1\left(\frac{n}{k+|A|}\right)$.

Proof: By Lemma 4.3.7, we can partition $L_1 \cup L_2$ into 3 sets (A, B, C), such that for some $1 \le i \le \lambda_3(\frac{n}{k})/2 - 3$, we have:

1. $|A| \le \frac{2\epsilon n}{\lambda_1^{(i)}(\frac{n}{k})} = o(n)$.
2. $|\Gamma(B)| \le 9\epsilon n$.
3. $maxdeg(C) \le \lambda_1^{(i+3)}(\frac{n}{k})$.

Let $\hat{G}$ be as in Definition 4.3.4 (with respect to (A, B, C)). By Proposition 4.3.5,

$$P_G[\ I_G \setminus \Gamma(B)\ ,\ O_G \setminus \Gamma(B)\ ,\ A\ ] = P_{\hat{G}}[\ I_{\hat{G}}\ ,\ O_{\hat{G}}\ ,\ \emptyset\ ].$$

Since $\hat{G}$ is of depth 1, the right hand-side equals $|E_{\hat{G}}|$. By Lemma 4.3.6,

$$|E_{\hat{G}}| \le \epsilon n\lambda_1\left(\frac{n}{k+|A|}\right) .$$

Hence,

$$P_G[\ I_G \setminus \Gamma(B)\ ,\ O_G \setminus \Gamma(B)\ ,\ A\ ] \le \epsilon n\lambda_1\left(\frac{n}{k+|A|}\right) .$$

□

4.3.4 Higher depth

We are now ready to state and prove our main lemma. As mentioned above, we actually prove a stronger lemma that will be needed for the induction. The main lemma will then follow as an easy corollary. First note that the functions $\lambda_d(n)$ satisfy $\lambda_{2i}(n) = \theta(\lambda_{2i+1}(n))$. Hence, we can assume w.l.o.g. that the depth d is odd, otherwise we can just increase the depth by 1. (We couldn't prove better results for the even levels because of the constructions given in [13]).

Lemma 4.3.11 *Let G be a leveled graph of constant depth $d \ge 3$, (and assume that d is odd), with at most $\epsilon n\lambda_d(\frac{n}{k})$ edges, for some $1 \le k = o(n)$ and for some $0 < \epsilon < 1/3$. Then there exist a set V of vertices and a set J of inputs and outputs, such that:*

1. $|V| = o(n)$,

2. $|J| \leq 5\epsilon dn$,

3. $P_G[\, I_G \setminus J \ , \ O_G \setminus J \ , \ V \,] \leq \epsilon n \lambda_1 \left(\frac{n}{k+|V|}\right)$.

Proof: The proof is by induction on d. The base case $d = 3$ was proved in Lemma 4.3.10. For $d \geq 5$, let (A, B, C) be the partition of $L_1 \cup L_{d-1}$ from Lemma 4.3.7. Let $\hat{G}$ be the depth $d-2$ graph defined in Definition 4.3.4 (with respect to (A, B, C)). Then by Lemma 4.3.6,

$$|E_{\hat{G}}| \leq \epsilon n \lambda_{d-2}\left(\frac{n}{k+|A|}\right) .$$

Since $|A| = o(n)$, we have that $k + |A| = o(n)$ and hence the inductive assumption holds for $\hat{G}$. Hence for $\hat{G}$, there exist a set $\hat{V}$ of vertices and a set $\hat{J}$ of inputs and outputs, such that:

1. $|\hat{V}| = o(n)$,

2. $|\hat{J}| \leq 5\epsilon(d-2)n$,

3. $P_{\hat{G}}[\, I_{\hat{G}} \setminus \hat{J} \ , \ O_{\hat{G}} \setminus \hat{J} \ , \ \hat{V} \,] \leq \epsilon n \lambda_1 \left(\frac{n}{k+|A|+|\hat{V}|}\right)$.

Define $V = \hat{V} \cup A$ and $J = \hat{J} \cup \Gamma(B)$. Then since $|\hat{V}|, |A| = o(n)$ we have $|V| = o(n)$ (note that since the original depth is constant, the total number of induction steps is constant). Since $|\hat{J}| \leq 5\epsilon(d-2)n$ and $|\Gamma(B)| \leq 9\epsilon n$ we have that $|J| \leq 5\epsilon dn$.

Finally, by Proposition 4.3.5,

$$P_G[\, I_G \setminus J \ , \ O_G \setminus J \ , \ V \,] = P_{\hat{G}}[\, I_{\hat{G}} \setminus \hat{J} \ , \ O_{\hat{G}} \setminus \hat{J} \ , \ \hat{V} \,] \leq$$

$$\leq \epsilon n \lambda_1 \left(\frac{n}{k+|A|+|\hat{V}|}\right) = \epsilon n \lambda_1 \left(\frac{n}{k+|V|}\right) .$$

□

Our main lemma is now stated as the following corollary. Note that the requirement $\epsilon < 1/400$ is only needed for the case $d = 2$ (a weaker requirement is needed for $d > 2$).

Corollary 4.3.12 *Let G be a leveled graph of constant depth $d \geq 2$, with more than n vertices and less than $\epsilon n \lambda_d(n)$ edges, for some $0 < \epsilon < 1/400$. Then there exist a set V of vertices and a set J of inputs and outputs, such that:*

1. $\sqrt{n} \leq |V| = o(n)$,

2. $|J| \leq 5\epsilon dn$,

3. $P_G[\ I_G \setminus J\ ,\ O_G \setminus J\ ,\ V\ ] \leq \epsilon \frac{n^2}{|V|}$.

Proof: First note that (as mentioned above) Lemma 4.3.11 is correct also for even depth if we just require ϵ to be slightly smaller (a factor of 2 or 3 is enough). Since here we require $\epsilon < 1/400$ we can apply Lemma 4.3.11 for any depth larger than 2.

We apply Lemma 4.3.11 with $k = 1$. If $|V| \geq \sqrt{n}$ we are done. Otherwise, just add arbitrary vertices to V. Call the resulting set V'. We have

$$P_G[\ I_G \setminus J\ ,\ O_G \setminus J\ ,\ V'\ ] \leq P_G[\ I_G \setminus J\ ,\ O_G \setminus J\ ,\ V\ ] \leq$$

$$\leq \epsilon n \lambda_1 \left(\frac{n}{1+|V|} \right) \leq \epsilon n \sqrt{\frac{n}{1+|V|}} = \epsilon \frac{n^2}{\sqrt{n(1+|V|)}} \leq \epsilon \frac{n^2}{|V'|} \ .$$

This completes the proof for depth higher than 2. For $d = 2$, the corollary was already stated as Corollary 4.3.3. □

4.3.5 Graphs for matrix product

As mentioned in the introduction, the main function that we concentrate on in this work is matrix product. A circuit for matrix product has $2m^2$ inputs and m^2 outputs, where m is the dimension of each matrix. Therefore, we will assume here that our graph has $2m^2$ inputs and m^2 outputs and that the outputs are ordered as a matrix. We will refer to such a graph as a graph for matrix product. For convenience, we will prove here a lemma that will be specific for such graphs. The proof will follow easily by Corollary 4.3.12.

Denote by O_i the outputs in the i^{th} column of the output matrix. Denote by $[m]$ the set $\{1, ..., m\}$. We think of $[m]$ as the set of all output columns. For a subset $D \subset [m]$, denote by O_D the outputs in all the columns in D. That is, $O_D = \cup_{i \in D} O_i$.

Roughly speaking, our lemma will state that after removing from the graph a small set of inputs and outputs (I and O) and a set V of size k of intermediate nodes, one can find a set D of $10k/m$ output columns, such that there are no paths between the inputs and the outputs in O_D. For simplicity, we will not state the lemma for a general constant ϵ and just fix some constant that will be good enough.

Lemma 4.3.13 *Let G be a leveled graph for matrix product, of constant depth $d \geq 2$, with less than $\epsilon m^2 \lambda_d(m^2)$ edges, for $\epsilon = 1/(1000 \cdot d)$. Then there exist sets $V \subset V_G$, $D \subset [m]$, $O \subset O_D$, $I \subset I_G$, such that:*

1. $m \leq |V| = o(m^2)$.

2. $|D| \geq \frac{10|V|}{m}$.

3. *For every* $i \in D$, $|O_i \cap O| < \frac{1}{10}m$.

4. $|I| < \frac{1}{10}m^2$.

5. $P_G[\, I_G \setminus I \,,\ O_D \setminus O \,,\ V \,] = 0$.

Proof: Let V, J be the sets guaranteed by Corollary 4.3.12 with $n = m^2$. Define $\tilde{I} = J \cap I_G$ and $\tilde{O} = J \cap O_G$, and let $k = |V|$. Then

1. $m \leq |V| = k = o(n)$. (hence, requirement (1) is satisfied).

2. $|\tilde{I}|, |\tilde{O}| \leq (1/200) \cdot m^2$.

3. $P_G[\, I_G \setminus \tilde{I} \,,\ O_G \setminus \tilde{O} \,,\ V \,] \leq \epsilon \frac{m^4}{k}$.

Denote by $\tilde{D}$ the set of all $i \in [m]$ such that $|O_i \cap \tilde{O}| < (1/100) \cdot m$. Then by the bound we have on $|\tilde{O}|$ we know that

$$|\tilde{D}| \geq (1/2) \cdot m.$$

For every $i \in \tilde{D}$, let $P(i)$ be the total number of paths between outputs in $O_i \setminus \tilde{O}$ and inputs in $I_G \setminus \tilde{I}$ that do not pass through V. Denote by D the set of $\lceil 10k/m \rceil$ indices i with the smallest $P(i)$. (then by the definition of D, requirement (2) is satisfied). Denote $O = \tilde{O} \cap O_D$. (then by the fact that $D \subset \tilde{D}$, requirement (3) is satisfied). Since

$$\sum_{i \in \tilde{D}} P(i) \leq P_G[\, I_G \setminus \tilde{I} \,,\ O_G \setminus \tilde{O} \,,\ V \,] \leq \epsilon \frac{m^4}{k},$$

we have,

$$\sum_{i \in D} P(i) \leq \frac{|D|}{|\tilde{D}|} \sum_{i \in \tilde{D}} P(i) \leq \frac{\lceil 10k/m \rceil}{m/2} \cdot \epsilon \frac{m^4}{k} < (1/50) \cdot m^2.$$

Hence, if we denote by $\hat{I}$ the set of all inputs that are connected by a path (that do not pass through V) to some output in $O_D \setminus O$ we have

$$|\hat{I}| < (1/50) \cdot m^2.$$

Denote $I = \tilde{I} \cup \hat{I}$. (then by the bounds we have on $|\tilde{I}|, |\hat{I}|$, requirement (4) is satisfied). By the definition of $\hat{I}$, all paths between $O_D \setminus O$ and $I_G \setminus I$ pass through V. (hence requirement (5) is satisfied). □

4.4 Arithmetic Model

In this section we prove lower bounds on the size of circuits computing the product of two $m \times m$ matrices. The input is of size $n = 2m^2$, and it consists of two $m \times m$ matrices X, Y. The output is the matrix $Z = X \cdot Y$, i.e., there are m^2 outputs, and the $(i,j)^{th}$ output is:

$$z_{i,j} = \sum_{k=1}^{m} x_{i,k} \cdot y_{k,j} \ .$$

Each output $z_{i,j}$ is hence a bilinear form in X and Y.

Since the product of two matrices is a bilinear form, it is natural to consider bilinear arithmetic circuits for it (recall the definition from section 1.3). We will now define the *size* and *depth* of the circuit.

Definition 4.4.1 *For a bilinear circuit C, we denote by $s(C)$ (the size of C) the number of edges between the product gates and the outputs. We denote by $d(C)$ (the depth of C) the length of the longest directed path from a product gate to an output.*

Note that these definitions ignore all gates and edges below the product gates (i.e., between the inputs and the products). That is, we allow the circuit to get for free any number of linear functions in the variables of X, and any number of linear functions in the variables of Y. We only count the size and depth above the product gates.

As discussed in the introduction the requirement that the circuit is bilinear is not restrictive over fields of characteristic zero. However over finite fields, the bilinearity assumption may be restrictive. We prove lower bounds for the general case of arithmetic circuits over finite fields in section 4.5. It is also worth noting that all known algorithms for matrix product are by bilinear arithmetic circuits.

Our main tool in proving our lower bounds is Lemma 4.3.13. Since that lemma is stated for leveled graphs, we would like our circuit to be leveled. We hence assume that our circuit is a leveled bilinear arithmetic circuit. Since we consider leveled circuits, the depth of the circuit is just the number of levels above the product gates. Our main result in this section is the following lower bound:

Theorem 4.4.2 *Any leveled bilinear arithmetic circuit C of depth d, for the product of two $m \times m$ matrices, is of size:*

$$s(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega\left(\frac{1}{d} m^2 \lambda_d(m^2)\right) & d > 1 \end{cases}$$

In order to prove lower bounds for a non-leveled bilinear circuit, we just level it. We can do that by increasing it's size by a factor of d. We can then use the lower bounds for leveled circuits. We hence have the following corollary:

Corollary 4.4.3 *Any bilinear arithmetic circuit C of depth d, for the product of two $m \times m$ matrices, is of size:*

$$s(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega\left(\frac{1}{d^2} m^2 \lambda_d(m^2)\right) & d > 1 \end{cases}$$

In particular, this gives the following size-depth tradeoff: there is no linear size and constant depth bilinear arithmetic circuit for the product of two matrices. Note that the theorem is valid for any field. It just needs the bilinearity assumption.

After proving the main theorem, we will use the result of Baur and Strassen [3] to prove a lower bound on the size of bounded depth arithmetic circuits for the trace of the product of 3 matrices:

$$\sum_{i,j,k=1}^{m} x_{i,j} \cdot y_{j,k} \cdot z_{k,i} \ ,$$

which is a function with a single output.

Let us start with bilinear circuits of depth 1. The structure of such circuits is very simple. First they compute linear forms. Then there is one level of product gates computing bilinear forms. Finally, there is one level of m^2 plus gates computing the outputs. For the proof, we will use the following notation.

$$O_j = \{ \ z_{i,j} \mid i \in \{1, ..., m\} \ \} \ ,$$

i.e., O_j denotes the set of outputs of the j^{th} column of the output matrix.

Theorem 4.4.4 *Any leveled bilinear circuit C of depth 1, for the product of two $m \times m$ matrices, is of size:*

$$s(C) = \Omega(m^3) \ .$$

Proof: Since the circuit is of depth one, the outputs come right after the product gates. Each output is computed by a plus gate that adds the results of the product gates that are connected to it. We will show that there are at least m^2 edges connected to each output column O_j. Hence, since there are m output columns, the result follows.

Assume for a contradiction that an output column O_j is connected to $r < m^2$ product gates. Denote the functions computed by these gates by $M_1, ..., M_r$. For each M_k denote by $L_{k,1}(X), L_{k,2}(Y)$ the two linear functions that it multiplies. That is,

$$M_k(X, Y) = L_{k,1}(X) \cdot L_{k,2}(Y).$$

Since $r < m^2$, we can find a substitution to the matrix X such that:

1. $X \neq 0$.

2. For every $1 \leq k \leq r$, $L_{k,1}(X) = 0$.

Hence, for every $1 \leq k \leq r$,

$$M_k(X, Y) = 0.$$

Therefore, $O_j = 0$, no matter what Y is. At the other hand, $X \neq 0$, and hence we can find a substitution for the matrix Y such that no column of $X \cdot Y$ is all zero (a contradiction). $\square$

The main idea of the proof for depth 1 was the following: if there is a small number of edges in the circuit then one can find a substitution for the matrix X, such that a large number of outputs are forced to be 0 (no matter what Y is). The main idea of the proof for larger depth is the following: First apply Lemma 4.3.13 to transform the circuit into a circuit of depth 1. This is done by removing from the circuit a certain number of inputs, outputs and intermediate gates. Then use the argument for depth 1. However, since we remove from the circuit a certain number of nodes, we will need a more general argument. Roughly speaking, we will need to generalize the proof for depth 1 to the case where X and Y are restricted to certain subspaces of matrices. We will show that even if X and Y are restricted to (not too small) subspaces, their product can still not be computed by a small circuit of depth 1. We will use the following notations.

Definition 4.4.5 *For a matrix X, denote by $(X)_j$ the j^{th} column of X. For a linear subspace of matrices $\mathcal{A}$, denote*

$$(\mathcal{A})_j = \{\ (X)_j \mid X \in \mathcal{A}\ \}\ .$$

Since $\mathcal{A}$ is a linear subspace then so is $(\mathcal{A})_j$.

We clearly have:

Proposition 4.4.6 *For any linear subspace of matrices $\mathcal{A}$,*

$$dim(\mathcal{A}) \leq \sum_{j=1}^{m} dim((\mathcal{A})_j)\ .$$

We will also need the following lemma. Roughly speaking, the lemma claims that if X is a matrix of large rank and $\mathcal{B}$ is a subspace of matrices of high dimension, then for many columns j, $dim(\{\ (X \cdot Y)_j \mid Y \in \mathcal{B}\})$ is high.

Lemma 4.4.7 *Let X be an $m \times m$ matrix of rank $\geq \frac{2}{3}m$. Let $\mathcal{B}$ be a linear subspace of $m \times m$ matrices, such that $dim(\mathcal{B}) \geq m^2 - k$. Then for any subset of columns, D, of size $|D| \geq \frac{10k}{m}$, there exists a column $j \in D$ such that*

$$dim(\{ \ (X \cdot Y)_j \mid Y \in \mathcal{B} \ \}) \geq \frac{m}{2} \ .$$

Proof:

Let D be a subset of columns such that $|D| \geq \frac{10k}{m}$. We first show that there is a column $j \in D$ such that $dim((\mathcal{B})_j) \geq \frac{9}{10}m$. If this wasn't the case then by Proposition 4.4.6 we would have had:

$$m^2 - k \leq dim(\mathcal{B}) \leq |D| \left(\frac{9}{10}m - 1 \right) + (m - |D|)m = m^2 - \frac{m}{10}|D| - |D| < m^2 - k \ .$$

Let $j \in D$ be such that $dim((\mathcal{B})_j) \geq \frac{9}{10}m$. Since $rank(X) \geq \frac{2}{3}m$, we have

$$dim(\{ \ (X \cdot Y)_j \mid Y \in \mathcal{B} \ \}) = dim(\{ \ X \cdot v \mid v \in (\mathcal{B})_j \ \}) \geq rank(X) \ - \ (m - \ dim((\mathcal{B})_j)) \geq$$

$$\geq \frac{2}{3}m - \frac{1}{10}m > \frac{m}{2} \ .$$

□

In the proof for depth 1, we had $X \neq 0$. Since in the proof for higher depth Y will be restricted to a subspace of matrices, we will need X to satisfy a stronger condition. Namely, we need X to be of high rank. However, X itself will also be restricted to a subspace of matrices. Therefore, we want to show that in any subspace (of matrices) of high dimension, there is a matrix of high rank.

Lemma 4.4.8 *In any subspace of $m \times m$ matrices of dimension larger than $(2mr - r^2 + m)$, there is a matrix of rank at least r.*

Proof: We have two different proofs. The first is for finite fields and the second is for fields of characteristic zero. We wish to compare the number of matrices with rank at most r to the number of matrices in our linear subspace. If we prove that the number of matrices in the linear subspace is larger, then it must contain a matrix of rank larger than r.

- Assume that F is a finite field. Denote $|F| = q$. The number of elements in a subspace of dimension larger than $(2mr - r^2 + m)$ is larger than $q^{2mr - r^2 + m}$. We will now count the number of $m \times m$ matrices with rank at most r. Note that for every such matrix there are r rows, such that every row in the matrix is in their linear span. There are $\binom{m}{r}$ possible ways to choose these r rows. There are q^{mr} possible ways to choose the r vectors for these rows.

Every other row is in the linear span of these r rows, so it can be one of q^r vectors. Therefore the number of matrices of rank at most r is bounded from above by:

$$\binom{m}{r} q^{mr} (q^r)^{m-r} < q^{2mr-r^2+m} \ .$$

- Assume that F is a field of characteristic zero. Instead of counting, we will consider the dimension of the variety of matrices with rank at most r. The same argument as above shows that this variety is included in the union of $\binom{m}{r}$ varieties, each of dimension at most $mr + r(m-r)$, (as before, mr is for the freedom of choice of the first r vectors, and $r(m-r)$ is for spanning the other $m-r$ rows). Therefore the dimension is $2mr - r^2 < 2mr - r^2 + m$.

□

Before giving the formal proof for Theorem 4.4.2, let us first give a sketch of this proof. Let C be a leveled bilinear arithmetic circuit of depth d for the product of two $m \times m$ matrices. Let G be the leveled graph of depth d, corresponding to the graph of the circuit between the product gates and the outputs (i.e., the product gates are the inputs of the graph, the outputs are the outputs of the graph and the levels of the circuit between the product gates and the outputs are the levels of the graph). We would like to prove that

$$s(C) \geq \Omega\left(\frac{1}{d} m^2 \lambda_d(m^2)\right) \ .$$

Assume for a contradiction that $s(C) < \frac{1}{1000d} m^2 \lambda_d(m^2)$, or in other words

$$|E_G| < \frac{1}{1000d} m^2 \lambda_d(m^2) \ .$$

By Lemma 4.3.13, we can find a set of columns, D, and 3 sets of vertices, V, I, O, in the graph G, such that:

- $m \leq |V| = k = o(m^2)$.
- I is a small set of inputs.
- D is a set of $\lceil \frac{10k}{m} \rceil$ output columns.
- For every $i \in D$, $|O_i \cap O|$ is small (where O_i is the i^{th} output column).
- All the paths from $O_D \setminus O$ to the inputs pass through V or reach I.

We will derive a contradiction in 4 steps:

1. Since I is a small set of product gates, we can find a subspace of matrices, $\mathcal{A}$, such that for every matrix X in $\mathcal{A}$ all the gates in I output 0. The matrix X will be restricted to the subspace $\mathcal{A}$.

2. Note that once a matrix X is fixed, the nodes of V just compute linear functions in the variables of Y. Therefore, for every matrix $X \in \mathcal{A}$, we can find a subspace of matrices $\mathcal{B}_X$ such that for every pair $(X \in \mathcal{A}\ ,\ Y \in \mathcal{B}_X)$ all the gates in V output zero. Since V is a small set, the dimension of $\mathcal{B}_X$ is high. The matrix Y will be restricted to the subspace $\mathcal{B}_X$.

3. The subspace $\mathcal{A}$ is of high dimension, and for every X the subspace $\mathcal{B}_X$ is of high dimension. Therefore, we can find $X \in \mathcal{A}$ such that $dim((X \cdot \mathcal{B}_X)_j)$ is large for some $j \in D$ (this will follow from Lemma 4.4.7, and Lemma 4.4.8).

4. Since we restrict $X \in \mathcal{A}$ and $Y \in \mathcal{B}_X$, all the gates in V and I output zero. Therefore $X \cdot Y$ is computed by a circuit with no paths between $O_D \setminus O$ and $I_G \setminus I$. Hence, all the outputs in $O_D \setminus O$ must give zero. Because of the third step, this is a contradiction.

Let us now give the formal proof.

Proof of Theorem 4.4.2:

We already gave the proof for $d = 1$, so assume $d > 1$. Assume for a contradiction that we have a leveled bilinear arithmetic circuit C of depth d for the product of two $m \times m$ matrices, and such that

$$s(C) < \frac{1}{1000d} m^2 \lambda_d(m^2)\ .$$

Let G be the leveled graph of depth d between the product gates and the outputs of C (as explained above).

As before, denote by O_j the j^{th} output column. As before, for a set $D \subset [m]$ we denote $O_D = \cup_{i \in D} O_i$. By Lemma 4.3.13 there exist sets $V \subset V_G$, $D \subset [m]$, $O \subset O_D$, $I \subset I_G$, such that:

1. $m \leq |V| = o(m^2)$.

2. $|D| \geq \frac{10|V|}{m}$.

3. For every $i \in D$, $|O_i \cap O| < \frac{1}{10} m$.

4. $|I| < \frac{1}{10} m^2$.

5. $P_G[\ I_G \setminus I\ ,\ O_D \setminus O\ ,\ V\ ] = 0$.

Denote $k = |V|$. Hence, we have a set of outputs O_D consisting of at least $\frac{10k}{m}$ output columns, such that each of the output columns in O_D has a small intersection with O, and there are no paths between $O_D \setminus O$ and $I_G \setminus I$ that do not pass through V.

In the circuit C, the set I is just a set of product gates. Since

$$|I| < \frac{1}{10}m^2 \ ,$$

we can find a subspace of matrices $\mathcal{A}$ of dimension

$$dim(\mathcal{A}) \geq m^2 - |I| > \frac{9}{10}m^2 \ ,$$

such that for every $X \in \mathcal{A}$, all the product gates in I give zero.

Assume that a matrix X is fixed. All the functions computed by the vertices of V are now linear functions in the variables of Y. Therefore, for every matrix X, there is a subspace of matrices, $\mathcal{B}_X$, such that for every $Y \in \mathcal{B}_X$, all the gates in V give zero, and such that

$$dim(\mathcal{B}_X) \geq m^2 - |V| = m^2 - k \ .$$

Since

$$dim(\mathcal{A}) \geq \frac{9}{10}m^2,$$

by Lemma 4.4.8 we can find a matrix $X \in \mathcal{A}$, such that $rank(X) \geq \frac{2}{3}m$. We fix this X. Since $|D| \geq \frac{10k}{m}$ and $dim(\mathcal{B}_X) \geq m^2 - k$, we can apply Lemma 4.4.7 to get a column $j \in D$, such that

$$dim((X \cdot \mathcal{B}_X)_j) \geq \frac{m}{2} \ .$$

At the other hand, for $X \in \mathcal{A}$ and $Y \in \mathcal{B}_X$, all the product gates in I and all the gates in V output zero. Since all the paths to $O_D \setminus O$ pass through V or I, we get that for every $X \in \mathcal{A}$ and $Y \in \mathcal{B}_X$, all the outputs in $O_D \setminus O$ must give zero. Therefore, for every $j \in D$,

$$dim((X \cdot \mathcal{B}_X)_j) \leq |O_j \cap O| < \frac{m}{10} \ ,$$

(a contradiction). □

Theorem 4.4.2 gives a super-linear lower bound for a multi-output function. The following theorem of [3] (it is an immediate corollary of the results presented there) shows that we can also obtain a super-linear lower bound for a single-output function.

Theorem 4.4.9 *[3] Suppose that $f(x_1, ..., x_n)$ is computed by an arithmetic circuit of size s and depth d over a field of characteristic zero, then there is an arithmetic circuit of size $3s$ and depth $2d$ that computes $f, \frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n}$.*

(note that in this theorem the size and depth of a circuit are just the usual size and depth, i.e., the size is the number of edges and the depth is the length of the longest directed path).

Consider the following function:

$$f(X,Y,Z) = \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} x_{i,j}y_{j,k}z_{k,i} = trace(X \cdot Y \cdot Z) \ ,$$

where X, Y, Z are $m \times m$ matrices. Notice that

$$\frac{\partial f}{\partial z_{k,i}} = \sum_{j=1}^{m} x_{i,j}y_{j,k} = (X \cdot Y)_{i,k} \ .$$

Therefore the partial derivatives of f with respect to the $Z_{k,i}$'s are the outputs of the product of two matrices. If we take into account the price that we have to pay when transforming a circuit to a bilinear leveled one we get:

Theorem 4.4.10 *Every arithmetic circuit C of depth d, that computes the trace of the product of three $m \times m$ matrices, over a field of characteristic zero, is of size:*

$$\Omega\left(\frac{1}{d^2} m^2 \lambda_{6d}(m^2)\right) \ .$$

We can also generalize this result for circuits over finite fields, but we must make another assumption: the circuit computes

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} x_{i,j} \cdot y_{j,k} \cdot z_{k,i} \ ,$$

as a polynomial and not as a function. As mentioned in the introduction, over finite fields there are many polynomials that represent the same function. For example, $x^p - x = 0$ over a field with p elements. Therefore, we demand that the circuit computes this exact polynomial. In this case we can apply the theorem of Baur and Strassen as before and get the same lower bound as in the case of characteristic zero.

4.5 Circuits with Arbitrary Gates

In this section, we prove lower bounds for circuits with arbitrary gates over finite fields (recall the definition from section 1.1). Since the proofs are similar for all finite fields, we will only give in details the proofs for the field $GF(2)$, that is, the Boolean case. The proofs for other fields are only sketched.

We will mainly concentrate on matrix product over $GF(2)$. Our main result in this section is the following lower bound:

Theorem 4.5.1 *Any leveled Boolean circuit with arbitrary gates, C, of depth d, for the product of two $m \times m$ Boolean matrices over $GF(2)$, is of size:*

$$size(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega(\frac{1}{d}m^2\lambda_d(m^2)) & d > 1 \end{cases}$$

As before, in order to prove lower bounds for a non-leveled circuit, we just level it. We can do that by increasing it's size by a factor of d. We can then use the lower bounds for leveled circuits. We hence have the following corollary.

Corollary 4.5.2 *Any Boolean circuit with arbitrary gates, C, of depth d, for the product of two $m \times m$ Boolean matrices over $GF(2)$, is of size:*

$$size(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega(\frac{1}{d^2}m^2\lambda_d(m^2)) & d > 1 \end{cases}$$

Our proof for these lower bounds is quite general and can be applied for many other functions. In particular, the lower bound applies also for the following functions:

1. The product of two $m \times m$ matrices over $GF(p)$, where each element of the field is represented by $\lceil \log p \rceil$ bits.
2. All the functions considered in [34], e.g., the parallel prefix linear transformation over $GF(2)$.

As mention above, we define a similar model of circuits with arbitrary gates over any finite field. We prove similar lower bounds for this model for any finite field. In particular, for the field $GF(p)$, we prove a lower bound for the product of two $m \times m$ matrices, where the inputs take values in the field.

Let us start with a short sketch of the proof of Theorem 4.5.1. As before, the proof is based on Lemma 4.3.13. In all that comes below we use the notations of Subsection 4.3.5. We will apply Lemma 4.3.13 on the circuit C. By Lemma 4.3.13, there is a set of columns, $D \subset [m]$, and small sets of vertices I, O, V, such that, I is a set of inputs, O is a set of outputs, V is a set of intermediate gates, and such that,

$$P_G[\, I_G \setminus I \,,\, O_D \setminus O \,,\, V \,] = 0$$

(where O_D is the set of outputs corresponding to the set of columns D). The values of the outputs in $O_D \setminus O$ are hence determined by the outputs of the gates in V and by the values of the inputs in I. Therefore, for any fixed assignment for the inputs in I, the total number of possible values that the outputs in $O_D \setminus O$ can take is at most $2^{|V|}$ (which is the total number of values that the gates in V can output). Since V is a small set, $2^{|V|}$ is a relatively small number and we conclude that for any fixed assignment for the inputs in I the outputs in $O_D \setminus O$ can only get a small number of values. We will derive a contradiction by finding a fixed assignment for the inputs in I, such that the outputs in $O_D \setminus O$ can get many values.

Let us describe our assignment for I. First, we fix Y to be a matrix in which any minor of size $\frac{m}{2} \times |D|$ is of high rank. Therefore, there will be many vectors in the image of any such minor. After fixing Y, we set to zero all the inputs in I that come from the matrix X. That is, we allow X to be any matrix in which this certain set of entries is zero (the entries that appear in I). Since I is a relatively small set, there are many such matrices. More accurately, the set of all such matrices form a linear subspace of dimension $\geq m^2 - |I|$. The last step will be to show that after fixing Y as above, and after fixing to 0 all the inputs in I that come from X, there are still many possibilities for the product $X \cdot Y$. In particular, we will get that the number of possible values for the outputs in $O_D \setminus O$ is larger than $2^{|V|}$.

Thus, the first step is to show that there is a matrix Y, in which any minor of size $\frac{m}{2} \times |D|$ is of high rank. Note that it is not hard to prove that there exists a matrix Y, in which any such minor is of maximal rank. For our proof, however, it will be enough to have the weaker requirement that any such minor is of high rank.

Definition 4.5.3 *Let $Y = (y_{i,j})$ be an $m \times m$ matrix. For sets $\alpha, \beta \subset [m]$, denote*

$$(Y)_{\alpha,\beta} = (y_{i,j}) \text{ such that } i \in \alpha,\ j \in \beta\ ,$$

i.e., $(Y)_{\alpha,\beta}$ is the minor of Y with the set of rows α and the set of columns β.

Claim 4.5.4 *For any $l = o(m)$, there exists an $m \times m$ matrix Y (over $GF(2)$), such that for any $\alpha, \beta \subset [m]$, with $|\alpha| = \lceil \frac{m}{2} \rceil$ and $|\beta| = l$,*

$$rank((Y)_{\alpha,\beta}) \geq \frac{l}{2}\ .$$

Proof:
We assume for convenience that m, l are even numbers. We will show that a random matrix Y satisfies the requirement of the lemma (with high probability). Let us first calculate the probability that a certain minor of size $\frac{m}{2} \times l$ is of rank $\leq \frac{l}{2}$. As in the proof of Lemma 4.4.8, the number of $\frac{m}{2} \times l$ matrices of rank $\leq \frac{l}{2}$ is at most

$$\binom{l}{\frac{l}{2}} \cdot 2^{\frac{m}{2} \cdot \frac{l}{2}} \cdot 2^{\frac{l}{2} \cdot (l - \frac{l}{2})} < 2^{\frac{1}{4}ml + \frac{1}{4}l^2 + l}\ .$$

Therefore, the probability that a certain minor of size $\frac{m}{2} \times l$ is of rank $\leq \frac{l}{2}$ is at most

$$2^{\frac{1}{4}ml + \frac{1}{4}l^2 + l} \cdot 2^{-\frac{m}{2}l} = 2^{-(\frac{1}{4}ml - \frac{1}{4}l^2 - l)}\ .$$

Hence, if Y is a random matrix the probability that some $\frac{m}{2} \times l$ minor is of rank $\leq \frac{l}{2}$ is at most

$$\binom{m}{\frac{m}{2}} \cdot \binom{m}{l} \cdot 2^{-(\frac{1}{4}ml - \frac{1}{4}l^2 - l)} \leq 2^{-(\frac{1}{4}ml - \frac{1}{4}l^2 - l - 2m)} < 1 \ .$$

Consequently, there is an $m \times m$ matrix Y, in which every $\frac{m}{2} \times l$ minor is of rank $> \frac{l}{2}$. $\square$

For a set of coordinates $\beta \subset [m]$, and an m-vector v, denote by v_β the restriction of v to β, i.e.,

$$v_\beta = (v_i)_{i \in \beta} \ .$$

For a set of coordinates $\alpha \subset [m]$, denote by $\mathcal{V}_\alpha$ the subspace of all vectors that have the value 0 in all the coordinates outside α, i.e.,

$$\mathcal{V}_\alpha = \{ \ v \in \{0,1\}^m \mid \forall i \in [m] \setminus \alpha \ , \ v_i = 0 \ \} \ .$$

We will now prove that if Y is the matrix guaranteed by Claim 4.5.4, for $l = |\beta|$, then for every vector space $\mathcal{V}_\alpha$ of high dimension, there are many different vectors in the set

$$\{ \ (v \cdot Y)_\beta \mid v \in \mathcal{V}_\alpha \ \}.$$

Claim 4.5.5 *Let $\beta \subset [m]$ be such that $|\beta| = o(m)$. Let $\alpha \subset [m]$ be such that $|\alpha| = \lceil \frac{m}{2} \rceil$. Let Y be an $m \times m$ matrix that satisfies the requirement of Claim 4.5.4 for $l = |\beta|$. Then, the number of different vectors of the form $(v \cdot Y)_\beta$, where $v \in \mathcal{V}_\alpha$, is at least $2^{\frac{|\beta|}{2}}$.*

Proof:

Since Y satisfies the requirement of Claim 4.5.4 for $l = |\beta|$, the minor $(Y)_{\alpha,\beta}$ is of rank at least $|\beta|/2$. Hence, the image of this minor is of dimension at least $|\beta|/2$. The image of this minor is just the set of all vectors of the form $(v \cdot Y)_\beta$, where $v \in \mathcal{V}_\alpha$. Hence, this set is of size at least $2^{\frac{|\beta|}{2}}$. $\square$

We are now ready to give the proof of Theorem 4.5.1.

Proof of Theorem 4.5.1:

For $d = 1$ the proof is trivial by the observation that every output column depends on all the variables in X. For larger depth, assume for a contradiction that $size(C) < (1/1000d) \cdot m^2 \lambda_d(m^2)$. Let G be the graph of the circuit. By Lemma 4.3.13 there exist sets $V \subset V_G$, $D \subset [m]$, $O \subset O_D$, $I \subset I_G$, such that:

1. $m \leq |V| = k = o(m^2)$.

2. $|D| = \left\lceil \frac{10k}{m} \right\rceil = o(m)$.

3. For every $i \in D$, $|O_i \cap O| < \frac{1}{10}m$.

4. $|I| < \frac{1}{10}m^2$.

5. $P_G[\, I_G \setminus I \ , \ O_D \setminus O \ , \ V \,] = 0$.

Since $P_G[\, I_G \setminus I \ , \ O_D \setminus O \ , \ V \,] = 0$, the values of the outputs in $O_D \setminus O$ are determined by the outputs of the gates in V and by the inputs in I.

Fix Y to be a matrix in which every $\lceil \frac{m}{2} \rceil \times |D|$ minor is of rank higher than $\frac{|D|}{2}$ (by Claim 4.5.4, there exists such a matrix). Denote by I_X the set of inputs in I that are variables of X. Obviously,

$$|I_X| \leq |I| < \frac{1}{10}m^2 \ .$$

Fix all the input variables in I_X to be 0. Since all the inputs in I are now fixed, the values of the outputs in $O_D \setminus O$ are determined by the outputs of the gates in V. Since there are at most 2^k possible values for the outputs of the gates in V, we conclude that after fixing Y and I_X as above, the outputs in $O_D \setminus O$ can get at most 2^k different values.

On the other hand, we only fixed less than $\frac{1}{10}m^2$ of the entries of X. Therefore, the number of rows of X, in which we fixed at most $\frac{m}{2}$ entries, is at least $\frac{8}{10}m$. For each one of these rows, we can apply Claim 4.5.5 (with $\beta = D$) and conclude that the outputs in the corresponding row in O_D can get at least $2^{\frac{|D|}{2}}$ possible values. Since the value of each of these $\frac{8}{10}m$ rows of X is independent of the values of the other rows, we conclude that the total number of different values that the outputs in O_D can get is at least

$$2^{\frac{|D|}{2} \cdot \frac{8}{10}m} = 2^{\frac{4}{10} \cdot |D| \cdot m} \ .$$

Since for every $i \in D$, $|O_i \cap O| < \frac{1}{10}m$, we get that

$$|O| < \frac{1}{10} \cdot |D| \cdot m \ .$$

Hence, the outputs in O can get at most

$$2^{\frac{1}{10} \cdot |D| \cdot m}$$

different values. Therefore, the outputs in $O_D \setminus O$ can get at least

$$2^{\frac{4}{10} \cdot |D| \cdot m} \ / \ 2^{\frac{1}{10} \cdot |D| \cdot m} = 2^{\frac{3}{10} \cdot |D| \cdot m} \geq 2^{3k}$$

different values (a contradiction). $\square$

As mentioned above, we can also obtain similar lower bounds for circuits with arbitrary gates over any finite field. A circuit with arbitrary gates over a finite field $GF(p)$ is defined similarly to a circuit with arbitrary gates over $GF(2)$. The only difference is that the inputs take values in $GF(p)$ and every gate of in-degree k is labeled with an arbitrary function from $GF(p)^k$ to $GF(p)$. Note that in particular this model includes the model of arithmetic circuits over $GF(p)$.

Theorem 4.5.6 *Any circuit with arbitrary gates over $GF(p)$, C, of depth d, for the product of two $m \times m$ matrices over $GF(p)$, is of size:*

$$size(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega(\frac{1}{d^2} m^2 \lambda_d(m^2)) & d > 1 \end{cases}$$

The proof is similar to the proof of Theorem 4.5.1 with some minor modifications: We prove a version of Claim 4.5.4 to get a matrix with the same properties over $GF(p)$. Then we prove a version of Claim 4.5.5 for $GF(p)$. (Both proofs are similar to the original proofs). We then repeat the proof of Theorem 4.5.1 to get Theorem 4.5.6 (we have to make some minor changes, e.g., the outputs of the gates in V can get $p^{|V|}$ different values (rather than $2^{|V|}$) etc').

We can also prove similar lower bounds for Boolean circuits with arbitrary gates for the product of two $m \times m$ matrices over the field $GF(p)$, where each element of the field is represented by $\lceil \log p \rceil$ bits (note that the input to the circuit is of size $2m^2 \lceil \log p \rceil$). One way of proving this lower bound is by a reduction to Theorem 4.5.6. Just observe that $\{0,1\} \subset GF(p)$, and hence any Boolean circuit (with arbitrary gates) can be viewed as a circuit with arbitrary gates over $GF(p)$. Another way of proving this lower bound is by proving a version of Lemma 4.3.13 with different parameters (because of the $\lceil \log p \rceil$ factor), and then we can repeat the proof of Theorem 4.5.1 with minor modifications.

Theorem 4.5.7 *Any Boolean circuit with arbitrary gates, C, of depth d, for the product of two $m \times m$ matrices over $GF(p)$, is of size:*

$$size(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega(\frac{1}{d^2} m^2 \lambda_d(m^2)) & d > 1 \end{cases}$$

Chapter 5

Lower Bounds for Matrix Product, in Bilinear and Quadratic Circuits

5.1 Introduction

We prove that any *quadratic* circuit that computes matrix product over the field $GF(2)$ has at least $3m^2 - o(m^2)$ product gates, and that any *bilinear* circuit for matrix product over the field $GF(p)$ must have at least $(2.5 + \frac{1.5}{p^3-1})m^2 - o(m^2)$ product gates.

From now on we will use the notation MP_m to denote the problem of computing the product of two $m \times m$ matrices.

5.1.1 Known Lower Bounds

In contrast to the major advances in proving upper bound, the attempts to prove lower bounds on the size of bilinear circuits that compute MP_m were less successful. Denote by $q_*(MP_m)$ and $bl_*(MP_m)$ the number of product gates in a smallest quadratic circuit for MP_m, and in a smallest bilinear circuit for MP_m respectively. We also denote by $bl_{tot}(MP_m)$ the total number of gates in a smallest bilinear circuit for MP_m. In 78' Brocket and Dobkin proved that $bl_*(MP_m) \geq 2m^2 - 1$ over any field [40]. This lower bound was later generalized by Lafon and Winograd to a lower bound on $q_*(MP_m)$ over any field [25]. In 89' Bshouty showed that over $GF(2)$, $q_*(MP_m) \geq 2.5m^2 - O(m \log m)$ [8]. Recently Bläser proved a lower bound of $2m^2 + m - 3$ on $q_*(MP_m)$ over any field [6]. In [5] Bläser proved that $bl_*(MP_m) \geq 2.5m^2 - 3m$ over any field.

In [36] it is shown that any bounded depth circuit for MP_m, over any field, has a super linear (in m^2) size. Notice however, that the best known circuits for MP_m have depth $\Omega(\log m)$.

5.1.2 Bilinear Rank

An important notion that is highly related to the problem of computing matrix product in bilinear circuits is the notion of bilinear rank.

A bilinear form in two sets of variables x, y is a polynomial in the variables of x and the variables of y, which is linear in the variables of x and linear in the variables of y. Clearly each output of MP_m is a bilinear form in $x = \{x_{i,j}\}$, $y = \{y_{i,j}\}$. The bilinear rank of a set of bilinear forms $\{ b_1(x,y), \ldots, b_m(x,y) \}$ is the smallest number of rank 1 bilinear forms that span $b_1, \ldots, b_m$, where a rank 1 bilinear form is a product of a linear form in the x variables and a linear form in the y variables. We denote by $R_F(b_1, \cdots, b_m)$ the bilinear rank of $\{ b_1, \ldots, b_m \}$ over the field F. For further background see [9, 16].

We denote by $R_F(MP_m)$ the bilinear rank over F of the m^2 outputs of matrix product, i.e. it is the bilinear rank of the set $\{\sum_{k=1}^{m} x_{i,k} \cdot y_{k,j}\}_{i,j}$ over F.

The following inequalities are obvious (over any field).

- $q_*(MP_m) \leq bl_*(MP_m) \leq 2q_*(MP_m)$.
- $R_F(MP_m) = bl_*(MP_m)$.
- The following inequality is less obvious, but also not so hard to see.

  $$bl_*(MP_m) \leq bl_{tot}(MP_m) \leq \text{poly}(\log m) \cdot bl_*(MP_m) \ .$$

  I.e. up to polylogarithmic factors, the number of product gates in a smallest bilinear circuit for MP_m, over any field F, is equal to the total number of gates in the circuit.

5.1.3 Results and Methods

We prove that any *quadratic* circuit that computes MP_m over the field $GF(2)$ has at least $3m^2 - o(m^2)$ product gates (i.e. $q_*(MP_m) \geq 3m^2 - o(m^2)$ over $GF(2)$). We also prove that over the field $GF(p)$ every *bilinear* circuit for MP_m must have at least $(2.5 + \frac{1.5}{p^3-1})m^2 - o(m^2)$ product gates (i.e. $bl_*(MP_m) \geq (2.5 + \frac{1.5}{p^3-1})m^2 - o(m^2)$ over $GF(p)$). Both of these results actually hold for the bilinear rank as well.

The proof of the lower bound over $GF(2)$ is based on techniques from the theory of linear codes. However, we cannot use known results from coding theory in a straightforward way, since we are not dealing with codes in which every two words are distant, but rather with codes on matrices in which the distance between two code words, of two matrices, is proportional to the rank of the difference of the matrices. The reduction from circuits to codes and the proof of the bound are given in section 5.3.

The proof of the second bound is based on a lemma proved by Bläser in [5]. We prove that in the case of finite fields we can use the lemma with better parameters than those used by Bläser. This result is proved in section 5.4.

5.1.4 Organization

In section 5.2 we present some algebraic and combinatorial tools that we need for the proofs of our lower bounds.

In section 5.3 we introduce the notion of linear codes of matrices, and prove our lower bound on bilinear circuits that compute MP_m over $GF(2)$. In section 5.4 we prove our lower bound on bilinear circuits that compute MP_m over $GF(p)$.

5.2 Algebraic and Combinatorial tools

In this section we present some algebraic and combinatorial tools that we will use.

The following lemma is an extremely weak variant of the famous Schwartz-Zippel lemma which shows that every non zero polynomial (non zero as a formal expression) over a large enough field has a non zero assignment in the field (see [44, 63]).

Lemma 5.2.1 *Let P be a polynomial of degree d in $x_1, \ldots, x_m$ over some field F, such that $d < |F|$, and such that at least one of the coefficients of P is not zero. Then we can find an assignment, $\rho \in F^m$, to the x_i's, such that $P(\rho_1, \ldots, \rho_m) \neq 0$.*

We say that two polynomials p, q in m variables are equivalent over a field F, if $p(x_1, \ldots, x_m) = q(x_1, \ldots, x_m)$ for any $x_1, \ldots, x_m \in F$. We denote $p \equiv q$ if p and q are equivalent over F (we omit F from the notation as the field that we deal with will be clear from the context).

Lemma 5.2.2 *Let P be a polynomial of degree d in the variables $x_1, \ldots, x_m$ over a field F. If $P \not\equiv 0$ then we can find an assignment, $\rho \in F^m$, to the x_i's such that at most d of the ρ_i's get a nonzero value, and such that $P(\rho_1, \ldots, \rho_m) \neq 0$.*

Proof: P is equal (as a function) to a polynomial $\bar{P}$ in which the degree of each variable is at most $|F| - 1$. We call $\bar{P}$ the reduction of P. Consider some monomial M in $\bar{P}$ whose coefficient is not zero. We assign all the variables that do not appear in M to zero. The resulting polynomial (after the assignment), is a polynomial in the variables of M, which is not the zero polynomial as it is a reduced polynomial which has a monomial with a non zero coefficient (M of course). Therefore according to lemma 5.2.1 there is some assignment to the variables of M, that gives this polynomial a nonzero value. Therefor we have found an assignment which gives nonzero values only to the variables of M (and there are at most d such variables) under which $P \neq 0$. $\square$

The following useful lemma, which is a straightforward implication of the previous lemma, is the key lemma in most of our proofs. The lemma deals with linear forms in m^2 variables. From now on we shall think about such linear forms as linear forms in the entries of $m \times m$ matrices.

Lemma 5.2.3 *Let $\mu_1, \ldots, \mu_{m^2}$ be m^2 linearly independent linear forms in m^2 variables over some field F. Let P be a polynomial of degree d in km^2 variables over F, i.e. we can view P as a polynomial $P(x_1, ..., x_k)$ in the entries of k matrices, $x_1, ..., x_k$, of size $m \times m$ each. Assume that $P \not\equiv 0$. Then we can find k matrices $a_1, \ldots, a_k \in M_m(F)$ such that $P(a_1, ..., a_k) \neq 0$ and such that there exist $m^2 - d$ linear forms among $\mu_1, ..., \mu_{m^2}$'s that vanish on all the a_i's.*

Proof: The idea of the proof is the following. Let $b_1, \ldots, b_{m^2}$ be the dual basis of $\mu_1, ..., \mu_{m^2}$, i.e. it is a basis of $M_m(F)$ satisfying $\forall i, j \;\; \mu_i(b_j) = \delta_{i,j}$. We wish to find k matrices, $a_1, ..., a_k$, such that $P(a_1, ..., a_k) \neq 0$, and such that there exist $b_{i_1}, ..., b_{i_d}$ that span all of them. If we manage to find such matrices, then since the b_i's are the dual basis to the μ_i's we will get that $m^2 - d$ of the μ_i's vanish on $a_1, ..., a_k$. The way to find such matrices that are contained in the span of a small subset of the b_i's, is based on lemma 5.2.2.

So let $b_1, \ldots, b_{m^2}$ be the dual basis to $\mu_1, \ldots, \mu_{m^2}$, i.e. $\forall i, j \;\; \mu_i(b_j) = \delta_{i,j}$. We now change the variables of P. Let $\alpha_{i,j} \;\; j = 1...k, \; i = 1...m^2$, be a set of km^2 variables. Denote $x_j = \sum_{i=1}^{m^2} \alpha_{i,j} b_i$. Thus $P(x_1, ..., x_k)$ can be viewed as a polynomial of degree d in the km^2 variables $\alpha_{i,j}$. Therefore $P \not\equiv 0$ as a polynomial in the $\alpha_{i,j}$'s. Hence, according to lemma 5.2.2 there exists an assignment, ρ, to the $\alpha_{i,j}$'s such that at most d of them get a nonzero value. Define $a_j = \sum_{i=1}^{m^2} \rho_{i,j} b_i$. Clearly $P(a_1, \ldots, a_k) \neq 0$. Since at most d of the $\rho_{i,j}$'s got non zero values, we see that there are at most d b_i's such that all the a_j's are linear combinations of them. Since the b_i's are the dual basis to $\mu_1, \ldots, \mu_{m^2}$ we get that there are at least $m^2 - d$ of the μ_i's that vanish on all the a_j's. Therefore $a_1, \ldots, a_k$ satisfy the requirements of the lemma. $\square$

The next lemma will enable us to translate properties of matrices over large fields of characteristic p to properties of matrices (of higher dimension) over $GF(p)$.

Lemma 5.2.4 *There exist an embedding, $\phi : GF(p^m) \hookrightarrow M_m(GF(p))$. That is there exist a mapping $\phi : GF(p^m) \mapsto M_m(GF(p))$ such that*

- *ϕ is a one to one linear transformation.*
- *$\phi(1) = I$, where I is the $m \times m$ identity matrix.*

- *ϕ is multiplicative, i.e. $\forall x, y \in GF(p^m)$ we have that $\phi(xy) = \phi(x) \cdot \phi(y)$.*

This embedding also induces an embedding $M_k(GF(p^m)) \hookrightarrow M_{mk}(GF(p))$.

This lemma is a standart tool in algebra, but for completness we give the proof.

Proof: $GF(p^m)$ is an m dimensional vector space over $GF(p)$. Each element $x \in GF(p^m)$ can be viewed as a linear transformation $x : GF(p^m) \mapsto GF(p^m)$ in the following way:

$$\forall y \in GF(p^m) \;\; x(y) = x \cdot y \; .$$

Clearly this is a linear transformation of $GF(p^m)$ into itself, as a vector space over $GF(p)$. Therefore, by picking a basis to $GF(p^m)$ we can represent the linear transformation corresponding to each $x \in GF(p^m)$ by a matrix $a_x \in M_m(GF(p))$. Thus, we have defined a mapping $\phi : GF(p^m) \mapsto M_m(GF(p))$ such that $\phi(x) = a_x$, and it is easy to verify that this mapping is an embedding of $GF(p^m)$ into $M_m(GF(p))$. The way to generalize it to an embedding of $M_k(GF(p^m))$ into $M_{mk}(GF(p))$ is the following. Let $a = (a_{i,j}) \in M_k(GF(p^m))$ be some matrix. Every entree of $a_{i,j}$ of a, is some element of $GF(p^m)$. We can now replace $a_{i,j}$ with the matrix $\phi(a_{i,j})$. Thus the resulting matrix will be a $km \times km$ matrix whose entries are in $GF(p)$. Again it is easy to verify that this is indeed an embedding of $M_k(GF(p^m))$ into $M_{mk}(GF(p))$. $\square$

In addition to the algebraic lemmas we also need the following combinatorial tools.

Definition 5.2.5 *Let F be a field, and let v, u be two vectors in F^t. We denote by $\mathbf{weight}(\mathrm{v})$ the number of nonzero coordinates of v. Let $\mathbf{d_H}(\mathrm{v},\mathrm{u}) = \mathbf{weight}(\mathrm{v}-\mathrm{u})$, i.e. $\mathbf{d_H}(\mathrm{v},\mathrm{u})$ is the number of coordinates on which u and v differ. $\mathbf{d_H}(\mathrm{v},\mathrm{u})$ is also known as the Hamming distance of u and v. We also denote by $\mathbf{agree}(\mathrm{u},\mathrm{v})$ the number of coordinates on which u and v are equal, i.e. $\mathbf{agree}(\mathrm{u},\mathrm{v}) = \mathrm{t} - \mathbf{d_H}(\mathrm{v},\mathrm{u})$.*

The next lemma shows that if a vector space contains a set of vectors such that every pair/triplet of them don't agree on many coordinates (i.e. their Hamming distance is large) then it is of large dimension. There are numerous similar lemmas in coding theory, and in particular the first part of our lemma is the famous Plotkin bound (see [26]).

Lemma 5.2.6
1. *In every set of k vectors in $GF(p)^t$, such that $p < k$, there are two vectors that agree on at least $(\frac{t}{p} - \frac{t}{k})$ coordinates.*
2. *In every set of k vectors in $GF(p)^t$, such that $2p < k$, there are three vectors that agree on at least $(\frac{t}{p^2} - \frac{3t}{pk})$ coordinates.*

Proof: We begin by proving the first claim. Let $v_1, \ldots, v_k$ be k vectors in $GF(p)^t$. We are going to estimate $\sum_{i<j} \mathbf{agree}(\mathrm{v_i}, \mathrm{v_j})$ in two different ways. On the one hand this sum is at most $\binom{k}{2}$ times the maximum of $\mathbf{agree}(\mathrm{v_i}, \mathrm{v_j})$. On the other hand consider a certain coordinate. For every $\alpha \in GF(P)$ denote by m_α the number of vectors among the v_i's that are equal to α on this coordinate. Clearly $\sum_{\alpha=0}^{p-1} m_\alpha = k$. The contribution of this coordinate to $\sum_{i<j} \mathbf{agree}(\mathrm{v_i}, \mathrm{v_j})$ is exactly $\sum_{\alpha=0}^{p-1} \binom{m_\alpha}{2}$. By convexity

$$\sum_{\alpha=0}^{p-1} \binom{m_\alpha}{2} \geq p \cdot \frac{1}{2} \cdot \frac{k}{p}\left(\frac{k}{p} - 1\right) = \frac{k(k-p)}{2p} .$$

We get that

$$\binom{k}{2} \cdot \max_{i<j}(\mathbf{agree}(\mathrm{v_i}, \mathrm{v_j})) \geq$$

$$\sum_{i<j} \mathbf{agree}(\mathrm{v_i}, \mathrm{v_j}) \geq \mathrm{t} \cdot \frac{\mathrm{k(k-p)}}{\mathrm{2p}} .$$

Therefore

$$\max_{i<j}(\mathbf{agree}(\mathrm{v_i}, \mathrm{v_j})) \geq \frac{\mathrm{t}}{\mathrm{p}} \cdot \frac{\mathrm{k-p}}{\mathrm{k-1}} \geq \frac{\mathrm{t}}{\mathrm{p}} \cdot \frac{\mathrm{k-p}}{\mathrm{k}} = \frac{\mathrm{t}}{\mathrm{p}} - \frac{\mathrm{t}}{\mathrm{k}} .$$

The proof of the second claim is similar. We give two different estimates to $\sum_{i<j<l} \mathbf{agree}(\mathrm{v_i}, \mathrm{v_j}, \mathrm{v_l})$ (the number of coordinates on which v_i, v_j, and v_l are the same). In the same manner as before we get that

$$\max_{i<j<l}(\mathbf{agree}(\mathrm{v_i}, \mathrm{v_j}, \mathrm{v_l})) \geq \frac{\mathrm{t}}{\mathrm{p^2}} \cdot \frac{\mathrm{k-p}}{\mathrm{k-1}} \cdot \frac{\mathrm{k-2p}}{\mathrm{k-2}} \geq \frac{\mathrm{t}}{\mathrm{p^2}} - \frac{\mathrm{3t}}{\mathrm{pk}} .$$

□

Corollary 5.2.7 *If $\{0,1\}^t$ contains k vectors $v_1, \ldots, v_k$, such that $2 < k$ and $\forall i \neq j$ $\mathbf{d_H}(\mathrm{v_i}, \mathrm{v_j}) \geq \mathrm{N}$, then $t \geq 2N - 4\frac{N}{k+2}$.*

Proof: According to lemma 5.2.6 there are two vectors, w.l.o.g. v_1 and v_2, such that $\mathbf{agree}(\mathrm{v_1}, \mathrm{v_2}) \geq \frac{\mathrm{t}}{2} - \frac{\mathrm{t}}{\mathrm{k}}$. Since $\mathbf{d_H}(\mathrm{v_1}, \mathrm{v_2}) = \mathrm{t} - \mathbf{agree}(\mathrm{v_1}, \mathrm{v_2})$ we get that

$$t - \left(\frac{t}{2} - \frac{t}{k}\right) \geq \mathbf{d_H}(\mathrm{v_1}, \mathrm{v_2}) \geq \mathrm{N}$$

and the result follows. □

5.3 Lower bound over GF(2)

In this section we prove our main theorems.

Theorem 5.3.1 $bl_*(MP_m) \geq 3m^2 - O(m^{\frac{5}{3}})$ *(in other words* $R_{GF(2)}(MP_m) \geq 3m^2 - O(m^{\frac{5}{3}})$*).*

The second theorem that we shall prove is a lower bound for quadratic circuits.

Theorem 5.3.2 $q_*(MP_m) \geq 3m^2 - O(m^{\frac{5}{3}})$. *I.e. the number of product gates in any quadratic circuit that computes the product of two* $m \times m$ *matrices over* $GF(2)$ *is at least* $3m^2 - O(m^{\frac{5}{3}})$*.*

Clearly theorem 5.3.2 imply theorem 5.3.1, but we first prove of theorem 5.3.1 as it is more intuitive and simple. We begin by introducing the notion of linear codes of matrices.

5.3.1 Linear Codes of Matrices

Definition 5.3.3 *A linear code of matrices is a mapping,*

$$\Gamma : M_m(GF(2)) \mapsto \{0,1\}^s ,$$

(for some s*) with the following properties:*

- Γ *is linear.*
- *For any matrix* a*,* $\mathbf{weight}(\Gamma(\mathrm{a})) \geq \mathrm{m} \cdot \mathrm{rank}(\mathrm{a})$*.*

From the linearity of Γ and the requirement on $\mathbf{weight}(\Gamma(\mathrm{a}))$ we get the following corollary.

Corollary 5.3.4 Γ *is a one to one mapping, and for any two matrices* a *and* b*,* $\mathbf{d_H}(\Gamma(\mathrm{a}), \Gamma(\mathrm{b})) \geq \mathrm{m} \cdot \mathrm{rank}(\mathrm{a} - \mathrm{b})$*.*

The following theorem shows that the dimension of the range of any linear code of matrices is large (i.e. s must be large).

Theorem 5.3.5 *Let* $\Gamma : M_m(GF(2)) \mapsto \{0,1\}^s$ *be a linear code of matrices, then* $s \geq 3m^2 - O(m^{\frac{5}{3}})$ *.*

Proof: Denote

$$\Gamma(a) = (\ \mu_1(a),\ \ldots,\ \mu_s(a)\ ) .$$

The proof is based on the following lemma that shows that we can find $k = m^{\frac{1}{3}}$ matrices, $a_1,\ \ldots,\ a_k \in M_m(GF(2))$, with the following properties.

- $\forall i \neq j$, $a_i - a_j$ is an invertible matrix.
- There are $m^2 - \binom{k}{2}m$ linear forms among the μ_i's that vanish on all the a_i's.

We state the lemma for every $k < 2^m$ but we apply it only to $k = m^{\frac{1}{3}}$.

Lemma 5.3.6 *For every m, k such that $k < 2^m$, and any $\mu_1, \ldots, \mu_{m^2}$ linearly independent linear forms in m^2 variables, over $GF(2)$, there are k matrices, $a_1, \ldots, a_k \in M_m(GF(2))$, such that for every $i \neq j$, $a_i - a_j$ is an invertible matrix, and such that $m^2 - \binom{k}{2}m$ of the μ_i's vanish on them.*

Proof: Consider the following polynomial P in k matrices:

$$P(a_1, \ldots, a_k) = \text{determinant}\left(\prod_{\mathrm{i<j}}(\mathrm{a_i} - \mathrm{a_j})\right) .$$

Clearly a set of k matrices $a_1, \ldots, a_k$ satisfy $P(a_1, \ldots, a_k) \neq 0$ iff all the matrices $a_i - a_j$ are invertible. In addition, it is easy to see that $deg(P) = \binom{k}{2}m$. Therefore if we show that $P \not\equiv 0$ over $GF(2)$, then according to lemma 5.2.3 we will get what we wanted to prove.

In order to show that $P \not\equiv 0$ we just have to prove the existence of k matrices, such that the difference of every two of them is invertible. Lemma 5.2.4 assures us that we can embed the field $GF(2^m)$ into $M_m(GF(2))$. Denote this embedding by $\Phi : GF(2^m) \hookrightarrow M_m(GF(2))$. We take k distinct elements in $GF(2^m)$, $x_1, \ldots, x_k$. Their images, $\Phi(x_1), \ldots, \Phi(x_k)$, are matrices in $M_m(GF(2))$ such that the difference of every two of them, $\Phi(x_i) - \Phi(x_j) = \Phi(x_i - x_j)$, is an invertible matrix. This is because the x_i's are distinct (i.e. $x_i - x_j \neq 0$), and every nonzero element in $GF(2^m)$ is invertible. Thus, $\Phi(x_1), \ldots, \Phi(x_k)$ are exactly the k matrices that we were looking for. This concludes the proof of the lemma. $\square$

We proceed with the proof of the theorem. Let $k = m^{\frac{1}{3}}$. Since Γ is a one to one mapping, there are m^2 independent linear forms among $\mu_1, \ldots, \mu_s$. Therefore we can use lemma 5.3.6 and get that there are k matrices $a_1, \ldots, a_k$ such that for every $i \neq j$ $a_i - a_j$ is invertible, and such that, w.l.o.g., $\mu_{s-r+1}, \ldots, \mu_s$ vanish on $a_1, \ldots, a_k$ for some $r \geq m^2 - \binom{k}{2}m \geq m^2 - m^{\frac{5}{3}}$.

Since the last r linear forms vanish on all the a_i's, we are going to restrict our attention only to the first $s - r$ linear forms. So from now on we only consider $\Gamma(a_i)$ restricted to its first $s - r$ coordinates.

Since each of the differences, $a_i - a_j$ $(\forall i \neq j)$, is an invertible matrix, we get that $\mathbf{d_H}(\Gamma(\mathrm{a_i}), \Gamma(\mathrm{a_j})) \geq \mathrm{m}^2$. Thus, $\Gamma(a_1), \ldots, \Gamma(a_k)$ are k vectors contained in $\{0,1\}^{s-r}$ (we consider only their first $s - r$ coordinates !) such that the hamming distance of every pair of them is at least m^2. Therefore according to corollary 5.2.7 we get that

$$s - r \geq 2m^2 - 4\frac{m^2}{k+2} .$$

Since $r \geq m^2 - m^{\frac{5}{3}}$ and $k = m^{\frac{1}{3}}$, we get that

$$s \geq 3m^2 - O(m^{\frac{5}{3}})$$

which is what we wanted to prove. This concludes the proof of the theorem. □

5.3.2 Proof of Theorem 5.3.1

Assume that $bl_*(MP_m) = m$. Let C be a smallest bilinear circuit for MP_m. Let

$$\mu_1(x) \cdot \eta_1(y), \; \ldots, \; \mu_s(x) \cdot \eta_s(y)$$

be the s bilinear forms computed in the product gates of C. We will show that these bilinear forms define in a very natural way a code on $M_m(GF(2))$. The code thus defined, will have the property that the dimension of the space into which the code maps $M_m(GF(2))$ is exactly s. Thus, according to theorem 5.3.5 we will get that $s \geq 3m^2 - O(m^{\frac{5}{3}})$, which is what we wanted to prove.

So we begin by defining a mapping from $M_m(GF(2))$ to $\{0,1\}^s$. Let $\Gamma : M_m(GF(2)) \mapsto \{0,1\}^s$ be the following mapping.

$$\Gamma(x) = (\mu_1(x), \; \ldots, \; \mu_s(x))) \; .$$

Notice that we ignore the η_i's in the definition of Γ. The next lemma shows that Γ is a linear code of matrices.

Lemma 5.3.7 *Γ is a linear transformation with the property that for every matrix $x \in M_m(GF(2))$,* $\mathbf{weight}(\Gamma(\mathrm{x})) \geq \mathrm{m} \cdot \mathrm{rank}(\mathrm{x})$.

Proof: Clearly Γ is a linear transformation from $M_m(GF(2))$ to $\{0,1\}^s$. So we only have to prove the claim about the weights. Let x be a matrix of rank r. Assume w.l.o.g. that $\mu_1(x) = \ldots = \mu_k(x) = 1$ and that $\mu_{k+1}(x) = \ldots = \mu_s(x) = 0$, i.e. $\mathbf{weight}(\Gamma(\mathrm{x})) = \mathrm{k}$. We shall show that $k \geq mr$. For every $y \in M_m(GF(2))$, the m^2 entries of $x \cdot y$ are functions of $\mu_1(x) \cdot \eta_1(y), \; \ldots, \; \mu_s(x) \cdot \eta_s(y)$. Since $\mu_{k+1}(x) = \ldots = \mu_s(x) = 0$, we get that $x \cdot y$ is a function of $\eta_1(y), \; \ldots, \; \eta_k(y)$. Therefore there are at most 2^k different matrices of the form $x \cdot y$. Since $\mathrm{rank}(\mathrm{x}) = \mathrm{r}$ we get that there are exactly 2^{mr} different matrices of the form $x \cdot y$. Therefore $k \geq mr$. This concludes the proof of the lemma. □

Therefore Γ is a linear code of matrices, so according to theorem 5.3.5 we get that $s \geq 3m^2 - O(m^{\frac{5}{3}})$ which is what we wanted to prove. This concludes the proof of theorem 5.3.1. ♣

5.3.3 Proof of Theorem 5.3.2

As in the proof of theorem 5.3.1 we will show that every quadratic circuit for MP_m, defines a code on $M_m(GF(2))$. The code thus defined, will have the property that s (i.e the dimension of the space into which the code maps $M_m(GF(2))$) is exactly the number of product gates in the circuit. Thus, according to theorem 5.3.5 we will get that $s \geq 3m^2 - O(m^{\frac{5}{3}})$, which is what we wanted to prove.

Let C be a quadratic circuit for MP_m. Assume that the product gates of C compute the quadratic forms $\mu_1(x,y) \cdot \eta_1(x,y), \ \ldots, \ \mu_s(x,y) \cdot \eta_s(x,y)$. Thus, each of the outputs $(x \cdot y)_{i,j}$ can be written as a sum of these quadratic forms:

$$(x \cdot y)_{i,j} = \sum_{k=1}^{s} \alpha_{i,j}^{(k)} \cdot \mu_k(x,y) \cdot \eta_k(x,y) \ ,$$

where $\alpha_{i,j}^{(k)} \in \{0,1\}$.

We would like to have a proof similar to the proof of theorem 5.3.1. In that proof we defined a code of matrices using the linear transformation $\mu_1, ..., \mu_s$. Unfortunately this method will fail here as μ_i is a linear function in both the variables of x and the variables of y and not just in the variables of x as in the proof of theorem 5.3.1. In order to overcome this obstacle we introduce a new set of variables $z = \{z_{i,j}\}_{i,j=1...m}$. We think about z as an $m \times m$ matrix. Define the following s linear forms in z:

$$\gamma_k(z) = \sum_{i,j} \alpha_{i,j}^{(k)} z_{i,j} \ , \quad k = 1, \ldots, s \ .$$

We get that

$$\begin{gathered}
\sum_{k=1}^{s} \mu_k(x,y) \cdot \eta_k(x,y) \cdot \gamma_k(z) = \\
\sum_{k=1}^{s} (\sum_{i,j} z_{i,j} \cdot \alpha_{i,j}^{(k)}) \cdot \mu_k(x,y) \cdot \eta_k(x,y) = \\
\sum_{i,j} z_{i,j} \sum_{k=1}^{s} \alpha_{i,j}^{(k)} \cdot \mu_k(x,y) \cdot \eta_k(x,y) = \\
\sum_{i,j} z_{i,j} \cdot (x \cdot y)_{i,j} = \mathbf{trace}(\mathrm{x} \cdot \mathrm{y} \cdot \mathrm{z}^{\mathrm{t}}) \ ,
\end{gathered} \tag{5.1}$$

where $(z^t)_{i,j} = z_{j,i}$. The computation that we just performed shows that the γ_k's that we introduced are quite natural. We also notice that z plays the same role in $\mathbf{trace}(\mathrm{x} \cdot \mathrm{y} \cdot \mathrm{z}^{\mathrm{t}})$ as x and y. These observations motivate us to try to repeat the proof of theorem 5.3.1 using the γ_k's instead of the μ_i's.

So define a linear mapping $\Gamma : M_m(GF(2)) \mapsto \{0,1\}^s$ by

$$\Gamma(z) = (\gamma_1(z),\ \ldots,\ \gamma_s(z)) \ .$$

The following lemma shows that Γ is indeed a linear code of matrices.

Lemma 5.3.8 Γ *is a linear mapping and it has the property that for every matrix z,* $\mathbf{weight}(\Gamma(\mathrm{z})) \geq \mathrm{m} \cdot \mathrm{rank}(\mathrm{z})$.

Proof: Clearly Γ is a linear mapping. So we only have to prove the claim about the weights. Let z_0 be a matrix of rank r, and assume w.l.o.g. that $\gamma_1(z_0) = \ldots = \gamma_k(z_0) = 1$ and $\gamma_{k+1}(z_0) = \ldots = \gamma_s(z_0) = 0$. We wish to prove that $k \geq mr$. From equation 5.1 we get that

$$\mathbf{trace}(\mathrm{x} \cdot \mathrm{y} \cdot \mathrm{z_0}^{\mathrm{t}}) = \sum_{\mathrm{i}=1}^{\mathrm{k}} \mu_{\mathrm{i}}(\mathrm{x}, \mathrm{y}) \cdot \eta_{\mathrm{i}}(\mathrm{x}, \mathrm{y}) \ .$$

We now consider the discrete derivatives of this equation. Let $e_{i,j}$ be the matrix of all zeros but 1 in the (i,j)'th place. Define

$$\frac{\partial}{\partial x_{i,j}} \mathbf{trace}(\mathrm{x} \cdot \mathrm{y} \cdot \mathrm{z_0}^{\mathrm{t}}) \stackrel{\mathrm{def}}{=}$$
$$\mathbf{trace}((\mathrm{x} + \mathrm{e_{i,j}}) \cdot \mathrm{y} \cdot \mathrm{z_0}^{\mathrm{t}}) - \mathbf{trace}(\mathrm{x} \cdot \mathrm{y} \cdot \mathrm{z_0}^{\mathrm{t}}) \ .$$

On the one hand

$$\mathbf{trace}((\mathrm{x} + \mathrm{e_{i,j}}) \cdot \mathrm{y} \cdot \mathrm{z_0}^{\mathrm{t}}) - \mathbf{trace}(\mathrm{x} \cdot \mathrm{y} \cdot \mathrm{z_0}^{\mathrm{t}}) =$$
$$\mathbf{trace}(\mathrm{e_{i,j}} \cdot \mathrm{y} \cdot \mathrm{z_0}^{\mathrm{t}}) = (\mathrm{z_0} \cdot \mathrm{y}^{\mathrm{t}})_{\mathrm{i,j}} \ .$$

On the other hand we have that

$$\begin{aligned} &\mathbf{trace}((\mathrm{x} + \mathrm{e_{i,j}}) \cdot \mathrm{y} \cdot \mathrm{z_0}^{\mathrm{t}}) - \mathbf{trace}(\mathrm{x} \cdot \mathrm{y} \cdot \mathrm{z_0}^{\mathrm{t}}) = \\ &\sum_{i=1}^{k} (\mu_i(x + e_{i,j}, y) \cdot \eta_i(x + e_{i,j}, y) - \mu_i(x,y) \cdot \eta_i(x,y)) = \\ &\sum_{i=1}^{k} (\mu_i(e_{i,j}, 0) \cdot \eta_i(x,y) + \mu_i(x,y) \cdot \eta_i(e_{i,j}, 0)) + \\ &+ \sum_{i=1}^{k} \mu_i(e_{i,j}, 0) \cdot \eta_i(e_{i,j}, 0) \ , \end{aligned} \tag{5.2}$$

where the last equality follows from the linearity of the μ_i's and the η_i's. Since $(z_0 \cdot y^t)_{i,j}$ is a linear form in y, we actually get that

$$(z_0 \cdot y^t)_{i,j} = \frac{\partial}{\partial x_{i,j}} \mathbf{trace}(\mathrm{x} \cdot \mathrm{y} \cdot \mathrm{z_0}^{\mathrm{t}}) =$$

$$\sum_{i=1}^{k}(\mu_i(e_{i,j},0)\cdot\eta_i(x,y)+\mu_i(x,y)\cdot\eta_i(e_{i,j},0))$$
$$\subset \mathrm{span}(\mu_\mathrm{i}(\mathrm{x},\mathrm{y}),\ \eta_\mathrm{i}(\mathrm{x},\mathrm{y}))$$

(since $(z_0\cdot y^t)_{i,j}$ is a linear form the third summand of equation 5.2 sums to 0). In the same manner we define

$$\frac{\partial}{\partial y_{i,j}}\mathbf{trace}(\mathrm{x}\cdot\mathrm{y}\cdot{\mathrm{z}_0}^{\mathrm{t}}) \stackrel{\mathrm{def}}{=}$$
$$\mathbf{trace}(\mathrm{x}\cdot(\mathrm{y}+\mathrm{e}_{\mathrm{i},\mathrm{j}})\cdot{\mathrm{z}_0}^{\mathrm{t}})-\mathbf{trace}(\mathrm{x}\cdot\mathrm{y}\cdot{\mathrm{z}_0}^{\mathrm{t}})\ .$$

We get that

$$(x^t\cdot z_0)_{i,j}=\frac{\partial}{\partial y_{i,j}}\mathbf{trace}(\mathrm{x}\cdot\mathrm{y}\cdot{\mathrm{z}_0}^{\mathrm{t}})=$$
$$\sum_{i=1}^{k}(\mu_i(0,e_{i,j})\cdot\eta_i(x,y)+\mu_i(x,y)\cdot\eta_i(0,e_{i,j}))$$
$$\subset \mathrm{span}(\mu_\mathrm{i}(\mathrm{x},\mathrm{y}),\ \eta_\mathrm{i}(\mathrm{x},\mathrm{y}))\ .$$

Denote by PD the set of all the discrete partial derivatives

$$\left\{\frac{\partial}{\partial x_{i,j}}\mathbf{trace}(\mathrm{x}\cdot\mathrm{y}\cdot{\mathrm{z}_0}^{\mathrm{t}}),\frac{\partial}{\partial \mathrm{y}_{\mathrm{i},\mathrm{j}}}\mathbf{trace}(\mathrm{x}\cdot\mathrm{y}\cdot{\mathrm{z}_0}^{\mathrm{t}})\right\}_{i,j}\ .$$

We just proved that PD is contained in the linear span of

$$\{\ \mu_i(x,y),\ \eta_i(x,y)\ \}_{i=1}^{k}\ ,$$

in the vector space of all linear forms in x, y. Therefore

$$\dim(\mathrm{span}(\mathrm{PD}))\leq$$
$$\leq \dim(\mathrm{span}\{\ \mu_\mathrm{i}(\mathrm{x},\mathrm{y}),\ \eta_\mathrm{i}(\mathrm{x},\mathrm{y})\ \}_{\mathrm{i}=1}^{\mathrm{k}})\leq 2\mathrm{k}\ . \tag{5.3}$$

We also showed that

$$\left\{\frac{\partial}{\partial x_{i,j}}\mathbf{trace}(\mathrm{x}\cdot\mathrm{y}\cdot{\mathrm{z}_0}^{\mathrm{t}}),\frac{\partial}{\partial \mathrm{y}_{\mathrm{i},\mathrm{j}}}\mathbf{trace}(\mathrm{x}\cdot\mathrm{y}\cdot{\mathrm{z}_0}^{\mathrm{t}})\right\}_{i,j}=$$
$$=\left\{\ (x^t\cdot z_0)_{i,j},\ (z_0\cdot y^t)_{i,j}\ \right\}_{i,j}\ .$$

Therefor, using our assumption that $\mathrm{rank}(\mathrm{z}_0)=\mathrm{r}$, we get that

$$\dim(\mathrm{span}(\mathrm{PD}))=$$
$$=\dim(\mathrm{span}\left\{\ (\mathrm{x}^\mathrm{t}\cdot\mathrm{z}_0)_{\mathrm{i},\mathrm{j}},\ (\mathrm{z}_0\cdot\mathrm{y}^\mathrm{t})_{\mathrm{i},\mathrm{j}}\right\}_{\mathrm{i},\mathrm{j}})=2\mathrm{mr}\ . \tag{5.4}$$

Combining equations 5.3 and 5.4 we get that $2k\geq 2mr$. □

Theorem 5.3.2 now follows from applying theorem 5.3.5 on the linear code of matrices Γ. ♣

5.4 Other Finite Fields

In this section we prove the following theorem.

Theorem 5.4.1 *The number of product gates in any bilinear circuit that computes the product of two $m \times m$ matrices over $GF(p)$ is at least $(2.5 + \frac{1.5}{p^3-1})m^2 - O(m^{\frac{7}{4}})$ (i.e. $bl_*(MP_m) \geq (2.5 + \frac{1.5}{p^3-1})m^2 - O(m^{\frac{7}{4}})$ over $GF(p)$).*

Let C be a bilinear circuit for MP_m over $GF(p)$. Assume that $\mu_1(x) \cdot \eta_1(y), \ldots, \mu_s(x) \cdot \eta_s(y)$ are the bilinear forms computed in the product gates of C. The following lemma of Bläser is the main tool in the proof of the theorem.

Lemma 5.4.2 *[5] Let $[a, b] = ab - ba$. If there are two matrices a, b such that $[a, b]$ is an invertible matrix, and such that there are t linear forms among $\mu_1, \ldots, \mu_s$ such that each of them vanish on I, a, b then*

$$s \geq t + 1.5m^2 \ .$$

We are going to prove that we can find a, b such that $(1 - \frac{1}{p^3})m^2 + \frac{s}{p^3} - O(m^{\frac{7}{4}})$ linear forms among $\mu_1, \ldots, \mu_s$ vanish on I, a, b , and such that $[a, b]$ is invertible.

Proof of Theorem 5.4.1: We begin by proving that (w.l.o.g.) many of the μ_i's vanish on I. The following lemma shows that we can always find an invertible matrix such that many of the μ_i's vanish on it. As before we assume that $\mu_1, \ldots, \mu_{m^2}$ are independent linear forms.

Lemma 5.4.3 *There exists an invertible matrix c, such that at least $(1 - \frac{1}{p})m^2 + \frac{s}{p} - O(m^{\frac{5}{3}})$ of the μ_i's vanish on it, where $m^2 - O(m^{\frac{5}{3}})$ of the μ_i's that vanish on it are among $\mu_1, \ldots, \mu_{m^2}$.*

Proof: An analog of lemma 5.3.6 over $GF(p)$ guarantees that we can find $k = m^{\frac{1}{3}}$ matrices, $a_1, \ldots, a_k \in M_m(GF(p))$, such that for $i \neq j$ $a_i - a_j$ is invertible, and such that $m^2 - \binom{k}{2}m$ linear forms among $\mu_1, \ldots, \mu_{m^2}$ vanish on all of them. Denote $r = m^2 - \binom{k}{2}m$. And assume w.l.o.g. that $\mu_1, \ldots, \mu_r$ vanish on all the a_i's.

Let us consider the following k vectors in $GF(p)^{s-r}$:

$$\Gamma(a_i) \stackrel{\text{def}}{=} (\mu_{r+1}(a_i), \ldots, \mu_s(a_i)) \ , \ i = 1 \ldots k \ .$$

As in the proof of theorem 5.3.1, we get that since $\forall i \neq j$ $a_i - a_j$ is an invertible matrix, then $\mathbf{d_H}(\Gamma(\mathrm{a_i}), \Gamma(\mathrm{a_j})) \geq \mathrm{m}^2$. According to lemma 5.2.6, two of these vectors agree on at least $\frac{s-r}{p} - \frac{s-r}{k}$ coordinates. Assume that $\Gamma(a_1)$ and $\Gamma(a_2)$ are these vectors. Denote $c = a_1 - a_2$. We have that

c is an invertible matrix, such that the first $r = m^2 - \binom{k}{2}m$ linear forms (which are independent) vanish on it, and such that all the linear forms that $\Gamma(a_1)$ and $\Gamma(a_2)$ agree on, vanish on it as well. Therefore there are at least

$$r + \frac{s-r}{p} - \frac{s-r}{k}$$

linear forms that vanish on c. Since $r = m^2 - \binom{k}{2}m$, and $k = m^{\frac{1}{3}}$, we get that at least

$$(1 - \frac{1}{p})m^2 + \frac{s}{p} - O(m^{\frac{5}{3}})$$

linear forms vanish on c, $m^2 - O(m^{\frac{5}{3}})$ of them are among $\mu_1, \ldots, \mu_{m^2}$ (we assume for simplicity that $m^2 < s < 10m^2$, as it will not change the results). This completes the proof of the lemma. □

The lemma doesn't tell us who c is, but using the *sandwiching method* we can assume that $c = I$: We know that $x \cdot y$ is computed using the bilinear forms

$$\mu_1(x) \cdot \eta_1(y), \ \ldots, \ \mu_s(x) \cdot \eta_s(y) \ .$$

We now do the following trick: $x \cdot y = (x \cdot c) \cdot (c^{-1} \cdot y)$, therefore $x \cdot y$ can be computed using the bilinear forms

$$\tilde{\mu_1}(x) \cdot \tilde{\eta_1}(y), \ \ldots, \ \tilde{\mu_s}(x) \cdot \tilde{\eta_s}(y) \ ,$$

where

$$\tilde{\mu_i}(x) \stackrel{\text{def}}{=} \mu_i(x \cdot c) \ \text{ and } \ \tilde{\eta_\mathrm{i}}(\mathrm{y}) \stackrel{\text{def}}{=} \eta_\mathrm{i}(\mathrm{c}^{-1} \cdot \mathrm{y}) \ .$$

Thus, if $\mu_i(c) = 0$ then we get that $\tilde{\mu_i}(I) = \mu_i(I \cdot c) = 0$. This trick is called sandwiching, for further background see [5, 12].

So by combining the sandwiching method and lemma 5.4.3 we get that we can assume w.l.o.g. that $(1 - \frac{1}{p})m^2 + \frac{s}{p} - O(m^{\frac{5}{3}})$ of the μ_i's, where $m^2 - O(m^{\frac{5}{3}})$ of them are among $\mu_1, \ldots, \mu_{m^2}$, vanish on I. The next lemma now assures us that we can find two matrices a, b that satisfy the requirements of lemma 5.4.2.

Lemma 5.4.4 *There are two matrices a, b such that $[a, b]$ is an invertible matrix and such that at least $(1 - \frac{1}{p^3})m^2 + \frac{s}{p^3} - O(m^{\frac{7}{4}})$ of the μ_i's vanish on I, a, b.*

Proof: The proof of this lemma is similar to the proof of lemma 5.4.3. Let $k = m^{\frac{1}{4}}$. The following lemma shows that we can find k matrices such that many of the μ_i's vanish on all of them and such that among their differences there are matrices satisfying the requirements of lemma 5.4.2.

Lemma 5.4.5 *For every m, k, such that $p^{\frac{m}{2}} > 4\binom{k}{3}$, and any $\mu_1, \ldots, \mu_{m^2}$ linearly independent linear forms, in m^2 variables, over $GF(p)$, there are k matrices, $a_1, \ldots, a_k$, such that $\forall i < j < l$, $[a_i - a_l, a_j - a_l]$ is invertible, and such that $m^2 - 2\binom{k}{3}m$ of the μ_i's vanish on all the a_i's.*

Proof: Again we use lemma 5.2.3. Let P be the following polynomial.

$$P(a_1, ..., a_k) = \text{determinant}\left(\prod_{\mathrm{i<j<l}} [\mathrm{a_i - a_l, a_j - a_l}]\right).$$

Clearly $deg(P) = 2\binom{k}{3}m$ (as a polynomial in the entries of the a_i's). Therefore if we will prove that $P \not\equiv 0$, i.e. that there exist k matrices on which P is not zero, then according to lemma 5.2.3 we are done. This is guaranteed by the following lemma.

Lemma 5.4.6 *If $p^{\frac{m}{2}} > 4\binom{k}{3}$ then there exist k matrices in $M_m(GF(p))$, $a_1, \ldots, a_k$ such that $\forall i < j < l$, $[a_i - a_l, a_j - a_l]$ is invertible.*

We prove the lemma only for m even. Clearly this will not affect theorem 5.4.1, as the lower bound for odd m follows from the lower bound for even m.

Proof: Consider the following polynomial in **4k** variables (i.e. it is a polynomial in k matrices over $\mathbf{M_2(GF(p))}$!).

$$P_2(x_1, ..., x_k) = \text{determinant}\left(\prod_{\mathrm{i<j<l}} [\mathrm{a_i - a_l, a_j - a_l}]\right).$$

P_2 is a polynomial of degree $d = 4\binom{k}{3}$ over $GF(p)$, in the entries of the a_i's. Clearly P_2 is not the zero polynomial (since it is a product of non zero polynomials). Consider the field $F = GF(p^{\frac{m}{2}})$. Since $d < |F|$ we get by lemma 5.2.1 that there are k matrices $\rho_1, ..., \rho_k \in M_2(F)$ such that $P_2(\rho_1, ..., \rho_k) \neq 0$. That is, $\forall i < j < l$ $[\rho_i - \rho_l, \rho_j - \rho_l]$ is an invertible matrix. According to lemma 5.2.4 we can embed $M_2(F)$ in $M_m(GF(p))$. Therefore there are k matrices in $M_m(GF(p))$ satisfying $\forall i < j < l$ $[a_i - a_l, a_j - a_l]$ is an invertible matrix, which is what we wanted to prove. □

This concludes the proof of lemma 5.4.5. □

We proceed with the proof of lemma 5.4.4. We now restrict our attention to the linear forms among $\mu_{m^2+1}, \ldots, \mu_s$ that vanish on I. We shall prove that three of the matrices guaranteed by lemma 5.4.5 agree on many of these linear forms (more formally on $\frac{s-m^2}{p^3} - O(m^{\frac{7}{4}})$ of them). Thus, if a_1, a_2, a_3 are these three matrices, then we get that $(1 - \frac{1}{p^3})m^2 + \frac{s}{p^3} - O(m^{\frac{7}{4}})$ linear forms vanish on I, $(a_1 - a_3)$, $(a_2 - a_3)$, and that $[(a_1 - a_3), (a_2 - a_3)]$ is an invertible matrix, which is what we wanted to prove.

So assume w.l.o.g. that the linear forms $\mu_{m^2+1}, \ldots, \mu_{m^2+r}$, vanish on I (beside those among $\mu_1, \ldots, \mu_{m^2}$ that vanish on it) where $r \geq \frac{s-m^2}{p} - O(m^{\frac{5}{3}})$. Let $a_1, \ldots, a_k$ be the matrices guaranteed by lemma 5.4.5. Consider the following vectors: $\forall 1 \leq i \leq k$,

$$v_i = (\mu_{m^2+1}(a_i), \ldots, \mu_{m^2+r}(a_i)) \in GF(p)^r .$$

According to lemma 5.2.6 three of these vectors, namely v_1, v_2, v_3, agree on at least $\frac{r}{p^2} - \frac{3r}{pk}$ coordinates. Therefore there are $\frac{r}{p^2} - \frac{3r}{pk}$ linear forms among $\mu_{m^2+1}, \ldots, \mu_{m^2+r}$ that vanish on $a_1 - a_3$ and $a_2 - a_3$. In addition there are $m^2 - 2\binom{k}{3}m$ linear forms among $\mu_1, \ldots, \mu_{m^2}$ that vanish on a_1, a_2, a_3, hence there are $m^2 - 2\binom{k}{3}m$ linear forms among $\mu_1, \ldots, \mu_{m^2}$ that vanish on $a_1 - a_3$ and on $a_2 - a_3$. Let $a = a_1 - a_3$, $b = a_2 - a_3$.

We get that there are $\frac{r}{p^2} - \frac{3r}{pk}$ linear forms among $\mu_{m^2+1}, \ldots, \mu_{m^2+r}$ that vanish on I, a, b. Since $m^2 - O(m^{\frac{5}{3}})$ of the first m^2 μ_i's vanish on I, we get that at least $m^2 - 2\binom{k}{3}m - O(m^{\frac{5}{3}})$ of the first m^2 μ_i's vanish on I, a, b. Putting it all together we get that at least

$$m^2 - 2\binom{k}{3}m - O(m^{\frac{5}{3}}) + \frac{r}{p^2} - \frac{3r}{pk}$$

linear forms among $\mu_1, \ldots, \mu_s$ vanish on I, a, b. Since $r = \frac{s-m^2}{p} - O(m^{\frac{5}{3}})$, and $k = m^{\frac{1}{4}}$, we get that at least

$$(1 - \frac{1}{p^3})m^2 + \frac{s}{p^3} - O(m^{\frac{7}{4}})$$

of the μ_i's vanish on them. This concludes the proof of lemma 5.4.4. □

Putting everything together we get by lemma 5.4.2 and lemma 5.4.4 that:

$$s \geq 1.5m^2 + (1 - \frac{1}{p^3})m^2 + \frac{s}{p^3} - O(m^{\frac{7}{4}}) .$$

Therefore

$$s \geq (2.5 + \frac{1.5}{p^3 - 1})m^2 - O(m^{\frac{7}{4}}) .$$

This concludes the proof of theorem 5.4.1. □

Bibliography

[1] Miklós Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[2] Noga Alon and Ravi B. Boppana. The monotone circuit complexity of boolean functions. *Combinatorica*, 7(1):1–22, 1987.

[3] Walter Baur and Volker Strassen. The complexity of partial derivatives. *Theoretical Computer Science*, 22:317–330, 1983.

[4] Michael Ben-Or. Private communication.

[5] Markus Bläser. A $5/2n^2$-lower bound for the rank of $n \times n$-matrix multiplication over arbitrary fields. In IEEE, editor, *40th Annual Symposium on Foundations of Computer Science: October 17–19, 1999, New York City, New York,*, pages 45–50, 1109 Spring Street, Suite 300, Silver Spring, MD 20910, USA, 1999. IEEE Computer Society Press.

[6] Markus Bläser. Lower bound for the multiplicative complexity of matrix multiplication. *Journal of computational complexity*, 9:73–112, 2000.

[7] Ravi B. Boppana and Michael Sipser. *Handbook of Th. Comp. Science*, volume A, chapter Algebraic complexity theory, pages 758–803. Elsevier and MIT Press, 1990.

[8] Nader H. Bshouty. A lower bound for matrix multiplication. *SIAM Journal on Computing*, 18(4):759–765, August 1989.

[9] Peter Bürgisser, Michael Clausen, and Mohammad A. Shokrollahi. *Algebraic Complexity Theory*. Springer, 1997.

[10] Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progression. *Journal of Symbolic Computation*, 9:251–280, 1990.

[11] Laszlo Csanky. Fast parallel matrix inversion algorithms. *SIAM Journal on Computing*, 5:618–623, 1976.

[12] Hans F. de Groote. *Lectures on the complexity of bilinear problems.* Number 245 in LNCS. Springer, 1986.

[13] Danny Dolev, Cynthia Dwork, Nicholas Pippenger, and Avi Wigderson. Superconcentrators, generalizers and generalized connectors with limited depth (preliminary version). In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, pages 42–51, Boston, Massachusetts, 25–27 April 1983.

[14] Richard Ehrenborg and Gian Carlo Rota. Apolarity and canonical forms for homogeneous polynomials. *European J. Combinatorics*, 14:157–181, 1993.

[15] Merrick Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, apr 1984.

[16] Joachim von zur Gathen. Algebraic complexity theory. *Ann. Rev. Computer Science*, pages 317–347, 1988.

[17] Daniel H. Gottlieb. A certain class of incidence matrices. In *Amer. Math. Soc.*, volume 17, pages 1233–1237, 1966.

[18] Dimitry Grigoriev and Marek Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998.*, pages 577–582. ACM, 1998.

[19] Dimitry Grigoriev and Alexander A. Razborov. Exponential complexity lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. *Applicable Algebra in Engineering, Communication and Computing*, 10(6):465–487, 2000.

[20] András Hajnal, Wolfgang Maass, Pavel Pudlák, Márió Szegedy, and György Turán. Threshold circuits of bounded depth. *Journal of Computer and System Sciences*, 46(2):129–154, April 1993.

[21] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, pages 6–20, Berkeley, California, 28–30 May 1986.

[22] Russell Impagliazzo, Ramamohan Paturi, and Michael E. Saks. Size-depth tradeoffs for threshold circuits. *SIAM Journal on Computing*, 26(3):693–707, June 1997.

[23] Mark Jerrum and Marc Snir. Some exact complexity results for straight-line computations over semi-rings. Technical Report CRS-58-80, Univ. of Edinburgh, 1980.

[24] Oded Lachish and Ran Raz. Explicit lower bound of $4.5n - o(n)$ for boolean circuits. In *Thirty-Third Annual ACM Symposium on Theory of Computing Crete, Greece,July 6-8, 2001*, pages 399–408. ACM, 2001.

[25] Jean-Claude Lafon and Shmuel Winograd. A lower bound for the multiplicative complexity of the product of two matrices. Technical report, Centre de Calcul de L'Esplanade, U.E.R. de Mathematique Univ. Louis Pasteur, Strasbourg, France, 1978.

[26] Jacobus H. van Lint. *Introduction to coding theory.* Springer-Verlag, Berlin, second edition, 1992.

[27] Jacobus H. van Lint and Richard M. Wilson. *A course in combinatorics.* Cambridge University Press, Cambridge, 1992.

[28] Satyanarayana V. Lokam. Spectral methods for matrix rigidity with applications to size-depth tradeoffs and communication complexity. In *36th Annual Symposium on Foundations of Computer Science, Milwaukee, Wisconsin, 23-25 October 1995. IEEE Computer Society*, pages 6–15, 1995.

[29] Alexis Maciel. *Threshold circuits of small majority depth.* PhD thesis, McGill University, February 1995.

[30] John W. Milnor. *Topology from the differentiable viewpoint.* Princeton Univ. Pr., 1997.

[31] Eduard I. Nečhiporuk. A boolean function. *Sov. Math. Dokl.*, 7(4):999–1000, 1966.

[32] Noam Nisan. Lower bounds for non-commutative computation. In Baruch Awerbuch, editor, *Proceedings of the 23rd Annual ACM Symposium on the Theory of Computing*, pages 410–418, New Orleans, LS, May 1991. ACM Press.

[33] Noam Nisan and Avi Wigderson. Lower bound on arithmetic circuits via partial derivatives. *Computational Complexity*, 6:217–234, 1996.

[34] Pavel Pudlak. Communication in bounded depth circuits. *Combinatorica*, 14(2):203–216, 1994.

[35] Jaikumar Radhakrishnan, Pranab Sen, and Sundar Vishwanathan. Depth-3 arithmetic circuits for $s_n^2(x)$ and extensions of the graham-pollack theorem. In Sanjiv Kapoor and Sanjiva Prasad, editors, *Foundations of Software Technology and Theoretical Computer Science, 20th Conference, FST TCS 2000 New Delhi, India, December 13-15, 2000, Proceedings*, volume 1974 of *Lecture Notes in Computer Science*, pages 176–187. Springer, 2000.

[36] Ran Raz and Amir Shpilka. Lower bounds for matrix product, in bounded depth circuits with arbitrary gates. In *Thirty-Third Annual ACM Symposium on Theory of Computing, Crete, Greece, July 6-8, 2001*, pages 409–418. ACM, 2001.

[37] Ran Raz and Avi Wigderson. Monotone circuits for matching require linear depth. *J. Assoc. Comput. Mach.*, 39(3):736–744, 1992.

[38] Alexander A. Razborov. Lower bounds on the monotone complexity of some boolean functions. *Dokl. Akad. Nauk. SSSR*, 281(4):798–801, 1985. in Russian.

[39] Alexander A. Razborov. Lower bounds for the size of circuits with bounded depth with basis $\{\wedge, \oplus\}$. *Mat. Zametki*, pages 598–607, 1987. in Russian.

[40] David Dobkin Roger W. Brocket and. On the optimal evaluation of a set of bilinear forms. *Linear Algebra and its Applications*, 9:207–235, 1978.

[41] Bruce E. Sagan. *Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions.* Chapman & Hall, 1991.

[42] Michael Saks. Private communication.

[43] Clauss P. Schnorr. A lower bound on the number of additionsin monotone computations. *TCS*, 2:305–315, 1976.

[44] Jacob T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM*, 27(4):701–717, 1980.

[45] Eli Shamir and Marc Snir. Lower bounds on the number of multiplications and the number of additions in monotone computations. Technical Report RC-6757, IBM, 1977.

[46] Eli Shamir and Marc Snir. On the depth complexity of formulas. *Mathematical Systems Theory*, 13:301–322, 1980.

[47] Amir Shpilka. Affine projections of symmetric polynomials. In *16th Annual IEEE conference on Computational Complexity*, pages 160–171, 2001.

[48] Amir Shpilka. Lower bounds for matrix product. In *42nd Annual Symposium on Foundations of Computer Science*, 2001. To appear.

[49] Amir Shpilka and Avi Wigderson. Depth-3 arithmetic formulae over fields of characteristic zero. In *14th Annual IEEE conference on Computational Complexity*, pages 87–96, 1999.

[50] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, pages 77–82, New York City, 25–27 May 1987.

[51] Volker Strassen. Gaussian elimination is not optimal. *Numer. Math*, 13:354–356, 1969.

[52] Volker Strassen. Die berechnungskomplexiät von elementarsymmetrischen funktionen und von interpolationskoeffizienten. *Numer. Math*, 20:238–251, 1973.

[53] Volker Strassen. Vermeidung von divisionen. *Journal of Reine Angew. Math.*, 264:182–202, 1973.

[54] Volker Strassen. *Handbook of Th. Comp. Science*, volume A, chapter Algebraic complexity theory, pages 633–672. Elsevier and MIT Press, 1990.

[55] Prasoon Tiwari and Martin Tompa. A direct version of Shamir and Snir's lower bounds on monotone circuit depth. *Information Processing Letters*, 49(5):243–248, March 1994.

[56] Leslie G. Valiant. Graph-theoretic arguments in low-level complexity. In *Lecture notes in Computer Science*, volume 53, pages 162–176. Springer, 1977.

[57] Leslie G. Valiant. Completeness classes in algebra. In *ACM Symposium on Theory of Computing*, pages 249–261, 1979.

[58] Leslie G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8(2):189–201, April 1979.

[59] Leslie G. Valiant. Negation can be exponentially powerful. *TCS*, 12(3):303–314, 1980.

[60] Leslie G. Valiant. Reducibility by algebraic projections. *Logic and Algorithms. L'ensignment Mathmatique*, 30:365–380, 1982.

[61] Andrew C. Yao. Separating the polynomial hierarchy by oracles. In *IEEE Symposium on Foundations of Computer Science*, pages 1–10, 1985.

[62] Andrew C. Yao. On acc and threshold circuits. In *IEEE Symposium on Foundations of Computer Science*, pages 619–627, 1990.

[63] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Symbolic and algebraic computation (EUROSAM '79, Internat. Sympos., Marseille, 1979)*, pages 216–226. Springer, Berlin, 1979.