Foundations and Trends® in
Theoretical Computer Science
Vol. 4, Nos. 1–2 (2008) 1–155
© 2009 S. V. Lokam
DOI: 10.1561/0400000011

Complexity Lower Bounds using Linear Algebra

By Satyanarayana V. Lokam

Contents

1 Introduction **2**

1.1 Scope 2
1.2 Matrix Rigidity 3
1.3 Spectral Techniques 4
1.4 Sign-Rank 5
1.5 Communication Complexity 6
1.6 Graph Complexity 8
1.7 Span Programs 9

2 Matrix Rigidity **11**

2.1 Motivation 11
2.2 Explicit Lower Bounds 15
2.3 Sparse Factorizations and Constant Depth Circuits 23
2.4 Rigidity Lower Bounds Using Algebraic Dimensions 24
2.5 Quadratic Lower Bounds on Linear Circuits 33
2.6 Paturi–Pudlák Dimensions 35

3 Spectral Methods to Study Rank Robustness **40**

3.1 Singular Value Decomposition 41

3.2 Variants of Rigidity 43
3.3 Bounded Coefficient Complexity of Linear Transformations 51
3.4 Bounded Coefficient Complexity of some Bilinear Transformations 56

4 Sign-Rank and Other Complexity Measures of Sign Matrices 65

4.1 Sign-Rank 65
4.2 Forster's Lower Bound 67
4.3 Margin Complexity, Factorization Norms, and Discrepancy 72
4.4 Sign-Rank of an AC^0 function 79
4.5 Applications of Complexity Measures of Sign Matrices 85

5 Rank Robustness and Two-Party Communication Complexity 94

5.1 The Log-Rank Conjecture in Communication Complexity 94
5.2 Discrepancy and Randomized and Quantum Communication Complexity 100
5.3 Relating Probabilistic Communication Complexity to Sign-Rank and Margin Complexity 103
5.4 Matrix Rigidity and Quantifier Alternation in Communication Complexity 107

6 Graph Complexity and Projective and Affine Dimensions of Graphs 113

6.1 Graph Complexity 114
6.2 Projective and Affine Dimensions of Graphs 117
6.3 Lower Bounds on Depth-3 Graph Complexity 125

7 Span Programs: A Linear Algebraic Model of Computation **132**

7.1 Characterizing Span Program Size 133
7.2 Explicit Lower Bounds on Monotone Span Program Size 136

8 Conclusions and Open Problems **141**

Acknowledgments **148**

References **149**

Foundations and Trends$^{\circledR}$ in
Theoretical Computer Science
Vol. 4, Nos. 1–2 (2008) 1–155

DOI: 10.1561/0400000011

Complexity Lower Bounds using Linear Algebra

Satyanarayana V. Lokam

Microsoft Research India, Bangalore – 560080, India, satya@microsoft.com

Abstract

We survey several techniques for proving lower bounds in Boolean, algebraic, and communication complexity based on certain linear algebraic approaches. The common theme among these approaches is to study robustness measures of matrix rank that capture the complexity in a given model. Suitably strong lower bounds on such robustness functions of explicit matrices lead to important consequences in the corresponding circuit or communication models. Many of the linear algebraic problems arising from these approaches are independently interesting mathematical challenges.

1

Introduction

1.1 Scope

Understanding the inherent computational complexity of problems is of fundamental importance in mathematics and theoretical computer science. While rapid progress has been made on upper bounds (algorithms), progress on lower bounds on the complexity of explicit problems has remained slow despite intense efforts over several decades. As is natural with typical impossibility results, lower bound questions are hard mathematical problems and hence are unlikely to be resolved by *ad hoc* attacks. Instead, techniques based on mathematical notions that capture computational complexity are necessary.

This paper surveys some approaches based on linear algebra to proving lower bounds on computational complexity. Linear algebraic methods are extensively used in the study of algorithms and complexity. Our focus here is on their applications to lower bounds in various models of circuits and communication complexity. We further consider mainly classical — as opposed to quantum — models of computation. Linear algebra plays an obviously pervasive role in the study of quantum complexity. Indeed, some of the techniques studied in this paper have natural extensions to quantum models. However, to keep the scope of this survey narrow enough to limit its length, we restrict ourselves to classical models. Even within classical complexity theory, we do not touch upon several applications where linear algebra plays a critical role, most notably techniques related to spectra of graphs and coding theory. Our choice of topics is centered around the theme of deriving complexity lower bounds from lower bounds on ranks of matrices or dimensions of subspaces — often after the matrices or the subspaces are altered in various ways. Such a theme occurs in several contexts in complexity theory. The rough overall approach in this theme consists of (i) distilling a rank robustness or a dimension criterion to solve a lower bound problem in complexity, (ii) developing techniques to solve such linear algebraic problems, and (iii) exploring the consequent implications to complexity lower bounds.

In the remaining sub-sections of this section, we give a brief introduction and preview of the material to be presented in detail in the later sections.

1.2 Matrix Rigidity

The most classical notion of rank robustness is *matrix rigidity*. The rigidity of a matrix A for target rank r is the minimum Hamming distance between A and a matrix of rank at most r. Valiant [98] introduced this notion to define a criterion for proving superlinear size lower bounds on linear circuits of logarithmic depth. Linear circuits are algebraic circuits consisting only of gates that compute linear combinations of their inputs. This is a natural model for computing linear transformations. Given the ubiquitous role linear transformations play in computing, understanding the inherent complexity of explicit linear transformations is important. For example, a fascinating open question is whether the Fourier transform requires a superlinear number of arithmetic operations. Furthermore, no superlinear lower bounds are known on the algebraic circuit complexity of any explicit function of constant degree, even when the circuit depth is restricted to be logarithmic. Thus, a superlinear lower bound on the size of a log-depth linear circuit computing an explicit linear transformation would be significant. Valiant showed that if the rigidity of an $n \times n$ matrix A for target rank ϵn is at least $n^{1+\delta}$ for positive constants ϵ and δ, then a linear circuit of log-depth computing the linear transformation $x \mapsto Ax$ must have superlinear number of edges. Hence, proving sufficiently strong lower bounds on the rigidity of explicit matrices would yield important consequences in complexity theory. However, the best known lower bound on the rigidity of an explicit matrix is only $\Omega\big(\frac{n^2}{r}\log\frac{n}{r}\big)$ [31, 56, 93] for target rank r. This bound is known for several explicit matrices, including the Fourier transform matrix of a prime order. Using certain algebraic dimension arguments, rigidity lower bounds of $\Omega(n^2)$ (for target rank $r = \epsilon n$ for a constant $\epsilon > 0$) are proved in [59] for the matrices whose entries are square roots of distinct primes and for matrices whose entries are primitive roots of unity of the first n^2 prime orders. While these matrices are mathematically succinct enough to describe, they are not explicit enough since their entries live in number fields of exponentially large dimensions. In Section 2, we study the notion of rigidity and its application to lower bounds on linear circuits. We will give several lower bound proofs on the rigidity of various matrices and the implied circuit lower bounds. We will also review two notions of a geometric flavor [70] that are similar to rigidity and have applications to circuit lower bounds.

1.3 Spectral Techniques

Several rank robustness functions similar to rigidity have been defined in the literature and applied to derive lower bounds in complexity theory. In Section 3, we discuss several such variations. The simplest of them considers the ℓ_2-norm of changes needed to reduce the rank of a given matrix to a target rank (notions considered in Section 3 are defined over $\mathbb{R}$ or $\mathbb{C}$). This measure of rank robustness is effectively related to the singular values of the matrix and hence lower bounds are easily proved using spectral techniques [57]. Using spectral techniques, we can also prove lower bounds on the rigidity (in the sense of Valiant) of certain matrices. The most notable of these is an Hadamard matrix [26, 42], for which a lower bound of $\Omega(n^2/r)$ is known. Spectrum of a matrix is also related to values of its sub-determinants (volumes).

Lower bounds on these measures imply lower bounds on linear circuits (over $\mathbb{C}$) with *bounded coefficients*, i.e., the coefficients in the linear combinations computed by the gates in such a circuit are bounded by a constant. Algebraic circuits over $\mathbb{C}$ with bounded coefficients is a realistic restricted model of computation since real computers can only use arithmetic operations with bounded precision in a given step. Several lower bound results have been proved in the models of bounded coefficients [22, 24, 57, 69, 75, 83]. Indeed, a classical result in [65] gives an $\Omega(n \log n)$ lower bound on the size of a linear circuit with bounded coefficients (with no restriction on depth) computing the Fourier transform. In a more recent development, Raz [83] proved a remarkable lower bound of $\Omega(n^2 \log n)$ on $n \times n$ matrix multiplication in the model of *bilinear* circuits with bounded coefficients. Raz defines a geometric variant of ℓ_2-rigidity and uses spectral techniques to prove lower bounds on the linear circuits obtained when one of the matrices in the input to a bilinear circuit performing matrix multiplication is fixed. Subsequently, Bürgisser and Lotz [22] proved lower bounds on several bilinear transformations using spectral and volume techniques. We will describe these results as well in Section 3.

1.4 Sign-Rank

In Section 4, we study a rank robustness notion called the *sign-rank* of a matrix with ±1 entries. The sign-rank of a matrix A is the minimum rank of a real matrix each of whose entries agrees in sign with the corresponding entry of A. In other words, by making sign-preserving changes to A (changes are allowed to all entries of A), its rank cannot be brought down below its sign-rank. This notion was first introduced by Paturi and Simon [71] in the context of *unbounded error probabilistic communication complexity.* Proving nontrivial, i.e., superlogarithmic, lower bounds on sign ranks of explicit matrices remained a long-standing open question until Forster [28] achieved a breakthrough by proving that the sign-rank of an $n \times n$ Hadamard matrix is at least $\sqrt{n}$. Interestingly, Forster's result and subsequent techniques for proving lower bounds on sign-rank rely on spectral techniques. Forster also considers the notion of *margin complexity* from learning theory and uses the same techniques to prove lower bounds on margin complexity. Recent results in [53, 54, 55], give new insights into sign-rank, margin complexity, and discrepancy of sign matrices by studying them all in the framework of factorization norms of operators. This general approach reveals several connections among the various complexity measures of sign matrices and led to exciting new techniques to prove lower bounds on them. In particular, they show that the discrepancy and the margin of a sign matrix are within a constant factor of each other. Lower bounds on sign-rank, margin complexity, and discrepancy are immensely useful in proving lower bounds in a variety of models such as communication complexity, circuit complexity, and learning theory. We will discuss several such applications in Section 4. Very recently, Razborov and Sherstov [88] proved a very interesting lower bound on the sign-rank of a matrix constructed from a Boolean function in AC^0. As an immediate consequence, they show that Σ_2^{cc} (the communication complexity analog of the complexity class Σ_2) is not contained in the communication complexity class UPP defined by [71]. This solves a long-standing open question [5] in two-party communication complexity. The sign-rank lower bound of [88] also has interesting consequences to lower bounds on circuit complexity and learning theory. Their result combines Forster's main argument with a number of novel techniques including the use of the pattern matrix method [90]. These techniques usher in exciting new developments and are likely to find more applications.

1.5 Communication Complexity

Ever since Mehlhorn and Schmidt [63] proved the fundamental result that the log-rank of a 0–1 matrix is a lower bound on the two-party communication complexity of the associated Boolean function, the relation between rank, and more generally rank robustness, and communication complexity has been widely investigated and exploited. Yet, the most basic question of whether log-rank and communication complexity are polynomially related to each other is still open (this is also known as the log-rank conjecture). In this conjecture, rank over $\mathbb{R}$ is considered. We begin Section 5 by discussing what is known about this conjecture. Nisan and Wigderson [68] exhibit a Boolean matrix with communication complexity at least (log-rank)$^\alpha$, where $\alpha \approx \log_3 6$. They also show that, to prove that the communication complexity of *every* $\{0,1\}$-matrix is bounded above by some polynomial function of log-rank of the matrix, it suffices to show that every $\{0,1\}$-matrix of rank r contains a sufficiently large submatrix of rank at most, say $r/4$. On the other hand, Nisan and Wigderson [68] succeed in showing that low rank matrices have high discrepancy (under the uniform distribution) using spectral arguments. Proving upper bounds on discrepancy is a common and very useful method to prove lower bounds on *probabilistic* communication complexity. In the result mentioned earlier, Linial et al. [53] show that discrepancy (under any distribution) is at least $\Omega(r^{-1/2})$ for any rank-r $\{0,1\}$-matrix. The proof of this bound uses connections among discrepancy, rank, and factorization norms of matrices discussed in Section 4. Strengthening these connections, Linial and Shraibman [54] prove general lower bounds on the bounded error probabilistic and quantum communication complexity of a sign matrix in terms of a factorization norm, called the γ_2^α-norm, of the matrix. As we noted before, the log-sign-rank of a matrix is essentially equal to the unbounded error communication complexity of the matrix. We will also see that the communication complexity analog of PP is characterized by margin complexity. Thus rank robustness measures such as sign-rank and γ_2-norm of sign matrices yield lower bounds, sometimes even characterizations, of probabilistic communication complexity. Babai et al. [5] defined two-party communication complexity analogs of traditional complexity classes such as Σ_k, PH, PSPACE, etc. While analogs of low complexity classes such as P, NP, Co–NP, and BPP were all separated from each other in two-party communication complexity, questions such as PH versus PSPACE, Σ_2 vs. PH are still open. In [84] and [57], it was shown that sufficiently strong lower bounds on rigidity (over finite fields) and a variant of rigidity (over reals) with bounds on changes would separate PH and PSPACE in communication complexity. As mentioned before, a recent result in [88] separates Σ_2^{cc} from UPP by proving a strong lower bound on the sign-rank of an AC^0-function. We conclude that lower bounds on rank and rank robustness have significant consequences to various lower bound questions in two-party communication complexity.

1.6 Graph Complexity

Graph complexity was introduced by Pudlák et al. [79]. In this model, a graph — typically bipartite — on a vertex set V is constructed by starting from a family of basic graphs, e.g., complete bipartite graphs, on V and using the elementary operations of union and intersection on edge sets. The model of graph complexity is a common generalization of Boolean circuit and two-party communication complexity. In particular, proving sufficiently strong lower bounds on the complexity of explicit bipartite graphs would imply lower bounds on formula size complexity, branching program size, and two-party communication complexity of Boolean functions. Naturally, proving lower bounds in models of graph complexity is even harder than proving lower bounds in circuit and communication complexity models. However, graph–theoretic formulations of lower bound questions have the potential to lead to new insights. In particular, such formulations involving *linear algebraic* representations of graphs have led to new approaches to proving lower bounds on branching program size, formula size, and separation questions in two-party communication complexity. In Section 6, we review such approaches. A linear algebraic representation of a graph places a vector, or more generally a subspace, at each vertex of the graph such that the adjacency relations among vertices can be expressed, or implied, in terms of orthogonality or intersection properties of the associated subspaces. The lowest dimension in which such a representation can be realized for a given graph is the parameter of interest. Such representations have been extensively studied, e.g., in the context of the Shannon capacity of a graph [61], Colin de Verdière's invariant of graphs [45], etc. These and many similar combinatorial-linear algebraic problems are not only of inherent mathematical interest, but also have found numerous applications in algorithms and complexity. In Section 6, we define the affine and projective representations of graphs and pose questions about the lowest dimensions in which explicit graphs can be realized by such representations. Unfortunately, only weak lower bounds — $\Omega(\log n)$ for $n \times n$ explicit bipartite graphs — are known on these dimensions. Lower bounds exceeding $\Omega(\log^3 n)$ on the affine dimension of explicit graphs are needed to derive new lower bounds on the formula size of explicit Boolean functions. Pudlák and Rödl [76] showed that a lower bound on the projective dimension of a bipartite graph implies a lower bound on the branching program size of the associated Boolean function. In relating formula size of bipartite graphs (thereby deriving lower bounds on the associated Boolean functions) to affine dimension, Razborov [85] developed an elegant approach based on rectangle covers of matrices closely related to communication complexity. A rank robustness parameter (given a partially specified matrix, what is the minimum rank of a full matrix that completes it) plays a central role in establishing this connection. This same parameter and the underlying techniques are also used in characterizing the size of span programs in Section 7. Nontrivial lower bounds are known on graph complexity when we restrict the model to be of depth-3 graph formulas. In this case, building on polynomial approximations of the OR function and Forster's lower bound on the sign-rank of an Hadamard matrix, we show [58] $\tilde{\Omega}(\log^3 n)$ lower bounds on the depth-3 complexity of explicit bipartite graphs.

1.7 Span Programs

Karchmer and Wigderson [41] introduced a linear algebraic model of computation called *span programs.* A span program associates a subspace with each of the $2n$ literals of an n variable Boolean function. The result of its computation on a given input x is 1 if and only a fixed nonzero vector, e.g., the all 1's vector, is in the span of the subspaces "activated" by x. The sum of the dimensions of the subspaces is the *size* of the span program. Proving lower bounds on span program size of explicit Boolean functions is a challenging research direction since such results imply lower bounds on Boolean formulas, symmetric branching programs, algebraic proof systems, and secret sharing schemes. The model of span programs realizes the *fusion method* for proving circuit lower bounds [103]. Currently, superpolynomial lower bounds are known only on *monotone* span programs. Monotone span programs are more powerful than monotone Boolean circuits [6]. Hence, proving lower bounds on monotone span programs is more challenging. Early results in this area include a combinatorial criterion on certain bipartite graphs that led to $\Omega(n^{5/2})$ monotone size lower bounds [9]. Subsequently, Babai et al. [6] proved the first superpolynomial monotone lower bound of $n^{\Omega(\log n/\log\log n)}$ exploiting the criterion from [9] but using Paley-type graphs. The most striking result to date on span program size is by Gál [32] who proves a *characterization* of span program size in terms of a rank robustness measure originally introduced by Razborov [85] and referred to above in Section 1.6 and discussed in Section 6. Specializing this characterization to the monotone situation and using previously known lower bounds on the rank robustness measure of certain matrices derived from Paley-type bipartite graphs [85], Gál proved the best known lower bound of $n^{\Omega(\log n)}$ on monotone span programs. We discuss Gál's characterization and her lower bound in Section 7.

2

Matrix Rigidity

2.1 Motivation

Matrix rigidity was introduced by Valiant [98]; a similar notion was used, independently, by Grigoriev [34].

Definition 2.1. The rigidity of a matrix A over a field $\mathbb{F}$, denoted $\mathcal{R}_A^{\mathbb{F}}(r)$, is the least number of entries of A that must be changed in order to reduce the rank of A to a value at most r:

$$\mathcal{R}_A^{\mathbb{F}}(r) := \min\{|C| : \operatorname{rank}_{\mathbb{F}}(A + C) \leq r\},$$

where $|C|$ denotes the number of nonzero entries of C.

When the field of definition is clear from the context, we often omit the superscript $\mathbb{F}$.

Intuitively, the rigidity of A for target rank r is the *Hamming distance* between A and the set of all matrices of rank at most r.

It is easy to see that for any $n \times n$ matrix A over any field, $\mathcal{R}_A^{\mathbb{F}}(r) \leq (n-r)^2$. Valiant [98] also showed that for "almost all" $n \times n$ matrices A, $\mathcal{R}_A^{\mathbb{F}}(r) = (n-r)^2$ if $\mathbb{F}$ is infinite and $\mathcal{R}_A^{\mathbb{F}}(r) = \Omega((n-r)^2/\log n)$

if $\mathbb{F}$ is finite. What do we mean by "almost all" matrices over an infinite field? It is a Zariski open set, i.e., the complement of the solution set of a finite system of polynomial equations. Over a finite field it is interpreted in the usual counting sense, i.e., the fraction of matrices in the set goes to 1 as $n \to \infty$.

The main challenge about rigidity is to prove nontrivial lower bounds for *explicit* matrices. Specifically, Valiant posed the following question:

Open Question 2.1. Find an *explicit* infinite sequence of matrices $\{A_n\}$ such that $\mathcal{R}^{\mathbb{F}}_{A_n}(\epsilon n) \geq n^{1+\delta}$, where $A_n \in \mathbb{F}^{n\times n}$ for all n in an infinite subset of $\mathbb{N}$ and $\epsilon, \delta > 0$ are constants.

A digression on explicitness: The challenge of constructing "explicit" objects with "nice" properties is a familiar one in several fields such as combinatorics, computer science, and coding theory. Examples include expander graphs, hard Boolean functions, and good error-correcting codes. Often a random object is easily shown to be nice and the challenge is to give explicit constructions of objects that are at least approximately nice. Informally, explicitness refers to an algorithmic description of an infinite sequence of nice objects of increasing size. In our context, we may define $\{A_n\}$ to be explicit if for each n in the index set, given $1 \leq i, j \leq n$, there is a Boolean circuit of size polynomial in[1] n to construct $A_n(i,j)$. If we want *uniform* constructibility, we may also insist that testing membership of a natural number in the index set and a construction of the relevant circuit itself be efficiently doable. We also remark that sometimes we say *an* explicit object (as in *a* hard function, *a* good code or *a* rigid matrix) when we really mean an infinite family of them; the usage refers to a "generic" object from an infinite sequence parameterized by "size" n. Also by abuse of notation, we sometimes abbreviate $\{A_n\}$ by A. *End of digression.*

Why should we care about rigid matrices? Strong lower bounds on the rigidity of explicit matrices will lead to significant breakthroughs in complexity theory. In particular, Valiant proved that if $\mathcal{R}^{\mathbb{F}}_{A}(\epsilon n) \geq n^{1+\delta}$, then the linear transformation defined by A cannot be computed by arithmetic circuits of $O(n)$ size and $O(\log n)$ depth in which each gate computes a linear combination of its inputs. Computing linear

[1] A stronger definition would be to require a circuit of size polylog(n). But poly(n) size explicitness is already a challenge.

transformations is a basic computational task. Many fundamental problems such as computing the Fourier transform, polynomial evaluation and interpolation, computing prefix sums and encoding/decoding linear error-correcting codes involve computation of linear transformations. Computation of bilinear transformations, such as matrix multiplication, is closely related to the computation of linear transformations [8]. Proving a superlinear lower bound in a general model on any explicit linear transformation will thus be a big step forward in algebraic complexity. For example, a proof that $\mathcal{R}_A^{\mathbb{F}}(\epsilon n) \geq n^{1+\delta}$, where A is the Fourier transform matrix ($A = (\zeta^{ij})_{0 \leq i,j \leq n-1}$, where ζ is a primitive nth root of unity) would show that there can be no "ultra-fast Fourier transform" algorithm running in $O(n)$ time. Recall that the current best algorithm — the Fast Fourier Transform (FFT) algorithm — uses linear circuits of size $O(n \log n)$ and depth $O(\log n)$.

We now describe the model of linear circuits and Valiant's result relating rigidity to linear circuit complexity.

Definition 2.2. A linear circuit over a field $\mathbb{F}$ is a directed acyclic graph L in which each directed edge is labeled by a nonzero element of $\mathbb{F}$. If g is a gate with in-coming edges labeled by $\lambda_1, \ldots, \lambda_k$ from gates $g_1, \ldots, g_k$, then g computes $v(g) := \lambda_1 v(g_1) + \cdots + \lambda_k v(g_k)$, where $v(g_i) \in \mathbb{F}$ is the value computed at gate g_i.

Suppose L has n input gates (nodes with no in-coming edges) and m output gates (including nodes with no out-going edges). If we denote by $y_1, \ldots, y_m \in \mathbb{F}$ the values computed at the output gates of L starting with the values $x_1, \ldots, x_n \in \mathbb{F}$ at the input gates, then we will have $y = A_L x$, where $A_L \in \mathbb{F}^{m \times n}$; in other words, the circuit L computes the linear transformation given by the matrix A_L.

The size of a linear circuit L is defined to be the number of edges in L. The depth of L is defined to be the length of a longest path from an input node to an output node in L. When depth of L is $\Omega(\log n)$, we assume that each of its internal nodes (gates) has in-degree (fan-in) exactly 2; otherwise, the in-degree is at least 2.

The model of linear circuits is a natural model of computation for computing linear transformations. Furthermore, at the expense of constant factors in size complexity, any arithmetic circuit computing a linear transformation over $\mathbb{C}$ can be turned into a linear circuit [20, Theorem 13.1]. It is trivial to see that any linear transformation $\mathbb{F}^n \to \mathbb{F}^n$ can be computed by a linear circuit of size $O(n^2)$ and depth $O(\log n)$. It is a major challenge in complexity theory to prove superlinear lower bounds on the size of linear circuits, even of logarithmic depth, computing *explicit* linear transformations. In his seminal result [98], Valiant proved a criterion for such a complexity lower bound in terms of matrix rigidity.

Theorem 2.1. Suppose the linear transformation $x \mapsto Ax$ is computed by a linear circuit of size s and depth d in which each gate has fan-in two. Then, for any $t > 1$,

$$\mathcal{R}_A\left(\frac{s\log t}{\log d}\right) \leq 2^{O(d/t)} \cdot n. \tag{2.1}$$

In particular, if $\mathcal{R}_A(\epsilon n) \geq n^{1+\delta}$ for some constants $\epsilon, \delta > 0$, then any linear circuit of logarithmic depth computing $x \mapsto Ax$ must have size $\Omega(n \log\log n)$.

Proof. Proof of this theorem depends on a classical graph–theoretic lemma [27].

Lemma 2.2. Let $G = (V, E)$ be a directed acyclic graph in which all (directed) paths are of length at most d. Then by removing at most $|E|/\log d$ edges, we can ensure that all paths in the remaining graph have length at most $d/2$.

For simplicity, let us assume that d and t are powers of 2. By repeatedly applying the above lemma $\log t$ times, we can identify a set T of edges from the circuit graph G such that (i) $|T| \leq s\log t/\log d$ and (ii) after removing T from the circuit, all paths in the remaining circuit graph G' have length at most d/t. Let V' be the "tails" of the edges in T and let $b_1, \ldots, b_{|V'|}$ be the linear forms (in the inputs x_i) computed at the gates in V'. In the left-over graph G', any given output node is reachable from at most $2^{d/t}$ input nodes since in G' all paths are of length at most d/t.

A linear form a_i computed at an output node i is now a linear combination of the removed "intermediate" nodes b_k, $1 \leq k \leq |V'|$, and at most $2^{d/t}$ variables. Identifying a linear form with the corresponding (row) vector in $\mathbb{F}^n$, we see

$$a_i = \sum_{k=1}^{|V'|} \beta_{ik} b_k + c_i, \quad 1 \leq i \leq n,$$

where $\beta_{ik} \in \mathbb{F}$ and $c_i \in \mathbb{F}^n$ has at most $2^{d/t}$ nonzero entries. In matrix notation,

$$A = B_1 B_2 + C = B + C, \tag{2.2}$$

where $B_1 = (\beta_{ik}) \in \mathbb{F}^{n \times |V'|}$ and B_2 has b_k, $1 \leq k \leq |V'|$, as its rows. Clearly, $\mathrm{rank}(B) \leq |V'| \leq |T| \leq s \log t / \log d$ and each row of C has at most $2^{d/t}$ entries and hence $|C| \leq n2^{d/t}$. It follows that by changing at most $n2^{d/t}$ entries, the rank of A can be reduced to a value at most $s \log t / \log d$. This proves the theorem. □

In addition to the original motivation above, rigidity and many similar matrix functions that measure the "robustness" of the rank function have been discovered to have applications in several other models of computation such as communication complexity, branching programs, span programs, and threshold circuits. Since rigidity-like functions appear to be so fundamental, several researchers [28, 29, 31, 56, 58, 74, 77, 80, 81, 84, 93] studied this topic and explored its connections to other problems in complexity theory, algebra, and combinatorics. By now, many such connections have been discovered.

2.2 Explicit Lower Bounds

Many candidate matrices are conjectured to have rigidity as high as in Valiant's question. Examples include Fourier transform matrices, Hadamard matrices, Cauchy matrices, Vandermonde matrices, incidence matrices of projective planes, etc. Despite intense efforts by several researchers, Valiant's question still remains unsolved. The best known lower bound is $\mathcal{R}_A^{\mathbb{F}}(r) = \Omega\big(\frac{n^2}{r}\log\frac{n}{r}\big)$ proved for various matrices by Friedman [31] and Shokhrollahi et al. [93]; note that this bound reduces to a trivial $\Omega(n)$ when $r = \epsilon n$. We present below the proof from [93]. This proof yields the best known lower bounds for both finite and infinite fields. We will use ideas from the proof in [31], which seem to work only for finite fields, in Section 2.6 in the context of Paturi–Pudlák dimensions.

We will use the following lemma from extremal graph theory about the *Zarankiewicz problem* [16, Section IV.2, Theorem 10].

Lemma 2.3. Let G be an $n \times n$ bipartite graph. If G does not contain $K_{s,s}$ (complete bipartite graph with s vertices on either side) as a subgraph, then the number of edges in G is at most

$$(s-1)^{1/s}(n-s+1)n^{1-1/s} + (s-1)n.$$

Lemma 2.4. Let $r \geq \log^2 n$. If fewer than

$$\frac{n(n-r)}{2(r+1)}\log\frac{n}{r}$$

changes are made to an $n \times n$ matrix A, then some $(r+1) \times (r+1)$ submatrix of A remains untouched.

Proof. Think of A as an $n \times n$ bipartite graph with unchanged entries as edges. Then this bipartite graph has at least

$$n^2 - \frac{n(n-r)}{2(r+1)}\log\frac{n}{r}$$

edges. It is easy to check that, for $\log^2 n \leq r$, this quantity is more than the bound from Lemma 2.3 with $s = r+1$. Thus, the graph must have $K_{r+1,r+1}$ as a subgraph. This means the corresponding $(r+1) \times (r+1)$ submatrix remains untouched. □

This lemma can be applied to prove rigidity lower bounds for matrices in which *all* submatrices of sufficiently large size are nonsingular. We prove some of these bounds below. For simplicity in these bounds, we assume $r \leq n/2$ and replace $n(n-r)$ in Lemma 2.4 by $n^2/2$.

Theorem 2.5. Let $\mathbb{F}$ be a field containing at least $2n$ elements and let $x_1, \ldots, x_n$, and $y_1, \ldots, y_n$ be $2n$ distinct elements such that $(x_i + y_j) \neq 0$ for any i, j. Then, the *Cauchy matrix* C, given by $C := \big(1/x_i + y_j\big)_{i,j=1}^{n}$, has the rigidity lower bound

$$\mathcal{R}_C(r) \geq \frac{n^2}{4(r+1)} \log \frac{n}{r}$$

for $\log^2 n \leq r \leq n/2$.

Proof. It is well-known and easy to prove that the determinant of the Cauchy matrix C is given by

$$\det C = \frac{\prod_{i<j}(x_i - x_j) \prod_{i<j}(y_i - y_j)}{\prod_{i,j}(x_i + y_j)}.$$

Clearly, a $t \times t$ submatrix of a Cauchy matrix is itself a Cauchy matrix. Hence, all submatrices of all sizes of C are nonsingular. If fewer than $(n^2/4(r+1)) \log \frac{n}{r}$ elements of C are changed, then, by Lemma 2.4, at least one $(r+1) \times (r+1)$ submatrix remains unchanged. It follows that the altered matrix has rank at least $(r+1)$. Thus to bring the rank of C down to at most r, at least $(n^2/4(r+1)) \log \frac{n}{r}$ of its entries must be changed. □

Next, we show that matrices with rigidity $\Omega(\frac{n^2}{r} \log \frac{n}{r})$ can be constructed from asymptotically good error correcting codes. The well-known Tsfasman–Vlăduţ–Zink (TVZ) bound on algebraic geometry codes allows us to construct the appropriate codes.

Theorem 2.6. Let q be the square of a prime. Then for every rate R, there exists an infinite sequence $[n_t, k_t, r_t]_{t=1,2,\ldots}$ of codes over $\mathbb{F}_q$ such that the asymptotic rate $\lim_{t\to\infty} k_t/n_t$ is R and the asymptotic relative distance $\lim_{t\to\infty} d_t/n_t$ is δ such that

$$R + \delta \geq 1 - \frac{1}{\sqrt{q} - 1}.$$

We will choose $R = 1/2$ in the above theorem. This gives us, for infinitely many n, a $[2n, n, d]$ code with $d \geq (1 - \epsilon)n$, where $\epsilon := 2/(\sqrt{q} - 1)$. A generator matrix for such a code is a $2n \times n$ matrix over $\mathbb{F}_q$ and can be brought into the standard form $G = [I \mid A]$, where I is the $n \times n$ identity matrix and A is an $n \times n$ matrix over $\mathbb{F}_q$.

Theorem 2.7. Let A be an $n \times n$ matrix over $\mathbb{F}_q$ obtained from a generator matrix of a $[2n, n, (1 - \epsilon)n]$ code as described above. Then, for $\epsilon n \leq r \leq n/4$,

$$\mathcal{R}_A(r) \geq \frac{n^2}{8(r+1)} \log \frac{n}{(2r+1)}.$$

Proof. Let $s := r + 1$. We first claim that every $2s \times 2s$ submatrix of A must have a rank of at least s. Suppose not, and consider a $2s \times 2s$ matrix with s dependent rows. Then, a linear combination of the corresponding s rows of $G = [I \mid A]$ yield a code word of weight $s + n - 2s = n - s \leq (1 - \epsilon)n - 1$. However, this contradicts that the code has minimum distance $(1 - \epsilon)n$.

It follows that unless the changes hit every $2s \times 2s$ submatrix of A, we cannot reduce the rank of A to $\leq r$. By Lemma 2.4, there must be at least $\frac{n^2}{8s} \log \frac{n}{2s-1}$ changes for $\log^2 n \leq 2s \leq n/2$ to hit every $2s \times 2s$ submatrix. Since $s = r + 1$, the claimed rigidity lower bound on A must hold. □

We next consider the Fourier transform matrix.

Theorem 2.8. Let $F = (\omega^{ij})_{i,j=0}^{n-1}$, where ω is a primitive nth root of unity. Then, as n ranges over all prime numbers and $\log^2 n \leq r \leq n/2$,

$$\mathcal{R}_F(r) \geq \frac{n^2}{4(r+1)} \log \frac{n}{r}.$$

Proof. It is well-known that all submatrices of all orders of the Fourier transform matrix of a prime order (character table of $\mathbb{Z}/p\mathbb{Z}$) are nonsingular; a more recent proof can be found in [96]. Thus, the same proof as that of Theorem 2.5 gives the claimed lower bound. □

Remark 2.1. The main step in the above proof, that all minors of the Vandermonde determinant $\left|\omega^{ij}\right|$ are nonzero, was originally proved by Chebotarëv[2] in 1926. Since then, several different proofs of this result appeared; see [96] for some of the references. Tao's proof seems to be the most accessible.

Note that the Fourier transform matrix is a Vandermonde matrix. We also have the following theorem, where instead of powers of ω we use distinct positive reals.

Theorem 2.9. Let $G = (a_i^{j-1})_{i,j=1,n}$, where a_{ij} are distinct *positive reals.* Then, for $\log^2 n \le r \le n/2$,

$$\mathcal{R}_G(r) \ge \frac{n^2}{4(r+1)} \log \frac{n}{r}.$$

Proof. The same proof as above also works using the classical fact that any submatrix of such a matrix is also nonsingular (Descartes rule of signs; (see, e.g., [73, Part V, Section 1, Problem 36]). □

With no conditions on its generators, Shparlinsky (see [56]) proved an $\Omega(n^2/r)$ lower bound for a Vandermonde matrix.

Theorem 2.10. Let $V = (x_i^{j-1})_{i,j=1}^n$ be a Vandermonde matrix with distinct x_i over any sufficiently large field. Then, $\mathcal{R}_V(r) \ge (n-r)^2/(r+1)$.

Proof. For a given r, let $s := \mathcal{R}_V(r)$ be the minimum number of changes made to V so that the altered matrix has rank at most r.

[2] For a fascinating historical account of this and other results by Chebotarëv, see [94].

By an averaging argument, we can select $r + 1$ consecutive columns, $k, \ldots, k + r$, $1 \leq k \leq n - r$, such that the total number of changes in these columns is at most $s(r + 1)/(n - r)$. Ignore any row that has a change in these columns. That gives us $n - s(r + 1)/(n - r)$ rows with no changes in the columns $k, \ldots, k + r$; let's call these the "good" rows. Since the altered matrix has rank at most r, the columns $k, \ldots, k + r$ are linearly dependent, say, with coefficients $\alpha_k, \ldots, \alpha_{k+r}$, not all of which are zero. This means that the polynomial $\sum_{t=0}^{r} \alpha_{k+t} X^t$ has at least $n - s(r + 1)/(n - r)$ roots, namely, the x_i from V that correspond to the good rows selected above. But, this polynomial can have at most r roots. Thus,

$$r \geq n - \frac{s(r + 1)}{(n - r)}.$$

This gives $\mathcal{R}_V(r) = s \geq (n - r)^2/(r + 1)$ and the theorem is proved. □

Another family of explicit matrices conjectured to be highly rigid is that of Hadamard matrices. The best known lower bound on the rigidity of an Hadamard matrix is $\Omega(n^2/r)$ and was first proved by Kashin and Razborov [42]. Recently, a much simpler proof of the same bound for the family of Sylvester-type Hadamard matrices is given by Midrijānis [64]. Recall that a Sylvester matrix $S_k \in \{-1, +1\}^{2^k \times 2^k}$ is recursively defined by

$$S_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad S_k = \begin{bmatrix} S_{k-1} & S_{k-1} \\ S_{k-1} & -S_{k-1} \end{bmatrix}. \tag{2.3}$$

Theorem 2.11. Let $n := 2^k$ and S_k be the $n \times n$ Sylvester-type Hadamard matrix. Then, for r a power of 2, $\mathcal{R}_{S_k}(r) \geq n^2/4r$.

Proof. Let $r = 2^{t-1}$ for $t \geq 1$. Note that if we divide the rows and columns of S_k into (continuous) intervals of length $2r = 2^t$ each, then we get a tiling of S_k by $n^2/4r^2$ submatrices each of which is $\pm S_{k-t}$. It follows that each of these submatrices is nonsingular. Now, suppose there are less than $n^2/4r$ changes made to S_k. Then, at least one of the submatrices in the above tiling must have fewer than r entries. Since this is originally a $2r \times 2r$ full-rank submatrix, the altered submatrix has rank more than r because each change can reduce the rank by at most 1. □

In Section 3, we will see a proof of the same bound for a generalized Hadamard matrix using spectral techniques (cf. [26, Theorem 3.8]).

2.2.1 Limitations of the Combinatorial Proof Techniques

As illustrated by the proofs above, proofs of the best known lower bounds on rigidity of explicit matrices follow essentially the same strategy and consist of two steps: (1) all (or most) square submatrices of these matrices have full- (or close to full) rank; (2) if the number of changes is too small, then one such submatrix remains untouched. We observe [56] that both steps of this approach are severely limited in answering Open Question 2.1.

Any proof relying on the second step in the above strategy, i.e, using an untouched submatrix of rank $\Omega(r)$, is *incapable* of producing a lower bound better than $\Omega\left(\frac{n^2}{r}\log\frac{n}{r}\right)$. Indeed, using Lovász's result [60] relating fractional and integral vertex covers of a hypergraph, we can show that the bound in Lemma 2.4 is asymptotically tight. Specifically, consider the hypergraph whose vertices are the entries of an $n \times n$ matrix and hyperedges are all $(r+1) \times (r+1)$ submatrices. The degree d of this hypergraph is $\binom{n-1}{r}^2$. By regularity, the fractional cover number t^* of this hypergraph is $\binom{n}{r+1}^2 / \binom{n-1}{r}^2 = n^2/(r+1)^2$. Lovasz's bound (greedy algorithm) gives that for the size of an integral cover t, we have $t \le t^*(\ln d + 1)$. Such an integral cover clearly intersects every $(r+1) \times (r+1)$ submatrix. Hence,

$$\frac{n^2}{(r+1)^2}\left(1 + \left(\frac{n-1}{r}\right)^2\right) = O\left(\frac{n^2}{r+1}\log\frac{n}{r}\right)$$

changes suffice.

The best bound provable by the first step of the above approach is limited by linear size superconcentrators (of logarithmic depth). Valiant himself constructed integer matrices in which all submatrices of all sizes are nonsingular and yet having a rigidity of at most $n^{1+o(1)}$ to reduce the rank to $O(n)$. Recall that a *superconcentrator* is a graph

$G = (V, E)$ with n inputs $X \subseteq V$ and n outputs $Y \subseteq V$ such that for all subsets $S \subseteq X$, $|S| = k$, and all subsets $T \subseteq Y$, $|T| = k$, there are k vertex-disjoint paths from S to T. A remarkable result in [72] gives superconcentrators with $|E| = O(n)$ and depth $O(\log n)$.

Theorem 2.12. There exist an $n \times n$ matrices A with integer entries such that all submatrices of A of all sizes are nonsingular, and yet $\mathcal{R}_A(r) = n^{1+o(1)}$ for $r = \epsilon n$ for any $\epsilon > 0$.

Proof. Let G be a superconcentrator with cn edges, input nodes $x_1, \ldots, x_n$ and output nodes $y_1, \ldots, y_n$. Let the vertices of G be $v_1, \ldots, v_m$ in topological order. The fan-in (in-degree) of each v_i is 2. We will define the matrix A by defining a linear circuit computing the linear transformation $x \mapsto Ax$ from the graph G. We will identify each vertex with the linear form in $x_1, \ldots, x_n$ computed at that vertex. It is helpful to think of the circuit as given by a $(2n + m) \times n$ matrix with the columns indexed by the x_j, the first n rows labeled by the x_i, the last n rows by the y_i, and the remaining m rows by v_i. Each row gives a vector in $\mathbb{C}^n$ associated to the linear form in the x_i computed at the corresponding node. Thus, in the first n rows, we have the identity matrix whereas the last n rows give us the desired matrix A.

We define the rows of this matrix inductively. Suppose u_1 and u_2 are already defined and are inputs to a node v. We define the coefficients α and β at the node v so the gate computes $v = \alpha u_1 + \beta u_2$. Choose α and β as follows: for all r-tuples of x_j and all linear forms $w_1, \ldots, w_{r-1}$ already constructed, whenever $\{u_1, w_1, \ldots, w_{r-1}\}$ and $\{u_2, w_1, \ldots, w_{r-1}\}$ are sets of linearly independent forms when restricted to $x_{j_1}, \ldots, x_{j_r}$, $\{\alpha u_1 + \beta u_2, w_1, \ldots, w_{r-1}\}$ is also a linearly independent set of linear forms when restricted to $x_{j_1}, \ldots, x_{j_r}$.

Let $\det(v_{i_1}, \ldots, v_{i_r}; x_{j_1}, \ldots, x_{j_r})$ denote the determinant of the submatrix indexed by the rows $v_{i_1}, \ldots, v_{i_r}$ and columns $x_{j_1}, \ldots, x_{j_r}$. Then by linearity,

$$\begin{aligned}
&\det(v, w_1, \ldots, w_{r-1}; x_{j_1}, \ldots, x_{j_r}) \\
&\quad = \alpha \det(u_1, w_1, \ldots, w_{r-1}; x_{j_1}, \ldots, x_{j_r}) \\
&\qquad + \beta \det(u_2, w_1, \ldots, w_{r-1}; x_{j_1}, \ldots, x_{j_r}).
\end{aligned}$$

If the two determinants on the right-hand side are nonzero, there is a unique value for the ratio $\alpha : \beta$ that is forbidden by the requirement that the left-hand side determinant is nonzero. It follows that for every choice of $w_1, \ldots, w_{r-1}$ and $x_{j_1}, \ldots, x_{j_r}$, at most one choice of the ratio $\alpha : \beta$ is forbidden in defining the coefficients α and β in the gate $v = \alpha u_1 + \beta u_2$. Since there are only finitely many choices for the w_i's and the x_j's, we conclude we can always find integer values for α and β so the condition above is satisfied at v.

By the superconcentrator property of G, from any set of r inputs $x_{j_1}, \ldots, x_{j_r}$ to any set of r outputs $y_{i_1}, \ldots, y_{i_r}$, there are r vertex-joint paths. By considering sets of r parallel nodes at successive levels, we can see by induction that the sub-determinant given by the corresponding rows and the x_j's is nonsingular. We conclude that all submatrices of all orders are nonsingular in the matrix A. Since the graph of the circuit is of linear size, we can apply the graph–theoretic argument in the proof of Theorem 2.1 to express A as $A = B + C$ such that for any $\epsilon > 0$, $\text{rank}(B) \leq \epsilon n$ and $|C| \leq n \cdot 2^{\log^{1-\delta} n}$ for some $\delta = \delta(\epsilon) > 0$. It follows that $\mathcal{R}_A(\epsilon n) = n^{1+o(1)}$ for any $\epsilon > 0$. □

2.3 Sparse Factorizations and Constant Depth Circuits

Since strong lower bounds on rigidity seem to be difficult to obtain, a natural approach is to study restricted variants of the problem. We discuss one such combinatorial restriction in this section and some others later.

For a matrix A, define

$$w_2(A) := \min\{|B| + |C| : \ A = BC\},$$

where the minimum is taken over all matrix pairs B and C such that $A = BC$. Recall that $|B| + |C|$ is the total number of nonzero entries in B and C.

Note that $w_2(A)$ captures the minimum size of a depth-2 linear circuit computing the linear transformation is given by A. We can naturally generalize this definition to w_k and use it to prove lower bounds on depth-k linear circuits. In fact, we will do so in Section 3 (cf. Lemma 3.14 and Theorem 3.15) for the restricted model of *bounded* *coefficient* linear circuits. However, explicit lower bounds on w_k have not yet been used successfully to prove strong superlinear, i.e., $n^{1+\delta}$, lower bounds on constant depth linear circuits in general.

As usual, it is easy to see that for almost $n \times n$ matrices A, $w_2(A) = \Omega(n^2/\log n)$, and the challenge is to find explicit A for which $w_2(A) \geq n^{1+\delta}$ can be proved. The best known lower bound is $\Omega(n \log^2 n/\log\log n)$ and follows from the tight lower bound on depth-2 superconcentrators [82]. This lower bound holds for any matrix in which every square submatrix is nonsingular: examples include the Cauchy matrix $A = (1/(x_i + y_j))$ where x_i and y_j are all distinct and Fourier transform matrices of prime order (cf. Theorems 2.5 and 2.8). Pudlák [74] shows how to derive a lower bound on $w_2(A)$ from a very strong lower bound on $\mathcal{R}_A(r)$ for r in a suitable range.

We note that in the proof of the lower bound on $w_2(A)$ using the lower bound on depth-2 superconcentrators, only *connectivity* properties of the factorization $A = BC$ are used; associate a depth-2 graph with $A = BC$ with inputs being column indices of A (= column indices of C), outputs being row indices of A (= row indices of B), and the middle being the column indices of B (= row indices of C) and note that nonsingularity of a square submatrix A_{ST} of A implies that $|S|$ vertex-disjoint paths must exist between S and T (Menger's theorem). Since the lower bound of [82] is tight for depth-2 superconcentrators, this approach *cannot* yield better lower bounds on $w_2(A)$. This is similar to the limitation that exists even for the original rigidity problem as discussed in Section 2.2.1.

2.4 Rigidity Lower Bounds Using Algebraic Dimensions

We observed in the last two subsections that some commonly used approaches to the rigidity problem face certain fundamental limitations: we call this the *barrier of purely combinatorial approaches* since the proof techniques essentially use only the connectivity (superconcentrator-like) properties and combinatorial structure of submatrices of the candidate matrices appealing very little to their algebraic structure. We believe that any stronger lower bound proof on rigidity will have to penetrate this barrier by delving deeper into the *algebraic* structure of the candidate matrices.

In this subsection, we look at such an attempt. We can prove much stronger, even quadratic, lower bounds on the rigidity of some complex matrices by exploiting algebraic independence (in a limited sense) among their entries. In particular, we will prove a lower bound of $\mathcal{R}_V(r) = \Omega(n^2)$ for $r \leq \epsilon\sqrt{n}$, on the rigidity of Vandermonde matrices $V = (x_i^{j-1})_{1\leq i,j\leq n}$ with *algebraically independent* x_i [56]. We will also prove [59] that $\mathcal{R}_A(\epsilon n) = \Omega(n^2)$ for two matrix families: (i) $A = (\sqrt{p_{jk}})$ and (ii) $A = (\mathrm{e}^{2\pi i/p_{jk}})$, where p_{jk} are the first n^2 primes. These bounds hold over $\mathbb{C}$. These bounds *break through the combinatorial barriers* mentioned above. The result for Vandermondes provides a natural n-dimensional manifold in the space of all $n \times n$ matrices with $\Omega(n^2)$ rigidity for nonconstant r. The results for $(\sqrt{p_{jk}})$ and $(\mathrm{e}^{2\pi i/p_{jk}})$ are the first quadratic lower bounds known on the rigidity of a *non-generic* matrix over any field for rank bound $\Omega(n)$.

The proof for $(\sqrt{p_{jk}})$ uses an algebraic dimension concept introduced by Shoup and Smolensky [91] (SS-dimension, for short). To prove the rigidity lower bound on $(\mathrm{e}^{2\pi i/p_{jk}})$, we use a higher degree generalization of the SS-dimension. The SS-dimension was used in [91] to derive superlinear lower bounds on linear circuits of depths up to $\mathrm{poly}(\log n)$ for linear transformations defined by a Vandermonde matrix and its inverse. In their Vandermonde matrix $V = (x_i^{j-1})$, the x_i are either algebraically independent transcendentals or *superincreasing* integers $(x_i = 2^{n^i})$. We note that Shoup and Smolensky prove these superlinear lower bounds directly, *without* appealing to or proving any rigidity bounds on their matrices.

More generally, algebraic dimension arguments involving square roots of primes and roots of unity of prime orders have been used in the literature to replace generic or random elements to fool “low-degree” computations. They have been used to construct specific polynomials that are hard to compute [7, 20, 52]. Square roots of primes are also used to define hard instances (Swinnerton–Dyer polynomials) for certain polynomial factorization algorithms [101, Section 15.3]. Rational approximations of square roots of primes are used in [25], to reduce randomness in polynomial identity testing based on the Schwartz–Zippel

Lemma. This approach is extended in [23] to prove that the square roots of any set of rationals independent over squares (i. e., linearly independent in the $\mathbb{F}_2$-space $\mathbb{Q}^*/\mathbb{Q}^{*2}$) can be used in place of square roots of primes. The approach in [25] is generalized in [51] using square roots of irreducible polynomials to be applicable over arbitrary fields — in particular over finite fields.

2.4.1 The Shoup–Smolensky Dimensions

In this subsection, we define the various SS-dimensions we use and prove some preliminary lemmas relating them to matrix rank. Shoup and Smolensky originally used Definition 2.3 in [91].

Let p be a nonnegative integer, $P = (a_1, \dots, a_p)$ a sequence of elements ($a_i \in \mathbb{C}$), and $t \geq 0$.

Definition 2.3. The *restricted SS-dimension* of degree t of P over $\mathbb{Q}$, denoted by $\mathcal{D}_t(P)$, is the rank over $\mathbb{Q}$ of the set of all the $\binom{p}{t}$ products $\prod_{j=1}^{t} a_{i_j}$, where $1 \leq i_1 < i_2 < \cdots < i_t \leq p$.

Definition 2.4. The *unrestricted SS-dimension* of degree t of P over $\mathbb{Q}$, denoted by $\mathcal{D}_t^*(P)$, is the rank over $\mathbb{Q}$ of the set of all the $\binom{p+t-1}{t}$ products $\prod_{j=1}^{t} a_{i_j}$, where $1 \leq i_1 \leq i_2 \leq \cdots \leq i_t \leq p$.

So, in both definitions we take t-term products of elements of P; in the restricted case, repeated entries are not permitted.

The following inequalities are immediate.

$$\mathcal{D}_t(P) \leq \mathcal{D}_t^*(P) \tag{2.4}$$

$$\mathcal{D}_t(P) \leq \binom{p}{t} \tag{2.5}$$

$$\mathcal{D}_t^*(P) \leq \binom{p+t-1}{t}. \tag{2.6}$$

Let now $Q = (b_1, \dots, b_q)$ and $PQ = (a_i b_j : i = 1, \dots, p;\ j = 1, \dots, q)$. The following is immediate.

Observation 2.13.

$$\mathcal{D}_t^*(PQ) \leq \mathcal{D}_t^*(P)\mathcal{D}_t^*(Q). \tag{2.7}$$

(Note that the analogous statement for the *restricted* SS-dimension is false for all $t \geq 2$.)

We define the SS-dimensions $\mathcal{D}_t(A)$ and $\mathcal{D}_t^*(A)$ of a $k \times \ell$ matrix A over $\mathbb{C}$ as the corresponding dimensions of the list (in some order) of the $k\ell$ entries of the matrix.

Even though it is immaterial in what order we list the $k\ell$ entries of a matrix, simple inequalities link the SS-dimensions to matrix multiplication and matrix rank.

The following is an immediate consequence of Observation 2.13.

Observation 2.14. Let A and B be matrices over $\mathbb{C}$ such that the product AB is defined. Then,

$$\mathcal{D}_t^*(AB) \leq \mathcal{D}_t^*(A)\mathcal{D}_t^*(B). \tag{2.8}$$

Corollary 2.15. If the $k \times \ell$ matrix A has rank r (over $\mathbb{C}$, its field of definition), then

$$\mathcal{D}_t^*(A) \leq \binom{kr+t-1}{t}\binom{\ell r+t-1}{t}. \tag{2.9}$$

We indicate the proof. A can be written as $A = BC$, where B is a $k \times r$ matrix and C an $r \times \ell$ matrix over $\mathbb{C}$. Hence, a combination of Observation 2.14 and the trivial bound (2.6) yields the result.

Remark 2.2. By noting that one of the factors, B or C, can be taken to contain an $r \times r$ identity submatrix, it is possible to slightly improve the bound (2.9) to $\binom{kr+t-1}{t}\binom{\ell r-r^2+t}{t}$.

Definition 2.5. Let $P = (a_1, \ldots, a_m)$ be a sequence of complex numbers and $T \subseteq \mathbb{N}^m$ be a set of vectors of nonnegative integers. The *generalized SS-dimension* $\mathcal{D}_T(P)$ is defined to be the rank over $\mathbb{Q}$ of the set of monomials $\prod_{i=1}^m a_i^{e_i}$, where $\mathbf{e} := (e_1, \ldots, e_m) \in T$:

$$\mathcal{D}_T(P) := \dim \left\langle \prod_{i=1}^m a_i^{e_i} \ : \ \mathbf{e} \in T \right\rangle_{\mathbb{Q}}. \tag{2.10}$$

Note that we obtain $\mathcal{D}_t(P)$ by letting $T = \{\mathbf{e} : \sum e_i = t \text{ and } e_i \leq 1\}$ and we obtain $\mathcal{D}_t^*(P)$ by letting $T = \{\mathbf{e} : \sum e_i = t\}$.

We will also use the following special case of $\mathcal{D}_T(P)$.

Let $\mathbf{t} = (t_1, \ldots, t_m)$ be a vector of nonnegative integers. We define $\mathcal{D}_{\mathbf{t}}(P)$ by letting T consists of vectors whose ith coordinate is at most t_i:

$$\mathcal{D}_{\mathbf{t}}(P) := \dim \left\langle \prod_{i=1}^m a_i^{e_i} \ : \ 0 \leq e_i \leq t_i \right\rangle_{\mathbb{Q}}. \tag{2.11}$$

Notation: For a vector $\mathbf{t}$, let

$$\sigma(\mathbf{t}) = \sum_{i=1}^m t_i, \quad \text{and} \quad \pi(\mathbf{t}) = \prod_{i=1}^m (t_i + 1).$$

Note that, in general, $\mathcal{D}_T(P) \leq |T|$ and, in particular, $\mathcal{D}_{\mathbf{t}}(P) \leq \pi(\mathbf{t})$. The weaker upper bound $\mathcal{D}_{\mathbf{t}}(P) \leq \binom{|P|+\sigma(\mathbf{t})}{\sigma(\mathbf{t})}$ is sometimes useful as in the following lemma, i.e., analogous to Corollary 2.15.

Lemma 2.16. If a $k \times \ell$ matrix A has rank at most r, then

$$\mathcal{D}_{\mathbf{t}}(A) \leq \binom{kr + \sigma(\mathbf{t})}{\sigma(\mathbf{t})}\binom{\ell r + \sigma(\mathbf{t})}{\sigma(\mathbf{t})}.$$

2.4.2 Quadratic Lower Bounds on Rigidity

The first application of SS-dimension that we prove is a quadratic lower bound on the rigidity of a generic Vandermonde matrix when the target rank is $O(\sqrt{n})$.

Theorem 2.17. Let $V = (x_i^{j-1})_{1 \le i,j \le n}$ be a Vandermonde matrix, where the x_i are *algebraically independent over* $\mathbb{Q}$. Then

$$\mathcal{R}_V(r) \ge \frac{n(n - c \cdot r^2)}{2},$$

where $c > 0$ is an absolute constant.

In particular, there exists an $\epsilon > 0$ such that for every $r \le \epsilon\sqrt{n}$, $\mathcal{R}_V(r) \ge n^2/4$.

Proof. Let C be a matrix with the smallest number of nonzero entries such that $V - C$ has rank r. Thus, $\mathcal{R}_V(r) = |C|$, number of nonzero entries of C.

Denote by s_i the number of changes in the ith row of V. Let s denote the average of the s_i, $s := (s_1 + \cdots + s_n)/n$. Thus, $|C| = n \cdot s$.

There must be at least $n/2$ rows of V with no more than $2s$ changes in each of those rows. Fix $n/2$ such rows and call them "good" rows. (The factor of 2 here is not optimal; see Remark 2.3.)

We claim that

$$\mathcal{D}_{n/2}(V - C) \ge (n - 2s)^{n/2}. \tag{2.12}$$

To prove this claim, consider the products formed by taking unchanged entries of good rows, one entry per row. They are of the form $x_{i_1}^{j_1} x_{i_2}^{j_2} \cdots x_{i_t}^{j_t}$, where $i_1, \ldots, i_t$ are good rows and each j_k has at least $(n - 2s)$ possibilities. By the algebraic independence of the x_i, these products are linearly independent over $\mathbb{Q}$. Hence, the claim follows.

Since $V - C$ has rank r, Corollary 2.15 gives an upper bound on $\mathcal{D}_{V-C}$:

$$\mathcal{D}_t(V - C) \le \binom{nr + t}{t}^2. \tag{2.13}$$

Combining (2.12) and (2.13), we get,

$$(n - 2s)^{n/2} \le \binom{nr + n/2}{n/2}^2.$$

Since $\binom{n}{k} \le (ne/k)^k$, $\quad (n-2s)^{n/2} \le \left(\dfrac{(nr+n/2)e}{n/2}\right)^{2\cdot n/2}$.

$$\text{Hence,} \quad (n-2s) \le ((1+2r)e)^2$$
$$\le c\cdot r^2, \quad \text{for some constant } c>0.$$
$$\text{This gives us,} \quad s \ge \frac{(n-c\cdot r^2)}{2}.$$

Since $\mathcal{R}_V(r) = |C| = n\cdot s$, we have proved the theorem. □

Remark 2.3. The bound of Theorem 2.17 can be improved by a constant factor with a more careful choice of the parameters. Indeed, as pointed out by Tom Hayes (private communication), by selecting $t \approx (\sqrt{2}ern)^{2/3}$ in the proof above, it can be shown that $\mathcal{R}_V(r) \ge n^2 - O(n^{5/3}r^{2/3})$.

Theorem 2.18. Let A be an $n\times n$ matrix over $\mathbb{C}$ and $0\le r\le n$. Suppose, $\mathcal{D}_{nr}(A) = \binom{n^2}{nr}$, i.e., all products of nr distinct entries of A are linearly independent over $\mathbb{Q}$. Then,

$$\mathcal{R}_A(r) \ge n(n-16r). \tag{2.14}$$

Proof. Let C be a matrix such that $\mathcal{R}_A(r) = \mathrm{wt}(C)$ and rank $(A-C)\le r$. By Corollary 2.15, we have

$$\mathcal{D}_t(A-C) \le \mathcal{D}_t^*(A-C) \le \binom{nr+t}{t}^2. \tag{2.15}$$

In order to obtain a lower bound on $\mathcal{D}_t(A-C)$, let us consider all t-wise products of elements of A not affected by C. By assumption, these (as well as the products of all other t-tuples from A) are linearly independent over $\mathbb{Q}$; therefore,

$$\mathcal{D}_t(A-C) \ge \binom{n^2-\mathrm{wt}(C)}{t}. \tag{2.16}$$

Set $t = nr$ and combine inequalities (2.15) and (2.16). We use the inequality $\binom{n}{k} \geq (n/k)^k$.

$$\begin{aligned} \binom{n^2 - \mathrm{wt}(C)}{nr} &\leq \binom{2nr}{nr}^2 \\ \left(\frac{n^2 - \mathrm{wt}(C)}{nr}\right)^{nr} &\leq \left(2^{2nr}\right)^2 = 16^{nr} \\ \mathrm{wt}(C) &\geq n^2 - 16nr. \end{aligned}$$

We conclude that $\mathcal{R}_A(r) \geq n(n - 16r)$. □

Corollary 2.19. Let $P = \left(\sqrt{p_{ij}}\right)$, where p_{ij} are distinct primes for $1 \leq i, j \leq n$. Then, $\mathcal{R}_P(r) \geq n(n - 16r)$.

In particular, $\mathcal{R}_P(n/17) \geq n^2/17$.

To obtain Corollary 2.19, we combine Theorem 2.18 with the following result cf. ([12, 4, Ex. 2.1.41]). An integer is *square-free* if it is not divisible by the square of any prime number.

Theorem 2.20. The square roots of all positive square-free integers are linearly independent over $\mathbb{Q}$. In particular, for *distinct primes* $p_1, \ldots, p_m$, $[\mathbb{Q}(\sqrt{p_1}, \ldots, \sqrt{p_m}) : \mathbb{Q}] = 2^m$.

The next rigidity lower bound uses the Generalized SS-dimension.

Theorem 2.21. Let

$$Z := \left(\mathrm{e}^{2\pi i/p_{jk}}\right)_{1 \leq j,k \leq n},$$

where p_{jk} are the first n^2 distinct primes. Then, for $0 \leq r \leq n$, we have

$$\mathcal{R}_Z(r) \geq n(n - 9r),$$

assuming n is sufficiently large.

In particular, $\mathcal{R}_Z(n/10) \geq n^2/10$.

Proof. Let C be a matrix such that $\mathrm{wt}(C) = \mathcal{R}_Z(r)$ and $\mathrm{rank}(Z - C) \leq r$.

Let $m := n^2 - \mathrm{wt}(C)$ be the number of entries of Z "untouched" by C and $p_1, \ldots, p_m$ be the corresponding primes. Let

$$\begin{aligned} P &:= (\mathrm{e}^{2\pi i/p_1}, \ldots, \mathrm{e}^{2\pi i/p_m}) \quad \text{and} \\ \mathbf{t} &:= (p_1 - 1, \ldots, p_m - 1). \end{aligned}$$

We will consider $\mathcal{D}_{\mathbf{t}}(P)$. To get a lower bound, we use the following facts.

Lemma 2.22. $[\mathbb{Q}(\mathrm{e}^{2\pi i/n}) : \mathbb{Q}] = \varphi(n)$.

For a proof of this lemma, see, e.g., [50, Ch VI, Theorem 3.1].

Lemma 2.23. $\mathbb{Q}(\mathrm{e}^{2\pi i/a_1}, \ldots, \mathrm{e}^{2\pi i/a_m}) = \mathbb{Q}(\mathrm{e}^{2\pi i/\mathrm{lcm}(a_1,\ldots,a_m)})$.

This lemma is easy to prove.

It follows that

$$\mathcal{D}_{\mathbf{t}}(P) = [\mathbb{Q}(\mathrm{e}^{2\pi i/p_1}, \ldots, \mathrm{e}^{2\pi i/p_m}) : \mathbb{Q}] = \varphi(p_1 \cdot \cdots \cdot p_m) = \prod_{i=1}^{m} (p_i - 1). \tag{2.17}$$

On the other hand, the elements of P are entries of the matrix $Z - C$ of rank at most r. Thus, by Lemma 2.16, we have the upper bound

$$\mathcal{D}_{\mathbf{t}}(P) \leq \binom{nr + \sigma(\mathbf{t})}{nr}^2. \tag{2.18}$$

Since $\prod_{i=1}^m (p_i - 1) \geq m!$ and $\sigma(\mathbf{t}) = \sum_{i=1}^m (p_i - 1) = O(mn^2 \log n)$, and using the inequality $\binom{a}{b} \leq a^b$, we obtain from (2.17) and (2.18),

$$m! \leq \mathcal{D}_{\mathbf{t}}(P) \leq (nr + O(mn^2 \log n))^{2nr}. \tag{2.19}$$

Since $nr, m \leq n^2$, after taking logarithms of the above inequality, we obtain $m \log m \leq c_2 nr \log n$ for some constant $c_2 > 0$. Since $\mathcal{R}_A(r) \leq (n - r)^2$ in general, we note that $m = n^2 - \mathcal{R}_Z(r) \geq 2nr - r^2 \geq 2n - 1$ for $1 \leq r \leq n$. Using this in the last inequality, we have $m \leq 9nr$ assuming n is sufficiently large. □

2.5 Quadratic Lower Bounds on Linear Circuits

Combining Corollary 2.19 or Theorem 2.21 with Theorem 2.1, we get the following circuit lower bound.

Theorem 2.24. Let $A = \left(\sqrt{p_{jk}}\right)$ or $A = \left(\mathrm{e}^{2\pi i/p_{jk}}\right)$, where p_{jk} are the first n^2 primes for $1 \le j,k \le n$. Then, any linear circuit of logarithmic depth computing $x \mapsto Ax$ must have size $\Omega(n \log\log n)$.

We will now see that using the SS-dimension, we can obtain quadratic lower bounds on linear circuits significantly improving those in Theorem 2.24 via Valiant's criterion. Indeed, such lower bounds were already proved by Lickteig [52]. Nevertheless, we present a proof of the circuit lower bound using the generalized SS-dimension since we feel it is simple, intuitive, and fits well within the framework of rigidity.

The following lemma generalizes an inequality from [91] on SS-dimension.

Lemma 2.25. Let L be a linear circuit over $\mathbb{C}$ computing the linear transformation $x \mapsto Ax$. Let s denote the size and d denote the depth of L. Define $\bar{s} := s/d$. For a set $T \subseteq \mathbb{N}^{n^2}$, define $\sigma(T) := \max_{\mathbf{t} \in T} \sum_{i=1}^{n^2} t_i$ and let $\mathcal{D}_T(A)$ be as in Definition 2.5. Then,

$$\mathcal{D}_T(A) \le \binom{\bar{s} + \sigma(T)}{\bar{s}}^d. \tag{2.20}$$

Proof. Let $(a_1, \ldots, a_m)$ be the sequence of entries of the matrix A in some order, where $m := n^2$. By abuse of notation, we use A to also denote this sequence. We want to estimate the $\mathbb{Q}$-linear dimension spanned by monomials of the form $a_1^{e_1} \cdots a_m^{e_m}$, where $\mathbf{e} \in T$.

Arrange the circuit L into d levels, where a gate g is at level k if the longest path to g from an input has k edges; the input nodes are at level 0. For $1 \le k \le d$, let L_k denote the set of labels on the edges that feed into gates at level k. Our first observation is that an entry $a := a_{ij}$ of the matrix A is equal to the sum of labels of all the paths from input j to output i, where the label of a path is defined to be the product of all the labels on its edges. Here and below, we suppress many subscripts for brevity. The intended summation ranges should be clear from the context. Thus, we have

$$a = \sum_p \lambda_1 \cdots \lambda_d,$$

where the sum ranges over all paths p from input j to output i and $\lambda_k \in L_k \cup \{1\}$ are "labels" on edges of the path p. We include 1 here, since an edge may not go between two consecutive levels and hence 1 may be used as λ_k for the "skipped" k. Hence, a monomial in the a's is given by

$$\begin{aligned}\prod_{i=1}^m a_i^{e_i} &= \prod_{i=1}^m \sum_{p_i} (\lambda_{i1} \cdots \lambda_{id})^{e_i} \\ &= \sum (\lambda_{11}^{e_1} \cdots \lambda_{m1}^{e_m}) \cdots (\lambda_{1d}^{e_1} \cdots \lambda_{md}^{e_m}).\end{aligned}$$

Note that each monomial $\lambda_{1k}^{e_1} \cdots \lambda_{mk}^{e_m}$ has labels from $L_k \cup \{1\}$ and may have repetitions. Since $\mathbf{e} \in T$, we may thus view it as a monomial of total degree at most $\sigma(T)$ on $|L_k|$ variables. Let $s_k = |L_k|$. There are at most $\binom{s_k+\sigma(T)}{s_k}$ such monomials. Hence, each monomial $\prod_{i=1}^m a_i^{e_i}$ in our space is an integer linear combination of products of d such monomials from L_k for $1 \le k \le d$. It follows that

$$\begin{aligned}\mathcal{D}_T(A) &\le \prod_{k=1}^d \binom{s_k + \sigma(T)}{s_k} \\ &\le \binom{\frac{1}{d}\sum_{k=1}^d s_k + \sigma(T)}{\frac{1}{d}\sum_{k=1}^d s_k}^d,\end{aligned}$$

where the last inequality is a consequence of the log-concavity of $f(x) = \binom{x+c}{x}$. Since $s = \mathsf{size}(L) = \sum_{k=1}^d s_k$, we have proved the claim. □

Corollary 2.26. Let $Z := \left(\mathrm{e}^{2\pi i/p_{jk}}\right)_{j,k=1}^n$ where p_{jk} are the first n^2 primes. Then, any arithmetic circuit computing the linear transformation $x \mapsto Zx$ must have size $\Omega(n^2)$.

Proof. Let $m := n^2$. We will apply the special case $\mathcal{D}_{\mathbf{t}}(Z)$ of $\mathcal{D}_T(Z)$ with $\mathbf{t} = (p_1 - 1, \ldots, p_m - 1)$. From the proof of Theorem 2.21, we know that $\mathcal{D}_T(Z) \geq m!$. On the other hand, $\sigma(T) = \sigma(\mathbf{t}) = \sum_{i=1}^{m}(p_i - 1) = \Theta(m^2 \log m)$. Since $s \leq n^2$ always, we have $\bar{s} \ll \sigma(T)$ and the easy estimate $\binom{\bar{s}+\sigma(\mathbf{t})}{\bar{s}} \leq (2\sigma(T))^{\bar{s}}$. Using these in (2.20), we get

$$m! \leq (2\sigma(T))^{\bar{s}d} \leq (c \cdot m^2 \log m)^s.$$

Taking logarithms on both sides of this inequality, we obtain $s = \Omega(m) = \Omega(n^2)$. □

Corollary 2.27. Let $P = \left(\sqrt{p_{ij}}\right)$, where p_{ij} are distinct prime integers. Then, any arithmetic circuit computing $x \mapsto Px$ must have size $\Omega(n^2/\log n)$.

Proof. Let $m := n^2$. Let $T := \{(e_1, \ldots, e_m) \ : \ 0 \leq e_i \leq 1\}$. From Theorem 2.20, $\mathcal{D}_T(P) = 2^m$. Since $\sigma(T) = m \geq s$, we obtain from (2.20),

$$2^m \leq (2m)^{\bar{s}d} = (2m)^s.$$

Taking logarithms proves the claim. □

The lower bounds in the corollaries above do not depend on the depth of the circuits. However, Shoup and Smolensky [91] exploit the dependence of (their special case of) inequality (2.20) on depth to derive lower bounds of $\Omega(dn^{1+1/d})$ for $d \leq \log n$, and $\Omega(n \log n/(\log d - \log\log n))$ for larger d, for Vandermonde matrices with algebraically independent generators.

2.6 Paturi–Pudlák Dimensions

Paturi and Pudlák [70] introduced certain dimensions of a subspace (which we call here inner and outer dimensions) that refine the notion of rigidity. They may give rise to new techniques to prove lower bounds on linear circuits.

A vector $v \in \mathbb{F}^n$ is said to be *s-sparse* if the number of nonzero coordinates in v (often denoted $\mathrm{wt}(v)$ or $|v|$) is at most s.

Let $V \subseteq \mathbb{F}^n$ be a subspace, where $\dim V = m$.

Definition 2.6. The *inner dimension* $d_V(s)$ for sparsity parameter s of V is the dimension of the largest subspace of V that can be generated by s-sparse vectors. In other words,

$$d_V(s) := \max\{\dim(U \cap V) : \\ U \subseteq \mathbb{F}^n \text{ has an } s\text{-sparse basis and} \dim U \leq \dim V\}.$$

Definition 2.7. The *outer dimension* $D_V(s)$ for sparsity parameter s of V is the dimension of the smallest subspace W that contains V and that can be generated by s-sparse vectors. In other words,

$$D_V(s) := \min\{\dim W : W \text{ has an } s\text{-sparse basis and } V \subseteq W\}.$$

Given a matrix $A \in \mathbb{F}^{m\times n}$, we let $\langle A\rangle$ to denote the row-space of A. Note that $\langle A\rangle$ is a subspace of $\mathbb{F}^n$ of dimension at most m. By abuse of notation, we denote $d_{\langle A\rangle}(s)$ and $D_{\langle A\rangle}(s)$ simply by $d_A(s)$ and $D_A(s)$.

To relate inner dimension to rigidity, we define the following version of rigidity of a matrix when the changes are counted on a per row basis.

$$\rho_A(s) := \min\{\text{rank}(A - C) \ : \text{ Each row of } C \text{ is } s\text{-sparse}\}.$$

Clearly, if $\rho_A(s) \leq r$, then $\mathcal{R}_A(r) \leq ns$. But lower bounds on $\rho_A(s)$ are sufficient to prove lower bounds on linear circuits *via* Valiant's criterion.

Lemma 2.28. For $A \in \mathbb{F}^{m\times n}$,

$$\rho_A(s) \geq \text{rank}\, A - d_A(s).$$

Proof. Let C be a matrix achieving $\rho_A(s)$, so $\text{rank}(A - C) = \rho_A(s)$ and each row of C has at most s nonzero entries.

Trivially, $\langle C\rangle + \langle A\rangle = \langle C\rangle + \langle A - C\rangle$.

Considering the left-hand side,

$$\dim(\langle C\rangle + \langle A\rangle) = \dim\langle A\rangle + \dim\langle C\rangle - \dim(\langle A\rangle \cap \langle C\rangle).$$

Considering the right-hand side,

$$\dim(\langle C\rangle + \langle A - C\rangle) \leq \dim\langle C\rangle + \dim\langle A - C\rangle = \dim\langle C\rangle + \rho_A(s).$$

Thus, we have $\dim\langle A\rangle - \dim(\langle A\rangle \cap \langle C\rangle) \leq \rho_A(s)$. Since $\dim\langle A\rangle = \mathrm{rank}(A)$ and $d_A(s) \geq \dim(\langle A\rangle \cap \langle C\rangle)$, we are done. □

Lemma 2.29. For any finite-dimensional subspace V,

$$d_V(s) + D_V(s) \geq 2\dim V.$$

Proof. Let $D := D_V(s)$ and $d := d_V(s)$. Let $W \supseteq V$ be an optimal subspace realizing $D_V(s)$ with a sparse basis $\{w_1, \ldots, w_D\}$. Consider the subspace $U \subseteq W$ with the basis $\{w_1, \ldots, w_m\}$, where $m = \dim V$. Clearly, $U + V \subseteq W$ and $\dim(U \cap V) \leq d$. Hence,

$$D \geq \dim(U + V) = \dim U + \dim V - \dim(U \cap V) \geq 2m - d. \qquad \square$$

While there seems to be no obvious relation between $D_A(s)$ and rigidity, we can use $D_A(s)$ directly to prove lower bounds on log-depth linear circuits using a proof technique similar to that of Theorem 2.1.

Theorem 2.30. Suppose, the linear transformation $x \mapsto Ax$ for $A \in \mathbb{F}^{n\times n}$ is computed by a linear circuit of size L and depth k. Then, there exists a constant $c > 1$ such that for any integer $t > 0$,

$$D_A(c^{k/t}) \leq n + \frac{L \log t}{\log k}.$$

In particular, if, for some constant $\epsilon > 0$, $D_A(n^\epsilon) \geq n + \omega(n/\log\log n)$, then $x \mapsto Ax$ cannot be computed by log-depth linear circuits of size $O(n)$.

Paturi and Pudlák use a counting argument to show that for a random m-dimensional subspace V of $\mathbb{F}^n$, *where* $\mathbb{F}$ *is finite*, $|\mathbb{F}| = q$,

$$D_V(s) \geq n\left(1 - \frac{s \log_q n}{m}\right),$$

with high probability.

The proof of Theorem 2.17 actually yields a lower bound of $\rho_V(s) = \Omega(\sqrt{n-s})$ for a generic Vandermonde matrix. However, no nontrivial upper (lower) bounds on d (D) of a generic Vandermonde matrix are currently known.

In the following theorem, we use codes over $\mathbb{F}_2$ to prove nontrivial lower (upper) bounds on $D_V(s)$ $(d_V(s))$. It can be easily generalized to codes over an arbitrary finite field $\mathbb{F}_q$ in an obvious way (the constants depend on q).

Theorem 2.31. Let C be an $[n,k,d]$ linear code over $\mathbb{F}_2$. Then, for $s \leq d/2$,

$$D_C(s) \geq k + \frac{d}{2s}\log\left(\frac{2sk}{d}\right), \quad \text{and} \tag{2.21}$$

$$d_C(s) \leq k - \frac{d}{2s}\log\left(\frac{2sk}{d}\right). \tag{2.22}$$

Proof. The proof is based on the following main claim: *there exists a* $[D_C(s),k,d/s]$ *code* B*.* Consider a space W of dimension $D := D_C(s)$ containing C and having an s-sparse basis $\{w_1,\ldots,w_D\}$. Every vector $x \in C$ is a linear combination of vectors from W, i.e., there is a $y \in \mathbb{F}^D$ (unique since w_i are a basis) such that $x = \sum_{i=1}^{D} y_i w_i$. Let B be the set of all such y for all $x \in C$. Clearly, B is a subspace of dimension k. We note that the minimum weight of a nonzero vector in B is at least d/s and this will prove the claim. Indeed, let y^* be a minimum weight nonzero vector in B and let $x^* = \sum_{i=1}^{D} y_i^* w_i$. Since w_i are linearly independent, $x^* \neq 0$ and by the property of C, x^* is of weight at least d. Since each w_i has weight at most s, at least d/s of the coordinates y_i^* must be nonzero.

We apply the sphere packing bound for the code B. The Hamming balls of radius $d/2s$ around vectors in B must be nonintersecting and hence their union must contain at most 2^D vectors:

$$2^k \sum_{j=0}^{d/2s} \binom{D}{j} \leq 2^D.$$

Let $\Lambda(D, d/2s) := \sum_{j=0}^{d/2s} \binom{D}{j}$ denote the volume of a Hamming ball of radius $d/2s$ in $\mathbb{F}_2^D$. Then, we conclude from the above that $D \geq k + \log_2 \Lambda(D, d/2s)$. Since $\Lambda(D, d/2s) \geq \binom{D}{d/2s} \geq \binom{k}{d/2s} \geq (2sk/d)^{d/2s}$, (2.21) follows.

The bound on $d_C(s)$ in (2.22) is proved using similar argument. Let U be a k-dimensional subspace realizing $d_C(s)$. Thus, $\dim(U \cap C) = d_C(s)$. Similar to the claim above, we can show that C contains a $[k, d_C(s), d/s]$ code B'. Applying the sphere packing bound to B', we have

$$2^{d_C(s)} \Lambda(k, d/2s) \leq 2^k.$$

Simplifying as before, we obtain (2.22). □

Friedman uses (2.22) to prove the first nontrivial lower bound on the rigidity of a matrix over a finite field. Note that we proved essentially the same lower bound in Theorem 2.7.

Corollary 2.32. Let $A \in \mathbb{F}^{k \times n}$ be the generator matrix of an asymptotically good error correcting code C. Then, for $0 < r < k/2$,

$$\mathcal{R}_A(r) = \Omega\left(\frac{n^2}{r} \log \frac{n}{r}\right).$$

Proof. Since C is asymptotically good, we know that $k = \Omega(n)$ and the minimum distance d of C is also $\Omega(n)$. Thus, it follows from (2.22) that $d_C(s) \leq k - (d/2s)\log(2ks/d)$. From Lemma 2.28, $\rho_A(s) \geq k - d_C(s) \geq (d/2s)\log(2ks/d) = \Omega(n \log s/s)$. In other words, $s = \Omega\left(\frac{n}{r} \log \frac{n}{r}\right)$ changes must be made in *each row* to reduce rank of A to be at most r. Thus, $\mathcal{R}_A(r) \geq ks = \Omega\left(\frac{n^2}{r} \log \frac{n}{r}\right)$. □

3

Spectral Methods to Study Rank Robustness

As mentioned before, many functions similar to rigidity have several applications in complexity theory. To encompass this broad range of applications, it is helpful to think of *rank robustness* in general. By changing the nature of changes or by changing the definition of distance between a given matrix and the set of low-rank matrices, we get various notions of rank robustness. For example, when the matrix is over $\mathbb{R}$ or $\mathbb{C}$, we may consider ℓ_2-distance (instead of Hamming distance as in the original definition of rigidity) or we may consider only sign-preserving changes. Many results [21, 28, 29, 42, 55, 57, 58, 75, 83] successfully use such notions of rank robustness to derive lower bound results in various models. We discuss some of them in this section and the next.

The "nonsmoothness" of Hamming distance seems to make it difficult to attack the general rigidity problem using well-studied mathematical properties of matrices. On the other hand, several smoother variants of rigidity, e.g., $\ell_2\text{-Rig}, \mathrm{Rig}, \mathrm{msv}, \mathrm{Vol}$, studied in this section, seem to be easier to get a handle on using classical techniques from matrix analysis [13, 33, 95]. In particular, these robustness functions of matrix rank are easily bounded, or even characterized, in terms of the singular values of a matrix. More importantly, all these functions are very useful in deriving interesting lower bounds on computational complexity. Significant applications of this approach are lower bounds on the bounded coefficient complexity of linear and bilinear transformations, e.g., Fourier transform and matrix multiplication.

3.1 Singular Value Decomposition

We begin by recalling some basic facts about singular values of a complex matrix and state an important variational inequality about them.

Definition 3.1. Let $A \in \mathbb{C}^{m \times n}$. Then,

- The `Frobenius norm` of A is
$$\|A\|_F := \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2}.$$
- The `Spectral norm` of A, $\|A\|_2$, usually denoted $\|A\|$, is defined by
$$\|A\| := \max_{x \neq 0} \frac{\|Ax\|}{\|x\|},$$
where $\|v\| = v^* v$ for $v \in \mathbb{C}^n$.
- The ith `Singular value`, $\sigma_i(A)$, is defined by
$$\sigma_i(A) = \sqrt{\lambda_i(AA^*)}, \;\; 1 \leq i \leq \min\{m,n\},$$
where $\lambda_i(AA^*)$ denotes the ith largest eigenvalue of AA^*.

Fact 3.1. Let $A \in \mathbb{C}^{m \times n}$. Then, there exist unitary matrices $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ such that

$$U^* AV = \mathrm{diag}(\sigma_1, \ldots, \sigma_p), \quad \text{where } p = \min\{m,n\}.$$

A proof of this fact can be found in [33, Section 2.5].

The Courant–Fischer minimax theorem gives a characterization of singular values.

Fact 3.2. For $i = 1,\dots,\min\{m,n\}$,

$$\sigma_i(A) = \max_{\dim(S)=i} \min_{0\neq x\in S} \frac{\|Ax\|}{\|x\|} = \min_{\dim(T)=n-i+1} \max_{0\neq x\in T} \frac{\|Ax\|}{\|x\|},$$

where S ranges over all i-dimensional subspaces and T ranges over all $(n-i+1)$-dimensional subspaces of $\mathbb{C}^n$. □

Clearly, rank, Frobenius norm, and spectral norm of A are invariant under unitary transformations. Thus, Facts 3.1 and 3.2 imply the following.

Fact 3.3. Let $A \in \mathbb{C}^{n\times n}$. Then,

(1) $\mathrm{rank}(A) = r$ if and only if

$$\sigma_1(A) \geq \cdots \geq \sigma_r(A) > \sigma_{r+1}(A) = \cdots = \sigma_n(A) = 0.$$

(2) $\|A\|_F^2 = \sigma_1^2(A) + \cdots + \sigma_n^2(A)$.

(3) $\|A\| = \sigma_1(A)$. □

The following important inequality [38] is often useful.

Theorem 3.4 (Hoffman-Wielandt Inequality). Let A and B be matrices in $\mathbb{C}^{n\times n}$. Then,

$$\sum_{i=1}^{n} [\sigma_i(A) - \sigma_i(B)]^2 \leq \|A - B\|_F^2.$$

Hoffman and Wielandt [38] prove their result for eigenvalues of normal matrices using the Birkhoff–von Neumann characterization of doubly stochastic matrices. The theorem for singular values as stated here can be found in [33, Section 8.3].

3.2 Variants of Rigidity

We will first consider two variants of the rigidity function based on the ℓ_2-distance to low-rank matrices.

Definition 3.2. Let $A \in \mathbb{C}^{n\times n}$ and $\theta \geq 0$.

- *ℓ_2 variant of rigidity:* This is a variant of rigidity that measures the ℓ_2-norm of changes.

  $$\ell_2\text{-Rig}_A^2(r) := \min_B \left\{ \sum_{i,j} |a_{ij} - b_{ij}|^2 : \text{rank}(B) \leq r \right\}.$$

- *Bounded changes variant of rigidity:* This is a restricted variant of rigidity where the changes are bounded in absolute value by θ.

  $$\mathcal{R}_A(r,\theta) := \min_B \{|A - B| : \text{rank}(B) \leq r,\ \forall i,j\ |a_{i,j} - b_{ij}| \leq \theta\}.$$

  Recall that $|C|$ denotes the number of nonzero entries of the matrix C.

We can immediately characterize ℓ_2-Rig in terms of singular values.

Lemma 3.5. For any matrix $A \in \mathbb{C}^{n\times n}$,

$$\ell_2\text{-Rig}_A^2(r) = \sum_{i=r+1}^{n} \sigma_i(A)^2.$$

Proof. The lower bound is immediate from the Hoffman-Wielandt inequality, but it can also be proved directly.

Let B be a matrix achieving the minimum in the definition of $\ell_2\text{-Rig}_A^2(r)$. Let N_0 denote the null space of B. Choose a unit vector $\nu_0 \in N_0$ that achieves $\max_{0\neq x\in N_0} \frac{\|Ax\|}{\|x\|}$. By Fact 3.2, since $\dim(N_0) = n - r$, $\sigma_{r+1}^2(A) \leq \|A\nu_0\|^2$. Continuing for $i = 1,\ldots,(n-r-1)$, let $N_i := \{x \in N_{i-1} : x \perp \nu_0,\ldots,x \perp \nu_{i-1}\}$. Choose ν_i to be a unit vector in N_i

that achieves $\max_{0\neq x\in N_i}\frac{\|Ax\|}{\|x\|}$. Since $\dim(N_i)=n-r-i$, by Fact 3.2, $\sigma^2_{r+i+1}(A)\leq\|A\nu_i\|^2$. By orthonormality of $\nu_0,\ldots,\nu_{n-r-1}$ and since $B\nu_i=0$ for $0\leq i\leq(n-r-1)$, it follows that

$$\|A-B\|_F^2\geq\sum_{i=0}^{n-r-1}\|(A-B)\nu_i\|^2=\sum_{i=0}^{n-r-1}\|A\nu_i\|^2\geq\sum_{i=0}^{n-r-1}\sigma^2_{r+1+i}(A).$$

Equality follows by considering the rank r matrix $B:=\sum_{i=1}^{r}\sigma_i(A)u_i\cdot v_i^*$, where u_i and v_i^* are the i-th column and the i-th row of U^* and V, respectively, in Fact 3.1. □

We will prove a lower bound on $\mathcal{R}_A(r,\theta)$ for a specific matrix in the next subsection.

The following lemma gives a useful lower bound on the rank of a submatrix in terms of the spectral norm.

Lemma 3.6. For any submatrix B of a matrix A, $\mathrm{rank}(B)\geq\|B\|_F^2/\|A\|^2$.

Proof. From Fact 3.3, $\mathrm{rank}(B)\geq\|B\|_F^2/\|B\|^2$. Since B is a submatrix of A, $\|B\|\leq\|A\|$. □

Using Hoffman–Wielandt inequality, we can prove a generalization of the above lemma.

Lemma 3.7. Let $A,B\in\mathbb{C}^{n\times n}$. Then,

$$\mathrm{rank}(B)\geq\frac{\Re\langle A,B\rangle}{\|A\|\|B\|},$$

where $\langle A,B\rangle:=\mathrm{Tr}(AB^*)$ and $\Re x$ denotes the real part of a complex number x.

Proof. Using Theorem 3.4,

$$\|A-B\|_F^2\geq\sum_{i=1}^{n}(\sigma_i(A)-\sigma_i(B))^2$$

$$
\begin{aligned}
&= \|A\|_F^2 + \|B\|_F^2 - 2\sum_{i=1}^{n}\sigma_i(A)\sigma_i(B), \\
&\qquad\qquad\qquad \text{using Fact 3.3(ii)} \\
&\geq \|A\|_F^2 + \|B\|_F^2 - 2\ \mathrm{rank}(B)\,\|A\|\,\|B\|, \\
&\qquad\qquad\qquad \text{using Fact 3.3(i) and (iii).}
\end{aligned}
$$

Observe that for any matrix M, $\|M\|_F^2 = \mathrm{Tr}(MM^*)$.

Using this in the last inequality above, we get

$$
\begin{aligned}
2\ \mathrm{rank}(B)\,\|A\|\,\|B\| &\geq \|A\|_F^2 + \|B\|_F^2 - \|A-B\|_F^2 \\
&= \mathrm{Tr}(AB^*) + \mathrm{Tr}(BA^*) \\
&= 2\,\Re\,\mathrm{Tr}(AB^*),
\end{aligned}
$$

and the lemma is proved. □

3.2.1 Lower Bounds for a Generalized Hadamard Matrix

We now prove lower bounds on the three variants of rigidity for a "Generalized Hadamard Matrix."

Definition 3.3. An $n \times n$ *complex* matrix H is called a *Generalized Hadamard* matrix if (i) $|h_{ij}| = 1$ for all $1 \leq i, j \leq n$, and (ii) $HH^* = n\,I_n$, where H^* is the conjugate transpose of H and I_n is the $n \times n$ identity matrix.

Two important example of a generalized Hadamard matrix are (i) the *Sylvester type* Hadamard matrix (defined in (2.3)), and (ii) the Fourier transform matrix (see Theorem 2.8).

We note that for a generalized Hadamard matrix H, $\sigma_i(H) = \sqrt{n}$ for all $i, 1 \leq i \leq n$.

We first prove a lower bound on the rigidity of a generalized Hadamard matrix. This proof is from [26].

Theorem 3.8. For any generalized $n \times n$ Hadamard matrix H,

$$\mathcal{R}_H(r) \geq \frac{n^2}{4r}.$$

Proof. Let R be the minimum number of changes that brought the rank of H down to r. By a simple averaging argument, we can find $2r$ rows of H that contain a total of at most $2rR/n$ changes. If $n \leq 2rR/n$, then $R \geq n^2/2r$ and we are done. Hence, we can assume that $n - 2rR/n > 0$. Consider the $(n - 2rR/n)$ columns that contain no changes in the above set of rows. We thus get a $2r \times (n - 2rR/n)$ submatrix B that contains no changes and hence is a submatrix of H. By definition of R, this submatrix must have rank at most r. Applying Lemma 3.6, we get $r \geq \operatorname{rank}(B) \geq 2r(n - 2rR/n)/n$, since $\|B\|_F^2$ is exactly the number entries in B. Rearranging this inequality, we get $R \geq n^2/4r$. □

Remark 3.1. Kashin and Razborov [42] also use spectral methods to prove an $\Omega(n^2/r)$ lower bound on the rigidity of a generalized Hadamard matrix. The essential claim in their paper is that a *random* $k \times k$ submatrix of a generalized Hadamard matrix has rank $\Omega(k)$.

Corollary 3.9. (to Lemma 3.5) For a generalized Hadamard matrix H, $\ell_2\text{-Rig}_H^2(r) = n(n-r)$.

Lemma 3.10. For a generalized Hadamard matrix H,

$$\mathcal{R}_H(r,\theta) \geq \frac{n^2(n-r)}{(\theta+1)(2n + r(\theta-1))}.$$

In particular,

(1) If $r(\theta - 1) \leq n$, then $\mathcal{R}_H(r,\theta) \geq n(n-r)/3(\theta+1)$.
(2) If $r(\theta - 1) \geq n$, then $\mathcal{R}_H(r,\theta) \geq n^2(n-r)/3r(\theta^2 - 1)$.

Proof. We will apply the Hoffman–Wielandt inequality (or the argument from the proof of Lemma 3.5) to H and a scaled version βB of the altered matrix B. Since $\operatorname{rank}(B) \leq r$, $\sigma_i(B) = 0$ for all $i > r$. Thus,

$$\|H - \beta B\|_F^2 \geq \sum_{i=r+1}^{n} \sigma_i^2(H) = n(n-r).$$

On the other hand, denoting the number of changes, i.e., $\mathrm{wt}(H - B)$, by R,

$$\begin{aligned}\|H - \beta B\|_F^2 &= \|(1-\beta)H - \beta(B - H)\|_F^2 \\ &\le (1-\beta)^2(n^2 - R) + (1+\beta\theta)^2 R \\ &= (1-\beta)^2 n^2 + R\left[(1+\beta\theta)^2 - (1-\beta)^2\right],\end{aligned}$$

where the inequality above follows from $\beta \ge 0$ (which we can assume w.l.o.g.) and $|b_{ij} - h_{ij}| \le \theta$.

Combining this upper bound with the lower bound $\|H - \beta B\|_F^2 \ge n(n-r)$,

$$R \ge \frac{n(n-r) - (1-\beta)^2 n^2}{(1+\beta\theta)^2 - (1-\beta)^2}.$$

Choosing $\beta = r/n$ and manipulating, we get

$$R \ge \frac{n^2(n-r)}{(\theta+1)(2n + r(\theta-1))}.$$

Now, (1) and (2) simply follow from this inequality using $r(\theta - 1) \le n$ and $r(\theta - 1) \ge n$, respectively. □

Remark 3.2.

- de Wolf [26] proves a lower bound very similar to the one in the lemma above using quantum arguments. As he observes, this bound captures the bound from [57] as (1) and the bound from [42] as (2).
- Pudlák [75] uses determinantal arguments to give a general lower bound, for $r \le n/2$,

$$\mathcal{R}_A(r,\theta) \ge (n-r)\left(\frac{|\det(A)|}{r^{r/2}}\right)^{2/(n-r)} \theta^{-O(1)}.$$

  In particular, he gets a bound of $\mathcal{R}_H(r,\theta) \ge \theta^{-O(1)} n(n-r)$ for a generalized Hadamard matrix. However, Razborov [86] explains how to obtain Pudlák's bound using the Hoffman–Wielandt inequality.

3.2.2 Geometric Variants of Rigidity

Raz [83] defines a geometric rigidity, given below, that is somewhat similar to ℓ_2-rigidity (Definition 3.2). He uses bounds on this function in his remarkable superlinear lower bound on matrix multiplication in the model of bilinear circuits with bounded coefficients. Subsequently, Bürgisser and Lotz [21] use the same function to prove superlinear lower bounds for cyclic convolution, polynomial multiplication, and polynomial division with remainder in the same model.

Definition 3.4. For a matrix $A \in \mathbb{C}^{m\times n}$,

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

where $a_i \in \mathbb{C}^m$ denotes the ith column of A and $\mathrm{dist}(x,V) := \min_{v\in V} \|x - v\|$ is the ℓ_2-norm in $\mathbb{C}^m$. The minimum is taken over all r-dimensional subspaces of $\mathbb{C}^m$.

Geometrically, if we think of A as a set of n points in $\mathbb{C}^m$, then $\mathrm{Rig}_r(A)$ seeks to minimize the maximum distance of a point in A from an r-dimensional subspace. For this reason, we often refer to $\mathrm{Rig}_r(A)$ as the *geometric rigidity* of A.

We first observe that $\mathrm{Rig}_r(A)$ and ℓ_2-rigidity we discussed before are related as follows.

Lemma 3.11. For $A \in \mathbb{C}^{m\times n}$ and for $1 \le r \le m$,

$$\ell_2\text{-}\mathrm{Rig}_A(r) \ge \mathrm{Rig}_r(A) \ge \frac{\ell_2\text{-}\mathrm{Rig}_A(r)}{\sqrt{n}}.$$

Proof. First, we prove the right-hand side inequality.

Let V be an r-dimensional subspace achieving $\mathrm{Rig}_r(A)$. Let P^*P the projection matrix for V, i.e., for $a \in \mathbb{C}^m$, P^*Pa gives the projection of a onto V. Note that P is an $r \times m$ matrix and that $\mathrm{dist}(a,V) = \|a - P^*Pa\|$. Now, define $B := P^*PA$, i.e., columns of B are projections of columns of A onto V. Clearly, $B \in \mathbb{C}^{m\times n}$ and, furthermore,

$\operatorname{rank} B \le r$. Hence,

$$\begin{aligned}\ell_2\text{-}\mathrm{Rig}_A^2(r) &\le \|A - B\|_F^2 \\ &= \sum_{i=1}^{n} \|a_i - P^* P a_i\|_2^2 \\ &\le n \max_i \operatorname{dist}^2(a_i, V) \\ &= n\,\mathrm{Rig}_r^2(A).\end{aligned}$$

To prove the left-hand side inequality, let B be a matrix achieving $\ell_2\text{-}\mathrm{Rig}_A(r)$, i.e., $\ell_2\text{-}\mathrm{Rig}_A^2(r) = \|A - B\|_F^2$ and $\operatorname{rank} B \le r$. Now, let V be an r-dimensional subspace containing the column space of B. Note that $\operatorname{dist}(a_i, V) \le \operatorname{dist}(a_i, b_i) = \|a_i - b_i\|$. Hence, we have

$$\begin{aligned}\ell_2\text{-}\mathrm{Rig}_A^2(r) &= \sum_{i=1}^{n} \|a_i - b_i\|_2^2 \\ &\ge \sum_{i=1}^{n} \operatorname{dist}^2(a_i, V) \\ &\ge \max_i \operatorname{dist}^2(a_i, V) \\ &\ge \mathrm{Rig}_r^2(A).\end{aligned}$$

□

Corollary 3.12. For a generalized Hadamard matrix H, $\mathrm{Rig}_r(H) \ge \sqrt{n - r}$.

The "volume" of a matrix was earlier used by Morgenstern [65] for proving lower bounds for the Fourier transform. The following definition from [21] is useful in generalizing Morgenstern's technique.

Definition 3.5. For a matrix $A \in \mathbb{C}^{m \times n}$ and an integer $r, 1 \le r \le m$,

(1) The r-volume $\mathrm{Vol}_r(A)$ of A is defined by

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

where the maximum is taken over all subsets $I \subseteq [n]$ of r columns of A and A_I denotes the submatrix of A consisting of columns indexed by I. Note that $\det(A_I^* A_I)^{1/2}$ is the volume of the parallelepiped defined by the columns of A_I.

(2) The rth mean square volume of A, $\mathrm{msv}_r(A)$ is defined by

$$\mathrm{msv}_r(A) := \left(\sum_{|I|=r} \det(A_I^* A_I) \right)^{1/2},$$

where I and A_I are as in (1) above.

Remark 3.3. The nice property of $\mathrm{msv}_r(A)$ defined in (2) above is that it is a unitarily invariant norm of a matrix A. In particular, it is well-known [13] that

$$\mathrm{msv}_r^2(A) = e_r(\sigma_1^2(A), \ldots, \sigma_m^2(A)),$$

where e_r is the rth elementary symmetric polynomial in m variables.

The next lemma connects $\mathrm{Vol}_r(A)$ and $\mathrm{Rig}_r(A)$:

Lemma 3.13.

$$\mathrm{Vol}_r(A) \geq (\mathrm{Rig}_r(A))^r.$$

Proof. We will pick columns $v_1, \ldots, v_r$ from A as follows. Pick v_1 to be a column with the largest length, i.e., that maximizes $|a_i^* a_i|$. In general, pick v_i that maximizes the distance, among the columns of A, to the subspace $\langle v_1, \ldots, v_{i-1} \rangle$. Clearly, a lower bound on $\mathrm{Vol}_r(A)$ is obtained by considering the submatrix of A given by $v_1, \ldots, v_r$. The corresponding determinant is at least $\|v_1'\|^2 \cdot \|v_2'\|^2 \cdots \|v_r'\|^2$, where $v_i - v_i'$ is the projection of v_i onto $\langle v_1, \ldots, v_{i-1} \rangle$ and hence $\|v_i'\| = \mathrm{dist}(v_i, \langle v_1, \ldots, v_{i-1} \rangle)$. By our choice of v_i, $\|v_i'\| \geq \mathrm{dist}(v_{i+1}, \langle v_1, \ldots, v_{i-1} \rangle) \geq \mathrm{dist}(v_{i+1}, \langle v_1, \ldots, v_i \rangle) = \|v_{i+1}'\|$. We conclude that $\mathrm{Vol}_r(A) \geq (\|v_r'\|)^r$.

On the other hand, consider the r-dimensional subspace $V = \langle v_1, \ldots, v_r \rangle$. Clearly, $\mathrm{Rig}_r(A) \leq \mathrm{dist}(a_i, V)$. By the choice of v_i again, and imagining continuing the selection for one more step, we note that $\mathrm{dist}(a_i, V) \le \|v'_{r+1}\| \le \|v'_r\|$. Hence, we have $\mathrm{Rig}_r(A) \le \|v'_r\|$.

It follows that $\mathrm{Vol}_r(A) \ge (\|v'_r\|)^r \ge (\mathrm{Rig}_r(A))^r$ proving the lemma. □

3.3 Bounded Coefficient Complexity of Linear Transformations

In studying the complexity of algebraic problems such as computing linear/bilinear transformations, polynomial evaluation/interpolation, counting the number of arithmetic steps (addition, multiplication, etc.) is a natural mathematical measure. In particular, this model allows multiplications by arbitrary scalars in the ground field (e.g., numbers of arbitrary value and precision if working over $\mathbb{R}$ or $\mathbb{C}$). But, as Morgenstern [65] and Chazelle [24] have observed, many practical algorithms for computing linear transformations such as FFT do not use multiplications by scalars of unbounded value. This implies that in the underlying linear circuits, the coefficients used at each gate computing a linear transformation are bounded by a constant. Morgenstern introduced the model of bounded coefficients and showed that any linear circuit with bounded coefficients computing the Fourier transform must have size $\Omega(n \log n)$. Since then, several authors [24, 57, 67, 75] have studied bounded coefficient complexity of linear transformations. In the same vein, bounded coefficient complexity of bilinear transformations has been studied in [21] (cf. Section 3.4) and [83].

Morgenstern used determinantal arguments to prove an $\Omega(n \log n)$ lower bound on the bounded coefficient complexity of the Fourier transform. We will give a more general proof of the statement based on [83]. We will also give a proof due to Pudlák which also uses bounds on the determinant, but in a different way from Morgenstern. His proof has the nice feature that it gives the strongest bounds known to date on *constant depth* circuits as well. The main lemma for his result is the following.

Lemma 3.14. Let $A \in \mathbb{C}^{n \times n}$. Suppose, $A = A_1 \cdots A_k$ (note: we place no restrictions on the dimensions of the A_i except for the obvious ones for the multiplications to make sense). Then,

$$|\det(A)| \leq \left(\frac{\|A_1\|_F^2}{n}\right)^{n/2} \cdots \left(\frac{\|A_k\|_F^2}{n}\right)^{n/2}.$$

Proof. The proof is by induction on k. The base case $k = 1$ follows from Hadamard inequality and the AM-GM inequality. So, consider $k > 1$ and assume that $A_1 \in \mathbb{C}^{n\times\ell}$ and $A_2 \in \mathbb{C}^{\ell\times\ell'}$, where w.l.o.g. $\ell \geq n$ (if $\ell < n$, then A is singular and there is nothing to prove). Furthermore, the rows of A_1 must be linearly independent. There exists a unitary matrix $Q \in \mathbb{C}^{\ell\times\ell}$ that maps the span of rows of A_1 into the span $\langle e_1,\ldots,e_n\rangle$. Let $B_1 := A_1Q$ and $B_2 := Q^*A_2$. Clearly, $A = B_1 \cdot B_2 \cdots A_k$. Since the columns $n + 1,\ldots,\ell$ of B_1 are all 0's (by definition of Q), we can retain the above equality if we replace rows $n + 1,\ldots,\ell$ of B_2 by all 0's rows. It follows that $A = B_1' \cdot B_2' \cdots A_k$, where B_1' (B_2') is the $n \times n$ ($n \times \ell'$) matrix obtained by omitting the last $\ell - n$ columns (rows) from B_1 (B_2). By applying the induction hypothesis to $B_2' \cdots A_k$, we obtain the upper bound

$$|\det(B_2' \cdots A_k)| \leq \left(\frac{\|B_2'\|_F^2}{n}\right)^{n/2} \cdots \left(\frac{\|A_k\|_F^2}{n}\right)^{n/2}.$$

Also from Hadamard inequality and the AM-GM inequality, $|\det(B_1')| \leq \left(\|B_1'\|_F^2/n\right)^{n/2}$. Combining these inequalities with the facts that $\det(A) = \det(B_1')\det(B_2' \cdots A_k)$, $\|A_1\|_F^2 = \|B_1\|_F^2 = \|B_1'\|_F^2$, and $\|A_2\|_F^2 = \|B_2\|_F^2 \leq \|B_2'\|_F^2$, we complete the proof of the lemma. □

In a *synchronous* circuit, edges go from a given level only to the next, i.e., all paths from inputs to outputs have the same length. Thus, a synchronous linear circuit may be expressed as a product of as many matrices as its depth. Pudlák [75] proves the following.

Theorem 3.15. The number of edges S in any *synchronous* depth-d linear circuit over $\mathbb{C}$ using coefficients bounded in absolute value by c computing the linear transformation $x \mapsto Ax$ given by $A \in \mathbb{C}^{n\times n}$ must satisfy

$$S \geq c^{-2} dn |\det(A)|^{2/dn}.$$

In particular, for a generalized Hadamard matrix H, we get alower bound of $c^{-2} dn^{1+1/d}$.

Remark 3.4. When the depth is $O(\log n)$, this gives an $\Omega(n \log n)$ lower bound, when restricted to synchronous circuits, for a generalized Hadamard matrix and in particular for the *Fourier transform.* This matches, asymptotically when c is a constant, the best known algorithm for computing the Fourier transform in the bounded coefficient model of linear circuits of logarithmic depth. The proof of Morgenstern, discussed later, however, gives a much better lower bound of $\log\det(A)/\log 2c$ *with no assumptions on the depth.*

Proof. A depth-d synchronous circuit can be viewed as consisting of d layers of intermediate nodes with ℓ_i nodes on level i (inputs at level 0, outputs at level d). The edges from level i to level $i+1$ define an $\ell_i \times \ell_{i+1}$ matrix A_{i+1}, where $\ell_0 = \ell_d = n$. We thus have $A = A_1 \cdots A_d$. We apply Lemma 3.14 to this factorization of A. Now, note that if s_i is the number of edges from level $i-1$ to level i, then, $\|A_i\|_F^2 \leq s_i c^2$. Using this in Lemma 3.14, we get

$$\begin{aligned} |\det(A)| &\leq \left(\frac{c^{2d} s_1 \cdots s_d}{n^d}\right)^{n/2} \\ &\leq \left(\frac{c^2}{n}\right)^{dn/2} \left(\frac{s_1 + \cdots + s_d}{d}\right)^{dn/2} \\ &\leq \left(\frac{c^2 S}{dn}\right)^{dn/2}. \end{aligned}$$

Here, we used the that $S = s_1 + \cdots + s_d$. The theorem readily follows from this last inequality. □

The bound from the theorem is minimized (as a function of d) for $d = 2\ln\det A/n$ and we get the following.

Corollary 3.16. Any *synchronous* linear circuit with coefficients bounded by c computing the linear transformation $x \mapsto Ax$ must have at least $2\mathrm{e} \ln \det A / c^2$ edges.

Both Raz [83] and Bürgisser and Lotz [21] also generalize the classical determinant-based technique [65] for proving lower bounds on linear circuits with bounded coefficients and express the lower bound in terms of the rigidity functions given in Definition 3.5.

Theorem 3.17. Let C be a linear circuit with coefficients bounded by $\theta \geq 1$, computing a linear transformation $x \mapsto A^* x$. Then, for $1 \leq r \leq n$,

(1) $\mathrm{size}(C) = \Omega(\log_{2\theta} \mathrm{Vol}_r(A))$.
(2) $\mathrm{size}(C) = \log_{2\theta} \mathrm{msv}_r(A) - O(n)$.

Proof. Let $s := \mathrm{size}(C)$. Sort the gates of C in a topological order and let g_i denote the ith gate in this order where $g_{-n+1}, \ldots, g_0$ are the input nodes and $g_1, \ldots, g_s$ are the linear gates such that $g_i = \lambda_{i1} g_{i1} + \lambda_{i2} g_{i2}$ with $i_1, i_2 < i$ in this order. Each g_i clearly computes a linear combination $z_i := v_i^* x$ of the inputs $x \in \mathbb{C}^n$. We will slightly abuse the notation and let C denote the matrix in $\mathbb{C}^{n \times (s+n)}$ whose ith column is v_i. By induction on j, we prove that $\mathrm{Vol}_r(C_j) \leq (2\theta)^j$ for the submatrix C_j given by the first $n + j$ columns of C. Since A is a submatrix of C, we have $\mathrm{Vol}_r(C) \geq \mathrm{Vol}_r(A)$ and we are done.

In the base case, $j = 0$, we have the identity matrix and the claim is trivial. For $j > 0$, let us observe that $v_j = \lambda_k v_k + \lambda_l v_l$ for $k, l < j$ and $|\lambda_k|, |\lambda_l| \leq 1$. We claim that $\mathrm{Vol}_r(C_j) \leq 2\theta \, \mathrm{Vol}_r(C_{j-1})$. Let B_I denote a submatrix of C_j given by the r columns indexed by elements of a set $I \subseteq [n]$, $|I| = r$, such that $\mathrm{Vol}_r^2(C_j) = \det(B_I^* B_I)$. We will show that $\det(B_I^* B_I) \leq 4\theta^2 \, \mathrm{Vol}_r^2(C_{j-1})$. If $j \notin I$, then by induction hypothesis, $\det(B_I^* B_I) \leq \mathrm{Vol}_r^2(C_{j-1})$. So, assume $j \in I$ and let $v_j = \alpha u + \beta w$, where $u = v_k$ and $w = v_l$ for $k, l < j$.

Let P (Q) be the same as B_I except that the column v_j of B_I is replaced by u (w, respectively). Then, it is easy to see by linearity of the determinant that

$$\begin{aligned}\det(B_I^*B_I) = &\ \alpha^*\alpha \det(P^*P) + \alpha^*\beta \det(P^*Q)\\ &+ \alpha\beta^* \det(Q^*P) + \beta^*\beta \det(Q^*Q).\end{aligned}$$

The first and the fourth determinants are clearly upper bounded by $\mathrm{Vol}_r^2(C_{j-1})$ since u and w are columns of C_{j-1}. The third and fourth determinants are conjugates of each other. We observe that $|\det(P^*Q)| \le \max\{|\det(P^*P)|, |\det(Q^*Q)|\}$. In fact $\det(P^*Q) = u^*w\times$ (an expression involving determinant polynomials in $v_s^*v_t$ for $s,t < j$). Now, $|u^*w| \le \max\{\|u\|^2, \|w\|^2\}$. Since each of the scalars in front of the four determinants in the above formula is bounded in absolute value by θ^2, we conclude that $|\det(B_I^*B_I)| \le 4\theta^2 \max\{|\det(P^*P)|, |\det(Q^*Q)|\}$ which by induction hypothesis is bounded above by $4\theta^2\,\mathrm{Vol}_r^2(C_{j-1})$.

It follows that $\mathrm{Vol}_r(C) \le (2\theta)^{\mathrm{size}(C)}$. Since $\mathrm{Vol}_r(A) \le \mathrm{Vol}_r(C)$, we have proved (1).

The bound in (2) follows from observing that

$$\binom{n}{r}^{-1} \mathrm{msv}_r^2(A) \le \mathrm{Vol}_r^2(A) \le \mathrm{msv}_r^2(A). \qquad \square$$

Combining Theorem 3.17, Lemma 3.13, and Remark 3.3, we obtain the following lower bound in terms of Rig and singular values of A, generalizing the classical lower bound due to Morgenstern [65].

Corollary 3.18. Let C be a linear circuit (no restrictions on depth) using coefficients bounded above in absolute value by $\theta \ge 1$. If C computes the linear transformation $x \mapsto Ax$ for $A \in \mathbb{C}^{m\times n}$, then, for $1 \le r \le n$,

(1) $\mathrm{size}(C) \ge r \log_{2\theta} \mathrm{Rig}_r(A)$.
(2) $\mathrm{size}(C) \ge \frac{1}{2} \log_{2\theta}(e_r(\sigma_1^2(A), \ldots, \sigma_n^2(A))) - O(n)$.

In particular, if A is a generalized Hadamard matrix, $\mathrm{size}(C) = \Omega(n \log n)$.

3.4 Bounded Coefficient Complexity of some Bilinear Transformations

In the previous subsections, we saw how rigidity-like functions can be used to prove lower bounds on linear circuits. In this subsection, we see applications of these functions to lower bounds on *bilinear* circuits. The first such application was discovered by Raz [83] who proved a remarkable lower bound of $\Omega(n^2 \log n)$ for multiplying two $n \times n$ matrices in the model of *bounded coefficient bilinear circuits.* Subsequently, Bürgisser and Lotz [21] followed a similar approach to prove lower bounds on convolution, polynomial multiplication, and polynomial division with remainder.

Definition 3.6. A bilinear circuit over a field $\mathbb{F}$ on two disjoint sets of variables X and Y works as follows:

(1) A linear part computes linear forms $L_i(X)$ and $L'_j(Y)$ in the variables X and Y, respectively.
(2) A bilinear part uses multiplication gates to compute products $L_i(X) * L'_j(Y)$ for i and j.
(3) Another linear part computes linear forms on the outputs of the multiplication gates.

In this section, we will only consider bilinear circuits over $\mathbb{C}$.

A bilinear circuit *with bounded coefficients* uses scalars bounded in absolute value by 1 in all scalar multiplications in the linear parts above. Note that the only nonscalar multiplications used are in the second part.

3.4.1 Bounded Coefficient Complexity of Matrix Multiplication

Note that given matrices X and Y, the product XY is a bilinear transformation of (X,Y). Hence, it can be computed by a bilinear circuit. On the other hand, every arithmetic circuit computing a bilinear transformation *over an infinite field* can be simulated by a bilinear circuit with at most a constant factor of loss in complexity [20]. This reduction also preserves the property of using bounded coefficients.

Before giving the technical details of the proof, we give an intuitive overview of Raz's proof. Let $B(X,Y)$ be a bilinear circuit with bounded coefficients computing the matrix product XY. Let $L(X) = (L_i(X))$ and $L'(Y) = (L'_j(Y))$ be the linear transformations computed in the first part. Note that L and L' are linear transformations on n^2 variables X and Y, respectively. If $\mathrm{Rig}_r(L')$ (or $\mathrm{Rig}_r(L)$) is at least n^ϵ for $r = \Omega(n^2)$, then we obtain an $\Omega(n^2 \log n)$ lower bound on the linear part itself by Corollary 3.18 and we are done. Hence, we may assume that $\mathrm{Rig}_r(L')$ is no more than n^ϵ for some small constant $\epsilon > 0$. Raz uses this to show the existence of a matrix Y such that: (i) $\mathrm{Rig}_r(Y) = \Omega(\sqrt{n})$ for $r = \Omega(n)$, and (ii) $\forall j$, $|L'_j(Y)| \leq O(n^\epsilon \log n)$ for $0 < \epsilon < 1$. To see the crucial role of this matrix, observe that if we fix Y, the circuit B computes a linear transformation on the matrix X and in fact becomes a linear circuit B_Y. Moreover, it is easy to see that the linear transformation computed by this circuit is given by $I \otimes Y$ and a lower bound on $\mathrm{Rig}_r(Y)$ from (i) implies one on $\mathrm{Rig}_{nr}(I \otimes Y)$. So, if B_Y were to involve only bounded coefficients, then we could use the results of the previous subsection to derive a lower bound on $\mathrm{size}(B_Y)$ and hence on $\mathrm{size}(B)$. However, even though B uses bounded coefficients, its restriction B_Y in general need not use bounded coefficients since the values $L'_j(Y)$ used by the multiplication gates in B (which become the scalars used in B_Y when we fix Y) can be unbounded. Here, Raz employs a neat trick using (ii). First, note that multiplication by a scalar θ can be replaced by $O(\log\theta)$ additions followed by multiplication by a scalar of absolute value at most 1. This by itself, however, cannot be directly applied since we will have to do this for all multiplication gates which might be as many as the size of B itself. Instead, Raz does the following: Let θ be the maximum (in absolute value) among all the scalars $|L'_j(Y)|$ to be multiplied. Replace the scalar multiplication by $L'_j(Y)$ at the inputs to the third part with a multiplication by $L'_j(Y)/\theta$, which is now bounded by 1 in absolute value. Now, (effectively) multiply each of the outputs (there are only n^2 of them) by the scalar θ using $O(\log\theta)$ additions and a multiplication by a scalar of value at most 1. Since $\theta = O(n^\epsilon \log n)$, this increases the size of the resulting circuit by at most $c_1 n^2 \log n$ for some constant $c_1 > 0$. If we can show that the linear circuit B_Y is of size at least $c_2 n^2 \log n$ for some constant $c_2 > c_1$, then we can conclude that the circuit B must have size at least $(c_2 - c_1)n^2 \log n$, thus proving the main lower bound. By choosing the parameters judiciously, we can achieve this goal.

Theorem 3.19. Let C be a bounded coefficient bilinear circuit computing the multiplication of two matrices in $\mathbb{C}^{n\times n}$. Then, $\text{size}(C) = \Omega(n^2 \log n)$.

Proof. We use the notation L and L' as above. For some constants ϵ_1 and ϵ to be determined later, if $\text{Rig}_{\epsilon_1 n^2}(L') \geq n^{\epsilon}$, then by Corollary 3.18, we obtain that the linear part on Y itself is of size at least $\epsilon\epsilon_1 n^2 \log n$ and we are done. Hence, we may henceforth assume that $\text{Rig}_{\epsilon_1 n^2}(L') < n^{\epsilon}$. Let $c_1 := \epsilon_1\epsilon$.

Referring ahead to Lemma 3.20, we know that in this case, there exists a $Y \in \mathbb{C}^{n\times n}$ such that

(1) For all $1 \leq j \leq k$, $|L'_j(Y)| \leq \delta n^{\epsilon} \log n$ for some δ that depends on ϵ, ϵ_1.
(2) For constants ϵ_3 and α, $\text{Rig}_{\epsilon_3 n}(Y) \geq \alpha\sqrt{n}$.

Fix the second input to C to be the matrix Y given by Lemma 3.20. The multiplication gates in the middle layer now become scalar multiplications by $L'_j(Y)$ of the linear forms $L_i(X)$ of the variable matrix X. Let $\theta_j := L'_j(Y)$ for $1 \leq j \leq k$ and $\theta := \max_j |\theta_j|$. Replace each scalar multiplication by θ_j with one by θ_j/θ. Then, multiply each output of the circuit by θ where this is accomplished by at most $\log\theta + 1$ repeated additions and a final multiplication by a scalar of absolute value at most 1. Let C' be the resulting circuit. Clearly, C' is a linear circuit equivalent to C restricted to the fixed second matrix Y. Moreover, C' uses scalar multiplications bounded in absolute value by 1. It is easy to see that C' computes a linear transformation on $X \in \mathbb{C}^{n^2}$ given by the matrix $I \otimes Y \in \mathbb{C}^{n^2\times n^2}$. Finally, since there are n^2 outputs and, by (1), $\theta \leq \delta n^{\epsilon} \log n$, $\text{size}(C')$ is bounded above by $\text{size}(C) + n^2(2\log\theta + 3) \leq \text{size}(C) + c_3 n^2 \log n$ for a constant c_3.

Our goal is now reduced to proving a lower bound on the linear circuit C'. It is easy to see that $\mathrm{Vol}_{r\cdot n}(I \otimes Y) \geq (\mathrm{Vol}_r(Y))^n$. By (2), and Lemma 3.13, $\mathrm{Vol}_r(Y) \geq (\alpha\sqrt{n})^r$ for $r = \epsilon_3 n$. Hence, $\mathrm{Vol}_{\epsilon_3 n^2}(I \otimes Y) \geq (\alpha\sqrt{n})^{\epsilon_3 n^2}$. Finally, by Theorem 3.17,

$$\mathrm{size}(C') \geq \log \mathrm{Vol}_{\epsilon_3 n^2}(I \otimes Y) \geq \epsilon_3 n^2 \log(\alpha\sqrt{n}) \geq c_4 n^2 \log n,$$

for a constant c_4.

By choosing the parameters, we can ensure that $c_4 > c_3$.

It follows that $\mathrm{size}(C) \geq \mathrm{size}(C') - c_3 n^2 \log n \geq c n^2 \log n$. □

Lemma 3.20. Let $L'_1, \ldots, L'_k$ be linear forms on $\mathbb{C}^{n^2}$ as above and $L' \in \mathbb{C}^{n^2 \times k}$ be the matrix with L'_j as its columns. Fix parameters ϵ, ϵ_1, c_1 and c_2. Then, there exists a matrix $Y \in \mathbb{C}^{n \times n}$ (also viewed as a vector in $\mathbb{C}^{n^2}$) such that

(1) For each j, $1 \leq j \leq k$,

$$|L'_j(Y)| \leq c_1 \,\mathrm{Rig}_{\epsilon n^2}(L')\sqrt{2\ln k + 4}.$$

(2)

$$\mathrm{Rig}_{\epsilon_1 n}(Y) \geq c_2\sqrt{n}.$$

Proof. Let $R := \mathrm{Rig}_{\epsilon n^2}(L')$. This means there exists a subspace $V \subset \mathbb{C}^{n^2}$ such that for all j, $\mathrm{dist}(L'_j, V) \leq R$. Let $V^\perp$ be the orthogonal complement of V. Let $L'_{j,V}$ and $L'_{j,V^\perp}$ be the projections of the vector L'_j on V and $V^\perp$, respectively. It follows that $\|L'_{j,V^\perp}\| \leq R$ for all j.

We prove the existence of Y through a probabilistic argument. Choose a matrix $W \in \mathbb{C}^{n \times n}$ by picking each entry independently from the Gaussian distribution with expectation 0 and variance 1, denoted $N(0,1)$. Let $W = W_V + W_{V^\perp}$ be the decomposition of W by projecting along V and $V^\perp$. We define $Y := W_{V^\perp}$, to be the projection of W on $V^\perp$.

Note that $L'_j(Y) = \langle L'_j, Y\rangle$ when treating L'_j as a vector in $\mathbb{C}^{n^2}$. Since $Y \in V^\perp$,

$$\langle L'_j, Y\rangle = \langle L'_{j,V^\perp}, Y\rangle = \langle L'_{j,V^\perp}, W\rangle.$$

It is well-known that if independent random variables X_i are distributed according to $N(\mu_i, \sigma_i^2)$, then $\sum_{i=1}^{t} \lambda_i X_i$ is distributed according to $N(\sum_i \lambda_i \mu_i, \sum_i \lambda_i^2 \sigma_i^2)$ and that if X is distributed according to $N(\mu, \sigma^2)$, then

$$\Pr[|X - \mu| > \theta\sigma] < \sqrt{\frac{2}{\pi}} \frac{\mathrm{e}^{-\theta^2/2}}{\theta}.$$

Since entries of W are independently distributed according to $N(0,1)$, it follows that $\langle L'_{j,V^\perp}, W\rangle$ is distributed according to $N(0, \|L'_{j,V^\perp}\|^2)$. Recall that $\|L'_{j,V^\perp}\| \leq R$. Hence,

$$\Pr[|\langle L'_{j,V^\perp}, W\rangle| > R\sqrt{2\ln k + 4}] < \sqrt{\frac{1}{\pi(\ln k + 2)}} \frac{\mathrm{e}^{-2}}{k}.$$

The above holds for each j, $1 \leq j \leq k$. Hence, we will have $|\langle L'_{j,V^\perp}, W\rangle| \leq R\sqrt{2\ln k + 4}$ for all j with probability $1 - \frac{\mathrm{e}^{-2}}{\sqrt{\pi(\ln k + 2)}} = 1 - o(1)$. Thus, a matrix Y satisfying (1) exists with probability $1 - o(1)$.

Toward proving (2), we first recall the well-known fact [92] that $\sigma_1(W)$ converges to $2\sqrt{n}$ almost surely. Hence, with probability $1 - o(1)$, $\sigma_1(W) \leq (2 + \epsilon)\sqrt{n}$, for a small enough $\epsilon > 0$.

Secondly, we recall that if $X_i \sim N(0,1)$, $1 \leq i \leq m$, are i.i.d. random variables, then $Q = \sum_i^m X_i^2$ is distributed according to the *chi-square density with m degrees of freedom*: $f(x,m) = \frac{2^{-m/2}}{\Gamma(m/2)} x^{m/2-1} \exp(-x/2)$. In particular, Q has expectation m, variance $2m$, and moment generating function $M_Q(t) = E[e^{tQ}](1 - 2t)^{-m/2}$ for $2t < 1$. Moreover, in the limit, Q approaches normal distribution . Hence, $\Pr[Q < (1 - \delta)m] < \epsilon$ for suitable ϵ and δ. It follows that $\|W\|_F^2 \geq (1 - \delta)n^2$ with high probability.

By Lemma 3.5 and Fact 3.3

$$\ell_2\text{-Rig}_r^2(W) \geq \sum_{i=r+1}^{n} \sigma_i^2(W) = \|W\|_F^2 - \sum_{i=1}^{r} \sigma_i^2(W).$$

By the above estimates, with high probability $\|W\|_F^2 \geq (1 - \delta)n^2$ and $\sigma_1(W) \leq (2 + \epsilon)\sqrt{n}$. For $r \leq cn$ for sufficiently small c, we then have, with high probability, $\ell_2\text{-Rig}_r^2(W) \geq (1 - \delta)n^2 - c(2 + \epsilon)^2 n^2 \geq c_1' n^2$.

Now, we use Lemma 3.11 to conclude that $\mathrm{Rig}_r(W) \geq \ell_2\text{-}\mathrm{Rig}_r(W)/\sqrt{n} \geq c_1\sqrt{n}$.

We still have to prove a lower bound on $\mathrm{Rig}_r(Y)$. Recalling $W = Y + W_V$, we note that for any rank-r matrix D, $\|Y - D\|_F \geq \|W - D\|_F - \|W - Y\|_F = \|W - D\|_F - \|W_V\|_F$. We will show that, with high probability, $\|W_V\|_F^2 \leq (1+\epsilon)\dim(V)n \leq c_2'n^2$ when $\dim(V) \leq r = cn$ for sufficiently small r. This will prove that $\|Y - D\|_F \geq c_1 n - c_2 n = c_3 n$ and by Lemma 3.11, $\mathrm{Rig}_r(Y) \geq c_3\sqrt{n}$.

Since (1) and (2) each holds with high probability(say, more than 3/4 each), we have shown that a matrix Y satisfying both (1) and (2) exists.

It remains to observe that $\|W_V\|_F^2$ is at most $(1+\epsilon)rn$ with high probability. Note that W_V is a projection of the Gaussian matrix W onto V. By choosing an orthonormal basis for V, we can ensure that each entry of W_V is a linear combination of Gaussian variables according to $N(0,1)$ by a unit vector, i.e., by λ_i where $\sum_i \lambda_i^2 = 1$. Thus, W_V is a collection of rn variables according to $N(0,1)$. By the same arguments that we used for W, we conclude that $\|W_V\|_F^2$ has the chi-square distribution with rn degrees of freedom with mean rn and variance $2rn$, and hence with high probability is at most $(1+\epsilon)rn$. □

3.4.2 Bounded Coefficient Complexity of Convolution

Bürgisser and Lotz [22] prove superlinear lower bounds on the bounded coefficient complexity of the bilinear forms defined by (cyclic) convolution, polynomial multiplication, and polynomial division with remainder.

Definition 3.7. Let $a = (a_0, \ldots, a_{n-1})$ and $b = (b_0, \ldots, b_{n-1})$ be vectors in $\mathbb{C}^n$. The cyclic convolution c of a and b is defined by

$$c_k = \sum_{i+j=k \bmod n} a_i b_j, \quad 0 \leq k \leq n-1.$$

Clearly, cyclic convolution is a bilinear form $\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}^n$. Convolution of a and b is often denoted by $a * b$.

If $f = \sum_{i=0}^{n-1} a_i x^i$ and $g = \sum_{j=0}^{n-1} b_j x^j$ are polynomials with a and b as coefficient vectors, then c gives the coefficients of the product h (also called convolution) of f and g in the ring $\mathbb{C}[X]/(X^n - 1)$.

The lower bounds on polynomial multiplication and division will be consequences, *via* well-known reductions, from those on convolution. The overall structure of the proof for convolution is similar to the one in Section 3.4.1 for matrix multiplication, so we will be terse.

Note that if we fix a, the convolution $a * b$ is just the linear transformation Ab given by the *circulant* matrix

$$A := \mathrm{circ}(a) := \begin{bmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & \cdots & a_{n-2} \\ \cdots & \cdots & \cdots & \cdots \\ a_1 & a_2 & \cdots & a_0 \end{bmatrix}$$

Consider a bilinear circuit computing $a * b$. Let the first part (cf. Definition 3.6) compute linear forms $L_i(a)$, $1 \le i \le k$, in a. Fix r, $1 \le r \le n$. Let $\mathrm{Rig}_{n-r}(L_1, \ldots, L_k) =: R$.

We will fix a (by a probabilistic argument) to satisfy the following two properties:

(1) $\forall i, 1 \le i \le k,, |L_i(a)| \le R\sqrt{2 \ln k + 4}$.
(2) For this a, the bounded coefficient linear circuit complexity of $x \mapsto Ax$, where A is the circulant matrix defined above is at least $(1/2)(n - r)\log n - cn$ for some constant c.

Given such an a, we can argue as in the proof of Theorem 3.19 and obtain the following:

Theorem 3.21. The bounded coefficient bilinear complexity of convolution is at least $(1/12)n \log n - O(n \log\log n)$.

To prove the existence of a satisfying (1) and (2), let $V \subseteq \mathbb{C}^n$ be an $n/2$-dimensional subspace achieving $\mathrm{Rig}_{n/2}(L_1, \ldots, L_k) = R$ and $V^{\perp}$ be its orthogonal complement. As before we will pick a to be a standard Gaussian random vector in $V^{\perp}$.

Lemma 3.22. Let $a \in \mathbb{C}^n$ be a standard Gaussian vector from $V^\perp$ of dimension r chosen as above. Let C be a linear circuit with bounded coefficients computing $x \mapsto Ax$, where A is the circulant matrix given by a. Then,

$$\Pr\left[\operatorname{size}(C) \geq \frac{1}{2} r \log r - cr\right] > \frac{1}{2},$$

where $c \approx 3.73$.[1]

Proof. We will derive a lower bound on the mean square r-volume, $\operatorname{msv}_r(A)$, of the circulant matrix A and use (ii) of Theorem 3.17.

Let $D = (\zeta^{ij})_{i,j=0}^{n-1}$, where $\zeta = \exp(2\pi i/n)$ is a primitive nth root of unity, be the Discrete Fourier Transform (DFT) matrix. Recall that the eigenvalues of A are given by the Fourier transform Da of the sequence a. In particular, $AD = \operatorname{diag}(\lambda_0, \ldots, \lambda_{n-1})D$, where $\lambda = Da$. Since $\frac{1}{\sqrt{n}}D$ is a unitary matrix, $AA^* = \operatorname{diag}(\lambda_0^2, \ldots, \lambda_{n-1}^2)$ and it follows that $|\lambda_i|$ are the singular values of A. Hence

$$\operatorname{msv}_r^2(A) = \sum_{|I|=r} \prod_{i \in I} |\lambda_i|^2. \tag{3.1}$$

Let $\alpha := (1/\sqrt{n})\lambda = (1/\sqrt{n})Da$. So, $\operatorname{msv}_r^2(A) = n^r \sum_{|I|=r} \prod_{i\in I} |\alpha_i|^2$. Since a is a standard Gaussian vector from $V^\perp$ and $(1/\sqrt{n})D$ is unitary, α is also a standard Gaussian vector in some subspace U of the same dimension r. We now use a deviation inequality for the products of squares of Gaussian random variables. We skip the proof and refer the reader to [22].

Lemma 3.23. Let $Z = (Z_1, \ldots, Z_r)$ be a centered Gaussian vector in $\mathbb{C}^r$. Let $\Sigma_r := (\mathbb{E}[Z_i Z_j])_{1 \leq i,j \leq r}$ be the *covariance matrix* of Z. Let

[1] The value of c is computed from the expectations of $\log X^2$ and $\log^2(X^2 + Y^2)$, where X and Y are independent standard normal random variables. We refer to [22] for details.

δ (≈ 0.02) be a fixed a constant.[2] Then,

$$\mathbb{E}[|Z_1|^2\cdots|Z_r|^2] \geq \det\Sigma_r, \tag{3.2}$$

$$\Pr[|Z_1|^2\cdots|Z_r|^2 \geq \delta^r \det\Sigma_r] > \frac{1}{2}. \tag{3.3}$$

Since α is a standard Gaussian vector in the subspace U, there exists an orthonormal basis $b_1,\dots,b_r \in \mathbb{C}^n$ such that $\alpha = B\beta$, where $B \in \mathbb{C}^{n\times r}$ with b_i as columns and $\beta \in \mathbb{C}^r$ is a vector of independent standard Gaussian random variables. For $I \subseteq [n]$, $|I| = r$, let us denote $\alpha_I = (\alpha_i)_{i\in I}$. We then have $\alpha_I = B_I\beta$, where B_I is the (square) submatrix given by rows with indices in I. Furthermore, the covariance matrix $\Sigma_I = \mathbb{E}[\alpha_I\alpha_I^*] = B_IB_I^*$ by the independence of β_i. Thus, $\det\Sigma_I = |\det B_I|^2$.

By the Binet–Cauchy formula,

$$\det BB^* = \sum_{|I|=r} \det B_I \det B_I^* = \sum_{|I|=r} |\det B_I|^2.$$

But $BB^* = (\langle b_i, b_j^*\rangle)_{1\leq i,j\leq r}$ and b_i are orthonormal. Hence, $\det BB^* = \det(\langle b_i, b_j^*\rangle) = 1$. It follows that $\sum_{|I|=r} |\det B_I|^2 = 1$. Hence, there must exist an I, $|I| = r$ such that $\det\Sigma_I = |\det B|^2 \geq \binom{n}{r}^{-1}$. We apply Lemma 3.23 to α_I for such an I and conclude that $\Pr[\prod_{i\in I} |\alpha_i|^2 \geq \delta^r \binom{n}{r}^{-1}] > 1/2$.

In particular, we now obtain from (3.1), that the circulant matrix A has, with probability at least 1/2,

$$\mathrm{msv}_r^2(A) \geq n^r \delta^r \binom{n}{r}^{-1} \geq \left(\frac{\delta r}{\mathrm{e}}\right)^r,$$

using the inequality $\binom{n}{r} \leq (\mathrm{e}n/r)^r$.

Finally, using Theorem 3.17 and that δ is a constant, we conclude that $\mathrm{size}(C) \geq (1/2)r\log r - cr$ for a suitable constant c with probability at least 1/2. □

[2] The value of δ again arises from considerations alluded to in footnote 1.

4

Sign-Rank and Other Complexity Measures of Sign Matrices

In Sections 2 and 3, we considered rigidity and other robustness measures for arbitrary matrices over a field. But in applications to several lower bound problems, we need to consider robustness measures of Boolean matrices (entries being 0–1 or ± 1). By considering a ± 1-matrix (also called a sign matrix) over $\mathbb{R}$, several complexity measures have been defined in the literature and lower/upper bounds on such measures are used to derive interesting consequences in complexity theory. Such measures include sign-rank, margin complexity, discrepancy, etc., and the applications include communication complexity, learning complexity, Boolean and threshold circuit complexity and so on. In this section, we study complexity measures of sign matrices and several of their applications. In Section 5, we focus on applications of various measures of Boolean matrices to communication complexity.

4.1 Sign-Rank

Let us begin with the definition.

Definition 4.1. For $A \in \{-1,+1\}^{m\times n}$,

$$\text{sign-rank}(A) := \min\{\text{rank}(B) : \forall i,j \ \text{sign}(b_{ij}) = a_{ij}\}.$$

Thus sign-rank(A) measures the robustness of the rank of A under *sign-preserving* changes (every entry is allowed to be changed).

The sign-rank of a matrix has the following elegant geometric interpretation.

Definition 4.2. Two sets of vectors $X = \{x_1,\dots,x_m\}$ and $Y = \{y_1,\dots,y_n\}$ with $x_i, y_j \in \mathbb{R}^d$ are said to *realize* the matrix $A \in \{-1,+1\}^{m\times n}$ in dimension d if $\forall i,j$ $\text{sign}(\langle x_i, y_j\rangle) = a_{ij}$.

The pair (X,Y) is also called a d-dimensional *linear arrangement* for matrix A.

It is easy to see that sign-rank(A) is exactly equal to the *minimum dimension* d of a realization of A. By thinking of A as the incidence matrix of a set system (rows correspond to elements of a universe, columns to subsets, and an element is in a subset iff the corresponding entry is a -1), this is very closely related to a problem about *geometric realizations of set systems* as defined in [71] and [2]. In such a realization, the elements of the universe are represented by points $x_i \in \mathbb{R}^d$ and the sets in the set system by hyperplanes through the origin given by their normal vectors $y_j \in \mathbb{R}^d$. The ith element is in the jth set if and only if the corresponding vector x_i is on the “positive” side of the hyperplane with normal y_j. The goal is to realize a given set system in as small a dimension d as possible.

The definition of sign-rank first appeared (with the geometric interpretation) in the paper by Paturi and Simon [71] who introduced the model of *unbounded error probabilistic communication complexity* and showed that the amount of communication to compute a function $f : \{0,1\}^t \times \{0,1\}^t \to \{0,1\}$ in this model is exactly captured by sign-rank(A_f), where $A_f(x,y) = (-1)^{f(x,y)}$.

Unlike in many cases, it is not even clear that almost all matrices have high sign-rank. Alon et al. [2] prove the non-trivial result that sign-rank$(A) = \Theta(n)$ for almost all $A \in \{-1,+1\}^{n\times n}$. Their proof relies on results from real algebraic geometry, e.g., by Warren [102], on the number of sign patterns of a set of polynomials; a simple *linear algebraic* proof of Warren's theorem can be found in [89]. Since then, finding an explicit matrix with sign-rank more than logarithmic remained a challenge. Then, Forster [28] achieved a breakthrough by proving that any $n \times n$ Hadamard matrix must have sign-rank at least $\Omega(\sqrt{n})$. Forster's lower bound implies asymptotically optimal lower bounds on the communication complexity of the inner product mod-2 function in the Paturi–Simon model and also upper bounds on maximal margins of hyperplane classifiers in learning theory. His lower bound and some of its generalizations are also applied [29] to derive lower bounds on threshold circuits of depth-2 and OBDD's. These applications are discussed in detail in Section 4.5.

4.2 Forster's Lower Bound

Let $x_i, y_j \in \mathbb{R}^d$ be a minimal dimensional realization of the matrix A. We begin with two simple observations: (i) we can assume w.l.o.g. that x_i and y_j are unit vectors in $\mathbb{R}^d$, and (ii) since $\forall i,j$ sign$(\langle x_i, y_j\rangle) = a_{ij}$ is an open condition, we can assume w.l.o.g. that the vectors x_i (and y_j) are in *general position*, i.e., any d of them are linearly independent.

We will make use of the following linear algebraic lemma. We defer its proof to Section 4.2.1.

Lemma 4.1. Let $X = \{x_1, \ldots, x_m\}$ be vectors in $\mathbb{R}^d$ in general position. Then, there is a nonsingular transformation B of $\mathbb{R}^d$ such that $\widetilde{x_i} := Bx_i/\|Bx_i\|$ satisfy

$$\sum_{i=1}^{m} \widetilde{x_i}\widetilde{x_i}^T = \frac{m}{d} I_d.$$

We observe that in any realization of A by x_i, y_j, we can replace them w.l.o.g. by $\widetilde{x_i}$ as given in Lemma 4.1 and $\widetilde{y_j}$ defined by $\widetilde{y_j} := (B^T)^{-1}y_j/\|(B^T)^{-1}y_j\|$. Indeed,

$$\text{sign}\langle \widetilde{x_i}, \widetilde{y_j}\rangle = \text{sign}\langle Bx_i, (B^T)^{-1}y_j\rangle = \text{sign}\langle x_i, y_j\rangle.$$

Hence, we can assume, w.l.o.g., that the x_i are nicely balanced in the sense that they satisfy

$$\sum_{i=1}^{m} x_i {x_i}^T = \frac{m}{d} I_d. \tag{4.1}$$

Theorem 4.2. For $A \in \{-1,+1\}^{m\times n}$, sign-rank$(A) \geq \sqrt{mn}/\|A\|$. In other words, the minimum dimension A can be realized in is at least $\sqrt{mn}/\|A\|$.

Proof. Let $x_i, y_j \in \mathbb{R}^d$ be vectors in a minimal dimensional realization of A. Let H_j be the hyperplane passing through the origin in $\mathbb{R}^d$ with y_j as its normal vector. Let P_i be the point in $\mathbb{R}^d$ given by x_i.

We are interested in bounding the quantity

$$D := \sum_{j=1}^{n} \left(\sum_{i=1}^{m} \mathrm{dist}(P_i, H_j) \right)^2 = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} |\langle x_i, y_j \rangle| \right)^2.$$

As discussed above, w.l.o.g., we can assume that the x_i and the y_j are unit vectors in general position and that the x_i satisfy (4.1).

Fix a j. Now,

$$\begin{aligned}
\sum_{i=1}^{m} |\langle x_i, y_j \rangle| &\geq \sum_{i=1}^{m} (\langle x_i, y_j \rangle)^2 \quad \text{since } x_i, y_j \text{ are unit vectors.} \\
&= \sum_{i=1}^{m} y_j^T x_i x_i^T y_j = y_j^T \left(\sum_{i=1}^{m} x_i x_i^T \right) y_j \\
&= y_j^T \frac{m I_d}{d} y_j \quad \text{by (4.1)} \\
&= \frac{m}{d} y_j^T I_d y_j = \frac{m}{d}.
\end{aligned}$$

It follows that $D \geq nm^2/d^2$. We will next show that $D \leq m\|A\|^2$. Combining these two bounds, we obtain that $d \geq mn/\|A\|$ and the theorem is proved.

Since x_i and y_j realize A, $|\langle x_i, y_j \rangle| = a_{ij} \langle x_i, y_j \rangle$. Hence, for any j,

$$\sum_{i=1}^{m} |\langle x_i, y_j \rangle| = \sum_{i=1}^{m} a_{ij} \langle x_i, y_j \rangle = \left\langle y_j, \sum_{i=1}^{m} a_{ij} x_i \right\rangle \leq \left\| \sum_{i=1}^{m} a_{ij} x_i \right\|,$$

by Cauchy–Schwartz, since y_j is a unit vector.

Thus,

$$\begin{aligned} D &\leq \sum_{j=1}^{n} \left\| \sum_{i=1}^{m} a_{ij} x_i \right\|^2 = \sum_{j=1}^{n} \left(\sum_{k=1}^{m} a_{kj} x_k^T \right) \left(\sum_{l=1}^{m} a_{lj} x_l \right) \\ &= \sum_{1 \leq k,l \leq m} (x_k^T x_l) \sum_{j=1}^{n} a_{kj} a_{lj} = \sum_{1 \leq k,l \leq m} \langle x_k, x_l \rangle \cdot (AA^T)_{kl}. \end{aligned}$$

Observe that $\|A\|^2 I_m - AA^T$ and $(\langle x_k, x_l \rangle)_{1 \leq k,l \leq m}$ are both positive-semidefinite matrices. We now use the following theorem (see, e.g., [39, Corollary 7.5.4]) characterizing positive-semidefinite matrices.

Fact 4.3 (Fejer's Theorem). A matrix $P \in \mathbb{R}^{m \times m}$ is positive-semidefinite if and only if for all positive-semidefinite matrices $Q \in \mathbb{R}^{m \times m}$, $\sum_{1 \leq k,l \leq m} P_{kl} Q_{kl} \geq 0$.

Hence, $\sum_{1 \leq k,l \leq m} \langle x_k, x_l \rangle \cdot (\|A\|^2 I_m - AA^T)_{kl} \geq 0$. Using this we continue with our upper bound on D:

$$\begin{aligned} D &\leq \sum_{1 \leq k,l \leq m} \langle x_k, x_l \rangle \cdot \|A\|^2 (I_m)_{kl} = \|A\|^2 \sum_{1 \leq k \leq m} \langle x_k, x_k \rangle \\ &= \|A\|^2 m \quad \text{since } x_k \text{ are unit vectors.} \end{aligned}$$

This completes the proof. □

Corollary 4.4. For an $n \times n$ Hadamard matrix H, sign-rank(H) $\geq \sqrt{n}$.

4.2.1 Proof of Lemma 4.1

For a vector $x \in \mathbb{R}^d$ (written as a column vector), let $\mathrm{M}(x)$ denote the rank-1 matrix xx^T in $\mathbb{R}^{d \times d}$. For a set of vectors $X = \{x_1, \ldots, x_m\} \subseteq \mathbb{R}^d$, let $\mathrm{M}(X)$ denote the $d \times d$ matrix $\sum_{i=1}^{m} \mathrm{M}(x_i) = \sum_{i=1}^{m} x_i x_i^T$. Define a set of vectors X to be *nice* if $\mathrm{M}(X) = (m/d) I_d$.

For a set of vectors X, let $\mathrm{N}(X)$ denote the normalizations of vectors in X, i.e., $\mathrm{N}(X) = \{x_i / \|x_i\| \ : \ x_i \in X\}$. For a linear transformation B

on $\mathbb{R}^d$, $B(X)$ denotes the set of images of elements of X under B. Then, the lemma can be restated as follows: *If X is a set of vectors in $\mathbb{R}^d$ in general position, then there exists a nonsingular linear transformation B on $\mathbb{R}^d$ such that* $\mathrm{N}(B(X))$ *is nice.*

The proof is based on the following two lemmas.

Lemma 4.5. Let $X \subseteq \mathbb{R}^d$ be a set of $m \geq d$ vectors in general position. Either the matrix $\mathrm{M}(X)$ is equal to $(m/d)I_d$ or there is a nonsingular linear transformation B on $\mathbb{R}^d$ such that the matrix $\mathrm{M}(\mathrm{N}(B(X)))$ has its smallest eigenvalue strictly greater than that of $\mathrm{M}(X)$.

Proof. If $|X| = d$, then by applying a nonsingular linear transformation B, we can ensure that $\mathrm{M}(X) = I_d$ and the lemma is proved. So, let us assume that $|X| > d$.

Let λ be the smallest eigenvalue of $\mathrm{M}(X)$. Note that $\mathrm{M}(X)$ is a positive semidefinite matrix and hence all its eigenvalues are nonnegative. In fact, it is easy to see that λ is at least 1 (for instance, map, say, the first d vectors onto the canonical basis vectors e_i of $\mathbb{R}^d$). Further, the sum of eigenvalues of $\mathrm{M}(X)$ is $\mathrm{tr}(\mathrm{M}(X)) = \sum_{i=1}^m \sum_{j=1}^d x_{i,j}^2 = \sum_{i=1}^m \|x_i\|^2 = m$. It follows that if $\lambda \geq m/d$, then all its eigenvalues are m/d. Hence, by applying a nonsingular linear transformation, we can write $\mathrm{M}(X)$ as $(m/d)I_d$ and we are again done.

So, we can assume that $\lambda < m/d$. Let s be the multiplicity of λ. We will show that by applying a nonsingular linear transformation B, we can ensure that the multiplicity of λ as the smallest eigenvalue of $\mathrm{M}(\mathrm{N}(B(X)))$ is strictly smaller than s. By repeatedly applying this step, we arrive at a matrix having λ as its smallest eigenvalue with multiplicity zero and no smaller eigenvalues; thus proving the lemma.

We can assume without loss of generality that $\mathrm{M}(X)$ is diagonalized: $\mathrm{M}(X) = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d)$, where $1 \leq \lambda = \lambda_1 = \lambda_2 = \cdots = \lambda_s < \lambda_{s+1} \leq \cdots \leq \lambda_d$. Let us define $B := \mathrm{diag}(\lambda_1^{-1/2}, \ldots, \lambda_d^{-1/2})$ and consider the multiplicity of λ as an eigenvalue of

$$C := \mathrm{M}(\mathrm{N}(B(X))) = \sum_{i=1}^m \frac{1}{\|Bx_i\|^2}(Bx_i)(Bx_i)^T.$$

Note that $\|Bx_i\|^2 = \sum_{j=1}^{d} x_{i,j}^2/\lambda_j \le 1/\lambda$ since $\lambda_j \ge \lambda$ and $\|x_i\|^2 = 1$. Also, $\sum_{i=1}^{m}(Bx_i)(Bx_i)^T = I_d$ by our assumption on $\mathrm{M}(X)$ and definition of B. It follows that $C - \lambda I_d$ is positive semidefinite. Hence, all eigenvalues of C are at least λ.

We need to show that the multiplicity of λ as an eigenvalue of C is strictly smaller than s. We will do this by showing that the rank of $C - \lambda I_d$ is strictly larger than $d - s$. To this end, we note that in the inequality $\|Bx_i\|^2 \le 1/\lambda$ observed above, equality holds if and only if x_i is an eigenvector of $\mathrm{M}(X)$ for eigenvalue λ. Since λ has multiplicity s, there are at most s vectors $x_i \in X$ such that this equality holds. For the remaining $m - s$ vectors from X, this inequality is strict. Note that $s < m$ since $\lambda < m/d$ and the sum of the eigenvalues of $\mathrm{M}(X)$ is m. Since

$$C - \lambda I_d = \sum_{i=1}^{m} \left(\frac{1}{\|Bx_i\|^2} - \lambda \right) (Bx_i)(Bx_i)^T,$$

for at least $m - s$ terms in the sum, the coefficients are strictly positive. Moreover, the vectors x_i (and hence Bx_i) corresponding to these terms are either linearly independent (if there are fewer than d of them) or in general position (if there are more than d of them). Hence, $C - \lambda I_d$ is a positive linear combination of at least $\min\{m - s, d\}$ rank-1 matrices of the form $y_i y_i^T$. It follows that the rank of $C - \lambda I_d$ is at least $\min\{m - s, d\}$. Since $m > d$, we conclude that the rank of $C - \lambda I_d$ is *strictly larger* than $d - s$, unless it is already d. □

Lemma 4.6. For a given finite set $X \subseteq \mathbb{R}^d$ of vectors in general position and an $\epsilon > 0$, the set of all matrices $A \in \mathbb{R}^{d \times d}$ such that (1) and (2) below hold is compact.

(1) $\|B\| = 1$.
(2) The smallest eigenvalue of $\mathrm{M}(\mathrm{N}(B(X)))$ is least $1 + \epsilon$.

The proof of this lemma is by a simple compactness argument and we refer the reader to [28] for the details.

Given Lemmas 4.5 and 4.6, we can now complete the proof of Lemma 4.1.

By reasoning as in the proof of Lemma 4.5, we can see that the smallest eigenvalue of $\mathrm{M}(X)$ is at least 1. If $\mathrm{M}(X)$ is already not of the form $(m/d)I_d$, then Lemma 4.5 shows that we can increase its smallest eigenvalue to at least $1 + \epsilon$ for some $\epsilon > 0$ after applying a nonsingular linear transformation B and normalizing.

The nonsingular transformation B above can be assumed to have $\|B\| = 1$ since replacing B by $B/\|B\|$ does not change $\mathrm{M}(\mathrm{N}(B(X)))$. By Lemma 4.6, the set of all such linear transformations is compact. The smallest eigenvalue of a positive semidefinite matrix is a continuous function of the matrix. Hence, the smallest eigenvalue of $\mathrm{M}(\mathrm{N}(B(X)))$ is a continuous function of B for the given X. It follows that the maximal value of the smallest eigenvalue is achieved by some B in this set. For this B, $\mathrm{M}(\mathrm{N}(B(X)))$ must be of the form $(m/d)I_d$. Otherwise, we can apply Lemma 4.5 again and find another B in this set to increase the smallest eigenvalue. This contradicts the maximality at B of the smallest eigenvalue of $\mathrm{M}(\mathrm{N}(B(X))$. □

4.3 Margin Complexity, Factorization Norms, and Discrepancy

The *margin* of an arrangement of hyperplanes and points arises in learning theory. We will discuss this application in more detail in Section 4.5.

Definition 4.3. Let $X = \{x_i,\ldots,x_m\}$ and $Y = \{y_1,\ldots,y_n\}$ be a sign realization by unit vectors of the matrix $A \in \{-1,+1\}^{m\times n}$ as in Definition 4.2. The *margin* of A is then defined to be the maximum margin among all its realizations:

$$\mathrm{margin}(A) := \max\left\{\min_{i,j} |\langle x_i, y_j\rangle| : X,Y \text{ realize } A\right\}.$$

We also define, following [55], the *margin complexity* $\mathrm{mc}(A)$ to be $\mathrm{margin}(A)^{-1}$.

When we view x_i as points and y_j as normals to hyperplanes through the origin, $\mathrm{margin}(A)$ is the minimum distance of any point from any hyperplane in the arrangement. Note that the dimension of the $\mathbb{R}$-space in which the x_i and y_j live is not relevant for this definition.

It is easy to see that for any $A \in \{-1,+1\}^{m\times n}$, $\mathrm{mc}(A)$ is at most $\min\{\sqrt{m},\sqrt{n}\}$.

Forster's result from Section 4.2 also proves a lower bound on the margin complexity of a sign matrix.

Theorem 4.7. For $A \in \{-1,+1\}^{m\times n}$, $\mathrm{mc}(A) \geq \sqrt{mn}/\|A\|$.

Proof. Recall the definition of D from the proof of Theorem 4.2:

$$D := \sum_{j=1}^{n}\left(\sum_{i=1}^{m}\mathrm{dist}(P_i,H_j)\right)^2 = \sum_{j=1}^{n}\left(\sum_{i=1}^{m}|\langle x_i,y_j\rangle|\right)^2.$$

Let us consider D in a realization by x_i and y_j achieving the maximal margin for A.

Clearly, $D \geq n(m\,\mathrm{margin}(A))^2$. In that proof, we also proved the upper bound $D \leq m\|A\|^2$. Combining these, we have $\mathrm{margin}(A) \leq \|A\|/\sqrt{mn}$. Since $\mathrm{mc}(A) = \mathrm{margin}(A)^{-1}$, we are done. □

Linial et al. [53] relate margin complexity to some factorization norms of sign matrices and discrepancy.

Given an operator $A : U \to V$, its *factorization norm through the normed space* W is obtained by writing $A = XY$, where $Y : U \to W$ and $X : W \to V$, that minimizes the product of the norms of operators X and Y. An example of such a factorization norm of interest to us is when $U = \ell_1^m$, $V = \ell_\infty^n$, and $W = \ell_2$.

Definition 4.4. Let $A : \ell_1^m \to \ell_\infty^n$. Then, we define $\gamma_2(A)$ to be

$$\gamma_2(A) = \min_{A=XY}\|Y\|_{1\to 2}\|X\|_{2\to\infty}.$$

It is easy to see that for a matrix $R \in \mathbb{R}^{m\times r}$, $\|R\|_{1\to 2}$ is the largest ℓ_2-norm of a row of R and that for a matrix $C \in \mathbb{R}^{s\times n}$, $\|C\|_{2\to\infty}$ is the largest ℓ_2-norm of a column of C. Hence, we have

$$\gamma_2(A) = \min_{A=XY}\max_{ij}\|x_i\|\,\|y_j\|, \tag{4.2}$$

where x_i, $1 \leq i \leq m$, are the rows of X and y_j, $1 \leq j \leq n$, are the columns of Y.

Definition 4.5. For a matrix $A \in \{+1, -1\}^{m\times n}$, the set $SP(A)$ is defined by

$$SP(A) := \{B \in \mathbb{R}^{m\times n} \text{ such that } \forall i, j\ a_{ij}b_{ij} \geq 1\},$$

i.e., the set of real matrices B that agree in sign with A entry-wise and have each entry of absolute value at least 1.

Lemma 4.8. $\mathrm{mc}(A) = \min_{B\in SP(A)} \gamma_2(B)$.

Proof. (sketch) Given a sign realization as in Definition 4.3 achieving $\mathrm{mc}(A)$, identify (by abuse of notation) the set of vectors X and Y with matrices X and Y with rows x_i and columns y_j, respectively. Let $B := \mathrm{mc}(A)XY$. Note that $B \in SP(A)$ and $\gamma_2(B) \leq \mathrm{mc}(A)$. This shows that $\min_{B\in SP(A)} \gamma_2(B) \leq \mathrm{mc}(A)$. For the other direction, let B achieve the minimum on the right-hand side. Let X and Y be the matrices realizing the minimum in $\gamma_2(B)$. Since $B \in SP(A)$, it can be seen that the rows and columns of X and Y, respectively give, after suitable normalization, a sign realization of A with a margin complexity at most $\gamma_2(B)$. □

Corollary 4.9. $\mathrm{mc}(A) \leq \gamma_2(A)$.

We recall the definition of a dual norm.

Definition 4.6. Let $A \in \mathbb{C}^{m\times n}$. Let $\|\cdot\|$ be a norm on a vector space V over $\mathbb{R}$ or $\mathbb{C}$. The dual norm $\|\cdot\|^*$ is defined by

$$\|y\|^* := \sup\{|\langle x, y\rangle| : \|x\| = 1\}.$$

In particular, for a matrix norm $\|\cdot\|$ on $\mathbb{C}^{m\times n}$ the dual norm $\|\cdot\|^*$ is defined by

$$\|A\|^* := \sup\{|\operatorname{tr} AC^*| : C \in \mathbb{C}^{m\times n}, \|C\| = 1\},$$

where the dual is defined with respect to the Frobenius inner product, $\langle A, C\rangle = \mathrm{tr}(AC^*)$, on $\mathbb{C}^{m\times n}$.

Using the dual norm γ_2^* of γ_2, Linial et al. [53] obtain a better lower bound on $\mathrm{mc}(A)$ than Forster's (Theorem 4.7).

Theorem 4.10. $\mathrm{mc}(A) \geq mn/\gamma_2^*(A)$.

Proof. Let $B \in \mathbb{R}^{m\times n}$ be a matrix with $\mathrm{mc}(A) = \gamma_2(B)$ as in Lemma 4.8. Define $C := B/\gamma_2(B)$ so $\gamma_2(C) = 1$. Since B and C have the same sign pattern, $\langle C, A\rangle = \sum_{ij} c_{ij}a_{ij} = (\sum_{ij} b_{ij}a_{ij})/\gamma_2(B) \geq mn/\gamma_2(B)$ since $B \in SP(A)$. Hence, by definition, $\gamma_2^*(A) \geq mn/\gamma_2(B)$. Since $\mathrm{mc}(A) = \gamma_2(B)$, we are done. □

The following inequality shows that the bound in Theorem 4.10 can be better than the bound in Theorem 4.7.

Lemma 4.11. For $A \in \mathbb{R}^{m\times n}$, $\gamma_2^*(A) \leq \sqrt{mn}\|A\|$.

Proof. Let $B \in \mathbb{R}^{m\times n}$ be such that $\gamma_2(B) = 1$. Using (4.2), let $X \in \mathbb{R}^{m\times t}$ and $Y \in \mathbb{R}^{t\times n}$ (the value of t is not important) be such that $B = XY$ and $\gamma_2(B) = 1 = \max_{ij}\|x_i\|\|y_j\|$, where x_i (y_j) is the ith row (jth column) of X (Y), $1 \leq i \leq m$, $1 \leq j \leq n$. Let $x^{(j)}$ be the jth *column* of X and $y^{(i)}$ be the ith *row* of Y for $1 \leq i, j \leq t$ and note that $B = XY = \sum_{j=1}^{t} x^{(j)}y^{(j)}$. Now,

$$\begin{aligned}
\langle A, B\rangle &= \left\langle A, \sum x^{(j)}y^{(j)}\right\rangle = \sum_j \langle A, x^{(j)}y^{(j)}\rangle \\
&= \sum_j y^{(j)}Ax^{(j)} \\
&\leq \|A\| \sum_j \|x^{(j)}\|\|y^{(j)}\| \quad \text{by definition of } \|A\| \\
&\leq \|A\| \sqrt{\left(\sum_j \|x^{(j)}\|^2\right)\left(\sum_j \|y^{(j)}\|^2\right)} \quad \text{by Cauchy–Schwartz}
\end{aligned}$$

$$= \|A\|\sqrt{\left(\sum_i \|x_i\|^2\right)\left(\sum_j \|y_j\|^2\right)}$$
$$\leq \|A\|\sqrt{m}\max_i \|x_i\|\sqrt{n}\max_j \|y_j\|$$
$$= \|A\|\sqrt{mn}\gamma_2(B) = \|A\|\sqrt{mn}. \qquad \square$$

In fact, Linial et al. exhibit a (random-like) matrix $A \in \{-1,+1\}^{n\times n}$ for which $\gamma_2^*(A) = O(n^{3/2})$, whereas $\|A\| = \Omega(n^{3/4})$. For this A, then, Forster's bound can give at most a lower bound on $\mathrm{mc}(A)$ of $O(n^{1/4})$ whereas the bound from Theorem 4.10 gives $\Omega(n^{1/2})$.

Finally, Linial et al. use a famous inequality due to Grothendieck to relate $\gamma_2^*(A)$ to an operator norm of A.

Theorem 4.12 (Grothendieck's Inequality). Let $A = (a_{ij})$ be a real matrix such that $|\sum_{ij} a_{ij}s_i t_j| \leq 1$ for every choice of reals that satisfy $|s_i|, |t_j| \leq 1$ for all i, j. Then, there exists an absolute constant K_G, where $1.5 < K_G < 1.8$, such that

$$\left|\sum_{ij} a_{ij}\langle x_i, y_j\rangle\right| \leq K_G,$$

for all unit vectors x_i and y_j in a real Hilbert space.

For an elementary proof of this inequality, see [14].

Lemma 4.13. For every $A \in \mathbb{R}^{m\times n}$,

$$\|A\|_{\infty\to 1} \leq \gamma_2^*(A) \leq K_G\|A\|_{\infty\to 1}.$$

Proof. Let x be such that $\|x\|_\infty = 1$ and $\|Ax\|_1 = \|A\|_1$. Consider the matrix $B = \mathrm{sign}(x)x^t$, where $\mathrm{sign}(x)$ is the column vector of signs of the coordinates of x. It is easy to see that $\gamma_2(B) = \|x\|_\infty = 1$ and that $\langle A, B\rangle = \mathrm{sign}(x)^t Ax = \|Ax\|_1$. Hence, $\gamma_2^*(A) \geq \langle A, B\rangle = \|A\|_{\infty\to 1}$.

The other direction is a restatement of Grothendieck's inequality. □

Corollary 4.14.

$$\mathrm{mc}(A) \geq \frac{mn}{K_G \|A\|_{\infty\to 1}}.$$

Using methods similar to the foregoing, Linial et al. also prove lower bounds on sign-rank.

Theorem 4.15. For any $A \in \{-1,+1\}^{m\times n}$,

$$\text{sign-rank}(A) \geq \frac{mn}{\gamma_2^*(A)}, \quad \text{and}$$

$$\text{sign-rank}(A) \geq \frac{mn}{K_G \|A\|_{\infty\to 1}}.$$

Using the well-known John's theorem [17, Section 4, Theorem 15] (also known as John's Lemma and related to the so-called Löwner–John ellipsoids), Linial et al. also show the following relation between rank and γ_2.

Lemma 4.16. For any $A \in \mathbb{R}^{m\times n}$,

$$\gamma_2^2(A) \leq \|A\|_{\infty\to 1}^2 \,\mathrm{rank}(A).$$

In particular, for a sign matrix A,

$$\gamma_2(A) \leq \sqrt{\mathrm{rank}(A)}.$$

Combined with Corollary 4.9 we have the following for any $m \times n$ sign matrix A:

$$\mathrm{mc}(A) \leq \gamma_2(A) \leq \sqrt{\mathrm{rank}(A)}. \tag{4.3}$$

The results for an Hadamard matrix show that these inequalities are in general tight.

As we note from the above, we have the same lower bounds on sign-rank(A) and mc(A) and they are proved using very similar techniques. Hence, one wonders if there is some relation between sign-rank(A) and mc(A). The best known relation is the following based on the Johnson–Lindenstrauss lemma.

Lemma 4.17. For $A \in \{-1,+1\}^{m\times n}$,

$$\text{sign-rank}(A) = O(\text{mc}(A)^2 \log(m+n)).$$

We will see a proof of this lemma in the context of probabilistic communication complexity in Section 5.3 (cf. Lemma 5.15).

Finally, Linial and Shraibman [55] show a close relationship between margin complexity and discrepancy.

Definition 4.7. Let $A \in \{-1,+1\}^{m\times n}$ and μ be a probability distribution on entries of A. For a combinatorial rectangle R in A, let R^+ and R^- are the positive and negative entries in R, respectively. Then,

$$\text{disc}_\mu(A) := \max |\mu(R^+) - \mu(R^-)|,$$

where the maximum is taken over all combinatorial rectangles R in A. The discrepancy of A, denoted disc(A), is then defined as the minimum of disc$_\mu$ over all distributions μ.

Theorem 4.18. For any $A \in \{-1,+1\}^{m\times n}$,

$$\frac{1}{8}\,\text{margin}(A) \le \text{disc}(A) \le 8\,\text{margin}(A).$$

Proof. (outline) The proof consists of the three main steps below. For details, see [55].

(1) Show that the margin changes by at most a factor of K_G, where K_G is the Grothendieck's constant, if we replace the arbitrary unit vectors in Definition 4.3 by normalized sign vectors, i.e., $x_i/\|x_i\|, y_j/\|y_j\|$, where $x_i, y_j \in \{-1,+1\}^k$ for some k. Denote this restricted variant of margin by margin$_\pm(A)$.

(2) Next, observe that

$$\operatorname{disc}(A) \leq \min_{P \in \mathcal{P}} \|P \circ A\|_{\infty \to 1} \leq 4 \operatorname{disc}(A),$$

where $\circ$ denotes the Hadamard product of two matrices and $\mathcal{P}$ denotes the set of matrices with nonnegative entries that sum up to 1.

(3) Finally, express margin$_\pm$ in the first step as the optimum of a linear program and use Linear Programming (LP) duality to show that

$$\operatorname{margin}_\pm(A) = \min_{P \in \mathcal{P}} \|P \circ A\|_{\infty \to 1}. \qquad \square$$

4.4 Sign-Rank of an AC^0 function

Recently, Razborov and Sherstov [88] proved a lower bound of $\exp(\Omega((\log n)^{1/3}))$ on the sign-rank of an $n \times n$ matrix A_f obtained by evaluating an AC^0 function f on all (x,y). This is a remarkable result both due to the new techniques it uses and several interesting consequences. In this subsection, we review the proof of this lower bound. We discuss its applications in the next subsection.

The matrix A_f is a $2^{t^3} \times 2^{t^3}$ ± 1 matrix given by evaluating the AC^0 function

$$f(x,y) := \bigwedge_{i=1}^{t} \bigvee_{j=1}^{t^2} (x_{ij} \wedge y_{ij}), \tag{4.4}$$

on all pairs $(x,y) \in \{0,1\}^{t^3} \times \{0,1\}^{t^3}$, i.e., $A_f(x,y) := (-1)^{f(x,y)}$. The main result from [88] is that sign-rank$(A_f) = 2^{\Omega(t)}$.

The first ingredient in the proof is a generalization of Forster's argument to matrices that may have a small number of zero entries and there is a lower bound on the magnitude of the remaining entries. The second new ingredient is the construction of the so-called *smooth orthogonalizing distributions* for Boolean functions. By "masking" a Boolean function with such a distribution, the Fourier coefficients can be controlled. The final ingredient is the derivation of an expression for the spectral norm of a *pattern matrix* in terms of the Fourier transform of the Boolean function defining that matrix. We will discuss these ingredients before proving the main theorem.

4.4.1 A Generalization of Forster's Argument

We define the sign-rank of an arbitrary real matrix A as the minimum rank of a real matrix B such that all nonzero entries of the entry-wise product $A \circ B$ are strictly positive:

$$\text{sign-rank}(A) := \min\{\text{rank}(B) : \forall i,j \ a_{ij} \neq 0 \Rightarrow a_{ij}b_{ij} > 0\}.$$

Clearly, we can assume that all the entries of A are bounded in absolute value by at most 1.

A simple generalization of Forster's bound from Theorem 4.2 was proved in [29]: for any matrix $A \in \mathbb{R}^{m\times n}$, sign-rank$(A) \geq \min_{xy}|A_{xy}| \cdot \sqrt{mn}/\|A\|$. Another generalization in [30] considers preserving signs for all but h of the entries of A: if $\tilde{A}$ is obtained from $A \in \{-1,+1\}^{m\times n}$ by changing at most h entries in an arbitrary fashion, then sign-rank$(\tilde{A}) \geq \sqrt{mn}/\|A\| + 2\sqrt{h}$. Razborov and Sherstov [88] prove the following hybrid generalization.

Theorem 4.19. Let $A \in \mathbb{R}^{m\times n}$ be such that for all but h of (i,j), $|a_{ij}| \geq \gamma$. Then,

$$\text{sign-rank}(A) \geq \frac{\gamma mn}{\|A\|\sqrt{mn} + \gamma h}.$$

Proof. Let sign-rank$(A) = d$. Let B be a matrix of rank d such that for all i,j, where $a_{ij} \neq 0$, $a_{ij}b_{ij} > 0$. We can further assume that

(1) $\|B\|_\infty \leq 1$.
(2) $\|B\|_F^2 = mn/d$.

Indeed, write $B = XY$, where $X \in \mathbb{R}^{m\times d}$ and $Y \in \mathbb{R}^{d\times n}$. Using arguments similar to those in Section 4.2, we can assume that the rows of X, $x_1,\ldots,x_m \in \mathbb{R}^d$, are unit vectors in general position and that the columns of Y, $y_1,\ldots,y_n \in \mathbb{R}^d$ are all unit vectors. Then, B is defined by $b_{ij} := \langle x_i, y_j\rangle$. Since x_i and y_j are unit vectors, for any i,j, $|b_{ij}| = |\langle x_i,y_j\rangle| \leq \|x_i\|\|y_j\| = 1$ by Cauchy–Schwartz. Thus, (1) holds. The calculation in the proof of Theorem 4.2 shows that for any j, $\sum_{i=1}^m(\langle x_i,y_j\rangle)^2 = m/d$. Thus, $\|B\|_F^2 = \sum_{j=1}^n\sum_{i=1}^m(\langle x_i,y_j\rangle)^2 = mn/d$. This shows (2).

The proof compares an upper bound and a lower bound on $\langle A,B\rangle := \sum_{ij} a_{ij}b_{ij}$ (cf. Lemma 3.7). We have,

$$\begin{aligned}\langle A,B\rangle &\geq \sum_{|a_{ij}|\geq\gamma} a_{ij}b_{ij} \quad \text{since } a_{ij}b_{ij} \geq 0 \text{ for all } i,j\\ &\geq \gamma\left(\sum_{ij}|b_{ij}| - h\right) \quad \text{since } \forall i,j\ |b_{ij}| \leq 1 \quad \text{and}\\ &\qquad\qquad\qquad\qquad a_{ij}b_{ij} = |b_{ij}| \text{ for all but } h \text{ of } (i,j)\\ &\geq \gamma(\|B\|_F^2 - h) \quad \text{by (1) above}\\ &\geq \gamma\Big(\frac{mn}{d} - h\Big) \quad \text{by (2) above.}\end{aligned}$$

On the other hand,

$$\begin{aligned}\langle A,B\rangle &= \sum_i \sigma_i(A)\sigma_i(B) \quad \text{from the proof of Lemma 3.7}\\ &\leq \|A\|\sum_i \sigma_i(B)\\ &\leq \|A\|\|B\|_F\sqrt{d} \quad \text{by Cauchy–Schwartz since}\\ &\qquad \sum_i \sigma_i^2(B) = \|B\|_F^2 \quad \text{and} \quad \sigma_i(B) = 0 \ \text{ for } i > \mathrm{rank}(B)\\ &= \|A\|\sqrt{mn} \quad \text{by (2) above.}\end{aligned}$$

Comparing the two bounds,

$$\gamma\Big(\frac{mn}{d} - h\Big) \leq \langle A,B\rangle \leq \|A\|\sqrt{mn}.$$

Hence,

$$d \geq \frac{\gamma mn}{\|A\|\sqrt{mn} + \gamma h}. \qquad \square$$

4.4.2 Smooth Orthogonalizing Distributions

We recall some basic facts about Fourier analysis of Boolean functions. Consider the space of functions $\mathcal{F} := \{f : \{0,1\}^n \to \mathbb{C}\}$ with the inner product:

$$(f,g) := 2^{-n} \sum_{x \in \{0,1\}^n} f(x)\overline{g(x)}.$$

The functions $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$, $S \subseteq [n]$, form an orthonormal basis for the space $\mathcal{F}$ with the above inner product. Hence, every function $f \in \mathcal{F}$ can be uniquely expressed as a linear combination of χ_S. The coefficients in this linear combination are called the *Fourier Coefficients* of f. In particular, for $S \subseteq [n]$, the Fourier coefficient $\widehat{f}(S)$ is given by

$$\widehat{f}(S) := 2^{-n} \sum_{x \in \{0,1\}^n} f(x)\chi_S(x).$$

Thus, we also have the inverse transform:

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x).$$

By orthogonality of the basis, we also have $\sum_{S \subseteq [n]} |\widehat{f}(S)|^2 = (f,f) = 2^{-n} \sum_{x \in \{0,1\}^n} |f(x)|^2$; this is known as *Parseval's Identity*.

Definition 4.8. A distribution μ on $\{0,1\}^n$ is *k-orthogonalizing* for a function $f : \{0,1\}^n \to \{-1,+1\}$ if

$$|S| < k \Rightarrow \mathbb{E}_{x \sim \mu}[f(x)\chi_S(x)] = 0.$$

A *smooth* distribution places substantial weight on all but a tiny fraction of the sample points.

The second ingredient of the main result constructs smooth orthogonalizing distributions for the function:

$$\mathrm{MP}_m(x) := \bigwedge_{i=1}^{m} \bigvee_{j=1}^{4m^2} x_{ij}.$$

Theorem 4.20. There is an $(m/3)$-orthogonalizing distribution μ for MP_m such that $\mu(x) \geq 8^{-m} 2^{-4m^3}/2$ for all $x \in \{0,1\}^{4m^3}$ such that $\mathrm{MP}_m(x) = -1$.

4.4.3 The Pattern Matrix Method for Lower Bounds on Sign-Rank

The pattern matrix method, introduced by Sherstov [90], is a powerful tool for proving lower bounds in communication complexity. This method is used in obtaining a formula (exact!) for the spectral norm of a pattern matrix defined by an AC^0 function f in terms of Fourier coefficients of f.

Let n and N be positive integers such that $n\,|\,N$. Partition $[N]$ into n intervals of length N/n each:

$$[N] = \left\{1,\ldots,\frac{N}{n}\right\} \dot{\cup} \cdots \dot{\cup} \left\{\frac{(n-1)N}{n} + 1,\ldots,N\right\}.$$

Let $\mathcal{V} := \mathcal{V}(N,n)$ denote the family of n-subsets $V \subseteq [N]$ that contain exactly one element from each of the above intervals. For $x \in \{0,1\}^N$ and $V = \{i_1,\ldots,i_n\} \in \mathcal{V}$, where i_j is in the jth interval in the above partition of $[N]$, let $x_{\restriction V} := (x_{i_1},\ldots,x_{i_n}) \in \{0,1\}^n$ denote the *projection of x onto V*.

Definition 4.9. Let $\phi : \{0,1\}^n \to \mathbb{R}$. The (N,n,ϕ)-*pattern matrix* A has rows indexed by $x \in \{0,1\}^N$ and columns indexed by pairs (V,w) where $V \in \mathcal{V}(N,n)$ and $w \in \{0,1\}^n$. Thus, A is a $2^N \times (N/n)^n 2^n$ real matrix. The entry indexed by $(x,(V,w))$ is given by evaluating the function ϕ at $x_{\restriction V} \oplus w$, i.e., the mod-2 sum of the projection of x onto V and w. That is,

$$A = \left(\phi(x_{\restriction V} \oplus w)\right)_{x\in\{0,1\}^N,(V,w)\in\mathcal{V}\times\{0,1\}^n}.$$

The final ingredient we need is a formula for the spectral norm of the pattern matrix A in terms of the Fourier transform of ϕ, [90].

Theorem 4.21. Let A be the (N,n,ϕ)-matrix as defined above. Then,

$$\|A\| = \sqrt{2^N\left(\frac{2N}{n}\right)^n} \max_{S\subseteq[n]}\left\{|\widehat{\phi}(S)|\left(\frac{n}{N}\right)^{|S|/2}\right\}.$$

We now have all the ingredients to prove the main result of this section.

Theorem 4.22. Let f be defined as in (4.4) and let $A_f = (f(x,y))_{x,y} \in \{-1,+1\}^{2^{t^3} \times 2^{t^3}}$, be the matrix obtained by evaluating f on all pairs of inputs (x,y). Then, sign-rank$(A_f) = 2^{\Omega(t)}$.

Proof. Let M be the (N,n,MP_m) pattern matrix. We first observe that M is a submatrix of A_f, by letting $t = cm$ for some large enough constant c. It suffices to take $c = 4N/n$ (later we will set $N = 10^6 n$). To see that M is a submatrix of A_f, note that $(z_{\restriction V})_{ij} \oplus w_{ij} = ((z_{\restriction V})_{ij} \wedge \overline{w_{ij}}) \vee (\overline{(z_{\restriction V})_{ij}} \wedge w_{ij})$. Now, consider two disjoint subsets of N bits each among the $2t$ bits of x and similarly for y. Map $z \in \{0,1\}^N$ onto an x that agrees with z on one of these two disjoint subsets and with $\overline{z}$ on the other. Next, map $(V,w) \in \mathcal{V} \times \{0,1\}^n$ onto a y, i.e., zero everywhere except the positions indexed by elements of $V \subseteq [N]$ in the two subsets. In the positions indexed by V set the bits in one of the subsets to those of w and in the other subset by $\overline{w}$. For such an (x,y) it is easy to see that $\mathrm{MP}_m(z_{\restriction V} \oplus w) = f(x,y)$. Thus, M is a submatrix of A_f. Hence, it suffices to prove that sign-rank$(M) = 2^{\Omega(m)}$.

Let P be the (N,n,μ) pattern matrix, where μ is as in Theorem 4.20. Since P is a nonnegative matrix, sign-rank$(M) \geq$ sign-rank$(M \circ P)$. Instead of on M, we prove a lower bound of $2^{\Omega(m)}$ on sign-rank$(M \circ P)$ since $M_\mu := M \circ P$ has a better-behaved Fourier transform than M, thanks to the smooth orthogonalizing aspect of μ for the function MP_m.

We will apply Theorem 4.19 to the matrix M_μ. Note that all but an $\exp(-m^2)$-fraction of entries of M_μ have absolute value at least $8^{-m}2^{-n}/2$ by the smoothness property of μ and since $|M(x,y)| = 1$. Hence, we take $\gamma := 8^{-m}2^{-n}/2$ and $h = |M|\exp(-m^2)$, where $|M_\mu| = |M|$ denotes the total number of entries in M, i.e., $2^N \cdot (2N/n)^n$. Hence, we have

$$\text{sign-rank}(M_\mu) \geq \frac{\gamma|M|}{\|M_\mu\|\sqrt{|M|} + \gamma h} \geq \min\left\{\frac{\gamma\sqrt{|M|}}{2\|M_\mu\|}, 2^{\Omega(m^2)}\right\}. \quad (4.5)$$

It remains to prove an upper bound on $\|M_\mu\|$. Let $\phi(x) := \mathrm{MP}_m(x)\mu(x)$ and note that M_μ is the (N,n,ϕ) pattern matrix. From definition, for any $S \subseteq [n]$, $|\widehat{\phi}(S)| \leq 2^{-n}\sum_{x\in\{0,1\}^n} |\phi(x)| = 2^{-n}$. Also, by the orthogonalizing nature of μ, $\widehat{\phi}(S) = 0$ for $|S| \le m/3$. Hence, in the expression for $\|M_\mu\|$ in Theorem 4.21, the max is at most $2^{-n}(n/N)^{m/6}$. It follows that $\|M_\mu\| \le \sqrt{|M|}2^{-n}(N/n)^{-m/6}$.

Substituting this upper bound on $\|M_\mu\|$ and the value of γ in (4.5), we obtain sign-rank$(M_\mu) \ge 8^{-m}(N/n)^{m/6}$. We take $N = 10^6 n$, so we get sign-rank$(M_\mu) = \exp(m)$ and the theorem is proved. □

4.5 Applications of Complexity Measures of Sign Matrices

4.5.1 Communication Complexity

The notion of sign-rank of a matrix was first introduced by Paturi and Simon [71] in the context of *unbounded error probabilistic communication complexity.* We will focus on communication complexity models in detail in Section 5. Hence, we postpone this application to Sections 5.3 and 5.4.

4.5.2 Learning Theory

In learning theory, a basic task is to design algorithms that take a training sample of input–output pairs $((\xi_1, f(\xi_1)), \ldots, (\xi_m, f(\xi_m)))$ of a function $f : \mathcal{D} \to \{-1, +1\}$, $\xi_i \in \mathcal{D}$, and *learn* some (approximate) representation f^* of f. The function f is often assumed to come from a *concept class* $\mathcal{F}$. Properties of $\mathcal{F}$ are exploited to make the learning algorithm efficient and the classifier f^* accurate. The domain $\mathcal{D}$ is often mapped, *via* a *kernel*, into $\mathbb{R}^d$ and the classifier f^* is represented by a hyperplane in $\mathbb{R}^d$. The sample inputs ξ_i are then mapped to points $x_i \in \mathbb{R}^d$ and an *exact*[1] classifier f^* would then correspond to a hyperplane with a normal $y \in \mathbb{R}^d$ such that x_i is on the positive side of the hyperplane, i.e., sign$\langle x_i, y\rangle > 0$ if and only of $f(\xi_i) = +1$. A future input $x \in \mathbb{R}^d$ is then classified by the learned hyperplane according to the sign of $\langle x, y\rangle$.

In large margin classifiers (e.g., Support Vector Machines (SVMs)), the objective of the classifier is to find a hyperplane that maximizes the margin, i.e., the distance to the hyperplane from any point from the

[1] Here we consider only exact classifiers.

training sample. Since scaling can artificially inflate this distance, it is natural to assume that the samples x_i and the normal y are all unit vectors in $\mathbb{R}^d$. Furthermore, we can assume that the hyperplane passes through the origin by translating. Then, for a given sample $X = \{x_i\}$ and a hyperplane with normal y, it is easy to see that the margin is given by $\min_i |\langle x_i, y\rangle|$. Since the unknown function f comes from the concept class $\mathcal{F} = \{f_1, \ldots, f_n\}$, we want to consider the overall minimum margin (worst case among all functions from $\mathcal{F}$) in an *arrangement* of hyperplanes (with normals y_j) for all the functions from $\mathcal{F}$ and for the sample x_i. By increasing the dimension by 1 and losing a factor of 2 in the margin, we can assume that hyperplanes all pass through the origin. The minimum margin of an arrangement (X, Y) is then given by

$$\operatorname{margin}(X,Y) = \min_{ij} |\langle x_i, y_j\rangle|\,.$$

Note that the dimension d of the space in which the arrangement is realized is not relevant for this complexity measure.

A concept class $\mathcal{F}$ on a given sample Ξ of inputs is naturally given by a sign matrix $A_{\Xi,\mathcal{F}} \in \{-1,+1\}^{m\times n}$, where $A_{\Xi,\mathcal{F}}(\xi, f) = f(\xi)$, $|\Xi| = m$, and $|\mathcal{F}| = n$. The performance of a learning algorithm with a given mapping $\kappa : \mathcal{D} \to \mathbb{R}^d$ for this concept class on this sample is $\operatorname{margin}(X,Y) = \min_{ij} |\langle x_i, y_j\rangle|$, where $X = \kappa(\Xi)$ and Y is the set of normals to the homogeneous hyperplanes learned by the algorithm. Hence, the performance of the *best* learning algorithm under the best mapping is given by

$$\operatorname{margin}(A_{\Xi,\mathcal{F}}) = \sup_{X,Y} \operatorname{margin}(X,Y) = \sup_{X,Y} \min_{ij} |\langle x_i, y_j\rangle|\,,$$

where the sup is over all unit vectors $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ in $\mathbb{R}^d$. This is the *margin* of the sign matrix A and defines, in some sense, the inherent learning complexity of A.

It is easy to show that there is a trivial arrangement of homogeneous halfspaces that realizes any $m \times n$ matrix A with a margin of at least $\max\{m^{-1/2}, n^{-1/2}\}$: if $m \leq n$, let x_i be the canonical unit vectors and y_j be the normalized column vectors of the matrix A. Ben-David et al. [11] show that a random $n \times n$ sign matrix has a margin of at most $O(n^{-1/2}\log^{1/2} n)$. Thus, for *almost all* $n \times n$ sign matrices the margin given by the trivial embedding is essentially optimal. Forster's theorem (Theorem 4.7) gives an explicit matrix for which the trivial embedding indeed gives the best possible margin.

Corollary 4.23 (to Theorem 4.7). Suppose, $A_{\Xi,\mathcal{F}}$ is given by an $n \times n$ Hadamard matrix H_n. Then, $\mathrm{margin}(H) \leq n^{-1/2}$.

There's another parameter, the *VC-dimension* of a concept class that also plays a prominent role in learning theory. Vapnik [99], in particular, showed that if the VC-dimension of a concept class is d, then the margin of any arrangement realizing the concept class is at most $d^{-1/2}$. The VC-dimension of the concept class given by an $n \times n$ Hadamard matrix is easily seen to be at most $\log n$. Thus, the upper bound based on VC-dimension is at most $(\log n)^{-1/2}$, whereas the above corollary gives the optimal $n^{-1/2}$.

4.5.3 Threshold Circuits

Our third application is to lower bounds on depth-2 threshold circuits in terms of minimal dimension.

Recall that a (linear) threshold gate with weights $w_1,\ldots,w_n$ and threshold w_0 outputs 1 on inputs $x_1,\ldots,x_n$ if and only if $w_1x_1 + \cdots + w_nx_n \geq w_0$. Proving superpolynomial lower bounds on the size of depth-2 threshold circuits for explicit functions when the threshold gates are allowed to use arbitrary weights is an interesting open question. Hajnal et al. [35] show an exponential lower bound for the `inner product mod-2` function when all the weights used in a depth-2 threshold circuit are polynomially bounded. Here, we will see a stronger result by showing that the restriction on the weights of the top gate can be removed. We use lower bounds on sign-ranks of explicit matrices to derive exponential lower bounds for some explicit functions including `inner product mod-2`. In fact, these lower bounds are exponential when the depth-2 circuit has a threshold gate (with unrestricted weights) at the top and either threshold gates with polynomial weights or gates computing arbitrary symmetric functions at the bottom.

The following theorem states that (loosely speaking) Boolean functions with “high” minimal dimension cannot be computed by “small” (somewhat technically restricted) depth-2 threshold circuits. Part (a) of this theorem strengthens the lower bound of [35] and Part (b) generalizes and strengthens the results of [19, 46, 47]. Note that for technical reasons we assume that the top threshold gate is ±1-valued whereas the threshold gates on the bottom level are $\{0,1\}$-valued.

Theorem 4.24. Let (f_n) be a family of Boolean functions, where $f_n : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ and let $A_f(x,y) = f(x,y)$ be the corresponding $2^n \times 2^n$ sign matrix. Suppose (f_n) is computed by depth-2 threshold circuits in which the top gate is a linear threshold gate (with *unrestricted* weights). Then the following holds:

(a) If the bottom level has s linear threshold gates using integer weights of absolute value at most W, then $s = \Omega\big(\frac{\text{sign-rank}(A_f)}{nW}\big)$. In particular, $s = \Omega\big(\frac{2^{n/2}}{nW}\big)$ for $f_n = \text{ip}_n$ (inner product mod-2).

(b) If the bottom level has s gates computing symmetric functions, then $s = \Omega\big(\frac{\text{sign-rank}(A_f)}{n}\big)$. In particular, $s = \Omega\big(\frac{2^{n/2}}{n}\big)$ for $f_n = \text{ip}_n$.

Note that the specific bounds for ip_n follow immediately from the general bound for f_n by applying Corollary 4.4.

Using Theorem 4.22, we obtain a separation result between depth-3 AC^0 and depth-2 threshold circuits.

Corollary 4.25. Let f_n be the depth-3 AC^0-function given by (4.4) with $n = t^3$. Then, depth-2 threshold circuits of the kind in (a) computing f_n must have size $2^{\Omega(n^{1/3})}/W$ and depth-2 threshold circuits of the kind in (b) computing f_n must have size $2^{\Omega(n^{1/3})}$.

The proof of this theorem is based on Theorem 4.2 and the two lemmas below.

Lemma 4.26. Let $G : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a threshold function where for $x, y \in \{0,1\}^n$, $G(x,y) = 1$ if and only if $\sum_{i=1}^n \alpha_i x_i + \sum_{i=1}^n \beta_i y_i \geq \mu$ for weights $\alpha_i, \beta_i, \mu \in \mathbb{Z}$. Then, G (viewed as a matrix) has rank at most $\min\{\sum_{i=1}^n |\alpha_i|, \sum_{i=1}^n |\beta_i|\} + 1$.

Proof. W.l.o.g. let $\sum_{i=1}^n |\alpha_i| \leq \sum_{i=1}^n |\beta_i|$. Let $\alpha_{\min}$ ($\alpha_{\max}$) be the minimal (maximal) value taken by $\sum_{i=1}^n \alpha_i x_i$ as x ranges over all possible inputs in $\{0,1\}^n$. As the weights are integers, this sum takes at most $\alpha_{\max} - \alpha_{\min} + 1$ distinct values. We partition the rows of G according to the weight contributed by x and, within each block of these rows, we partition the columns into two groups depending on whether or not the weight contributed by y together with that of any row of that block exceeds the threshold μ or not. Specifically, define the following sets of entries of G for all α such that $\alpha_{\min} \leq \alpha \leq \alpha_{\max}$:

$$S_{\alpha,0} := \left\{ (x,y) : \sum_{i=1}^n \alpha_i x_i = \alpha \text{ and } \sum_{i=1}^n \beta_i y_i < \mu - \alpha \right\}, \quad \text{and}$$

$$S_{\alpha,1} := \left\{ (x,y) : \sum_{i=1}^n \alpha_i x_i = \alpha \text{ and } \sum_{i=1}^n \beta_i y_i \geq \mu - \alpha \right\}.$$

Let $G_{\alpha,0}$ and $G_{\alpha,1}$ be (disjoint) submatrices of G defined by the entries $S_{\alpha,0}$ and $S_{\alpha,1}$, respectively. It is clear that $G_{\alpha,0}$ is an all-0 matrix and $G_{\alpha,1}$ is an all-1 matrix for any α. Furthermore, $G = \sum_\alpha (G_{\alpha,0} + G_{\alpha,1})$. Hence, by subadditivity of rank we see that the rank of G is at most the number of distinct values taken by α. The latter is bounded above by $\alpha_{\max} - \alpha_{\min} + 1 \leq \sum_{i=1}^n |\alpha_i| + 1$. □

Note that the same proof goes through even if we generalize the definition of G by setting $G(x,y) = 1$ if and only if $\sum_{i=1}^n \alpha_i x_i + \sum_{i=1}^n \beta_i y_i \in T$ for an *arbitrary* subset T of $\mathbb{Z}$. Specifically, we have the following corollary of the proof.

Corollary 4.27. Let $G : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a symmetric function in the sense that its value depends only on $\sum_{i=1}^n (x_i + y_i)$. Then, its matrix G as defined above has rank at most $n + 1$.

Lemma 4.28. Let $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ be a Boolean function computed by a depth-2 threshold circuit C with the top gate using *unrestricted weights* and each of the bottom gates using weights of absolute value at most W. Then, there is a real matrix F such that $\mathrm{sign}(F(x,y)) = f(x,y)$ for all $x,y \in \{0,1\}^n$ and $\mathrm{rank}(F) = O(snW)$, where s is the number of bottom gates.

Proof. Let the top gate of C have weights $\phi_1,\dots,\phi_s$ and threshold ϕ_0. Hence, we can write $f(x,y) = \mathrm{sign}(\sum_{i=1}^{s} \phi_i G_i(x,y) - \phi_0)$, where G_i are the functions computed by the bottom gates. Define the matrix

$$F := \phi_1 G_1 + \cdots + \phi_s G_s - \phi_0 J,$$

where J is the all 1's $2^n \times 2^n$ matrix. It is clear that $f(x,y) = \mathrm{sign}(F(x,y))$ for all $x,y \in \{0,1\}^n$. Moreover, $\mathrm{rank}(F) \le 1 + \sum_{i=1}^{s} \mathrm{rank}(G_i) \le 1 + s(1 + nW)$ using Lemma 4.26. □

Using Corollary 4.27, one can similarly prove that if f is computed by a depth-2 circuit with a threshold gate at the top and s symmetric gates at the bottom, then there is a matrix F of rank $O(sn)$ that sign-represents f.

Proof. [of Theorem 4.24]. By Lemma 4.28, if a depth-2 threshold circuit computes $f_n(x,y)$, then there is a matrix F_n such that $\mathrm{sign}(F_n(x,y)) = \mathrm{sign}(f_n(x,y))$ and $\mathrm{rank}(F_n) = O(snW)$. On the other hand, $\mathrm{rank}(F_n) \ge$ sign-rank(A_f). Comparing the upper and lower bounds on $\mathrm{rank}(F)$, we get $snW = \Omega(\text{sign-rank}(A_f))$. This proves Part (a) of the theorem. Part (b) is proved similarly by means of Corollary 4.27. □

4.5.4 Probabilistic OBDD's

We briefly recall the technical definitions concerning OBDD's. An *ordered binary decision diagram* (*OBDD*) over Boolean variables $x_1,\dots,x_n$ is given by a directed acyclic graph $G = (V,E)$. The graph G contains exactly one source (node without a proper predecessor) and two sinks (nodes without a proper successor). One of the sinks is labeled

-1 and the other-one is labeled $+1$. Each nonsink node is labeled by a Boolean variable such that, on each path from the source to a sink, each variable occurs exactly once and they always occur in the same order.[2] Each nonsink has two outgoing edges. One of them is labeled 0, the other one is labeled 1. Each OBDD G represents a Boolean function f_G which is evaluated on $a \in \{0,1\}^n$ as follows. Vector a enters G at the source and is then routed through G until it reaches a sink. More specifically, at a node labeled x_i, vector a is routed through the outgoing edge labeled a_i. According to this rule, a will be routed either to the (-1)-sink or to the $(+1)$-sink. In the former case, we have $f_G(a) = -1$; in the latter case $f_G(a) = 1$. The *size* of an OBDD is the number of its nodes.

A *probabilistic ordered binary decision diagram* (*POBDD*) G is defined analogously, except that the routing procedure becomes probabilistic. Loosely speaking, the deterministic transitions along edges labeled either 0 or 1 are replaced by random transitions. Formally, this can be achieved as follows. With each node v on layer l of G, we associate two probability distributions D_0 and D_1 that assign probabilities to all nodes on layer $l+1$. If v is labeled x_i and vector a is currently located at v, then a is routed to node w on layer $l+1$ with probability $D_{a_i}(w)$. Each POBDD G represents a collection $(B_G(x))$ of Bernoulli random variables in the obvious sense. We say that G *realizes* Boolean function $f : \{0,1\}^n \to \{-1,+1\}$ if $f(x) = 1$ implies that $\Pr(B_G(x) = 1) > 1/2$, whereas $f(x) = -1$ implies that $\Pr(B_G(x) = 1) < 1/2$. The *size* of a POBDD is the number of its nodes.

POBDD's are quite powerful because they need only a slight advantage over random guessing. We present here exponential lower bounds on the size of POBDD's for concrete functions. In the sequel, we present a function IP_n that is provably hard to realize with POBDD's. Our function IP_n embodies ip_n (the inner product modulo 2 function) as a subfunction. Note however that ip_n itself is easy to compute by OBDD's

[2] Since we say that each variable occurs exactly once (as opposed to at most once) on each path, we implicitly make the assumption that G is *layered*, where all nodes in the same layer are labeled by the same variable and edges go from one layer to the next one according to the fixed order of the variables. This assumption is made only for the sake of simple technical exposition.

(even for deterministic ones). We will define IP_n in such a way, that, no matter how the variable order in the OBDD is chosen, there will be a subfunction of the form ip_n such that the chosen variable order is actually "weird."

We move on to the definition of IP_n. We use $2n$ Boolean variables $z_1,\dots,z_{2n}$ and, in addition, $2n$ auxiliary Boolean variables $h_1,\dots,h_{2n}$. Assume that the Hamming weight of h is n and that the 0-entries occur in positions $i(1) < \cdots < i(n)$, whereas the 1-entries occur in positions $j(1) < \cdots < j(n)$. Then, we define

$$\mathrm{IP}_n(z,h) := (-1)^{\sum_{k=1}^{n} z_{i(k)} z_{j(k)}}. \tag{4.6}$$

If the Hamming weight of h differs from n we define $\mathrm{IP}_n(z,h) := 1$ by default.

Theorem 4.29. Any POBDD G that computes IP_n has size at least $2^{n/2}$.

Proof. Assume that the variables $z_1,\dots,z_{2n}$ appear in G in the order $z_{\sigma(1)},\dots,z_{\sigma(2n)}$. Let $i(1) < \cdots < i(n)$ be the sorted sequence of the elements of $\{\sigma(1),\dots,\sigma(n)\}$. Likewise, $j(1) < \cdots < j(n)$ denotes the sorted sequence of all the elements of $\{\sigma(n+1),\dots,\sigma(2n)\}$. Note that in G all $z_{i(k)}$-variables occur before all $z_{j(k)}$-variables. Fix a vector h of Hamming weight n such that $h_{i(k)} = 0$ and $h_{j(k)} = 1$ for $k = 1,\dots,n$. This leads to a subfunction of the form (4.6). Let V' be the set of nodes labeled $z_{j(1)}$. In what follows, we write input z in the form (x,y), where $x = (z_{i(1)},\dots,z_{i(n)})$ and $y = (z_{j(1)},\dots,z_{j(n)})$. For each $v \in V'$, denote by $p_v(x)$ the probability that $z = (x,y)$ is routed from the source to v. Denote by $q_v(y)$, the probability that z is routed from v to the (+1)-sink. We conclude that

$$\mathrm{IP}_n(x,y,h) = 1 \Leftrightarrow \sum_{v \in V'} p_v(x) q_v(y) > \frac{1}{2}. \tag{4.7}$$

Note that, for our fixed choice of h, the subfunction $\mathrm{IP}_n(x,y,h)$ is the inner product modulo 2 of x and y. Equation (4.7) realizes this subfunction by a $|V'|$-dimensional arrangement that uses half spaces induced by inhomogeneous hyperplanes. Adding one dimension, the hyperplanes can be made homogeneous (and we arrive at a linear arrangement). By Forster's Theorem 4.2, $|V| \geq |V'| + 1 \geq 2^{n/2}$. □

Our lower bound can be extended to k-POBDD's: k-POBDD's generalize POBDD's in the following way. On each path from the source to a sink the variables appear at most k times in an order, i.e., the same permutation of variables repeated k times, i.e., a k-POBDD can be partitioned into k layers such that each layer is read-once ordered according to the same variable ordering. From Example 5.11, we conclude that every probabilistic two-way protocol for IP_n has length at least $n/2 - 1$. Using this we obtain the following.

Corollary 4.30. Each k-POBDD, $k \geq 2$, computing IP_n has size at least $2^{\frac{n-2}{4k-2}-1}$.

Indeed, if there exists a k-POBDD P of size S then one can find a probabilistic communication protocol of $2k$ rounds that computes the same function as P and has length $(2k-1)\lceil \log S \rceil$. In this communication game, the input variables of processor Π_0 are the variables that are read in the first half of each layer of P and the input variables of processor Π_1 are the variables read in the second half of each layer, respectively. The computation can be performed in the way that the processors follow a path in P in turn such that each processor sends to the other one the number of the first node on the path not belonging to its sublayer. This well-known technique (e.g., [15]) yields the desired lower bound.

5

Rank Robustness and Two-Party Communication Complexity

5.1 The Log-Rank Conjecture in Communication Complexity

For every graph G, is it true that

$$\log \chi(\overline{G}) \leq (\log \operatorname{rank}(G))^{O(1)},$$

where $\chi(\overline{G})$ is the chromatic number of the complement of G and $\operatorname{rank}(G)$ is the rank of the adjacency matrix of G over $\mathbb{R}$? This combinatorial question is equivalent to a complexity question [62]:

> Let $f : \{0,1\}^t \times \{0,1\}^t \to \{0,1\}$ and let cc(f) denote the two-party communication complexity of f. Let $M_f(x,y) = f(x,y)$. Then, is it true that for all f
>
> $$\operatorname{cc}(f) \leq (\log \operatorname{rank}(M_f))^{O(1)}?$$
>
> Here also the rank of M_f is taken over $\mathbb{R}$.

It is well-known [48, Section 1.4] that $\operatorname{cc}(f) \geq \log \operatorname{rank}(M_f)$ (in this inequality, rank over any field would do). Thus, answering the above question positively would resolve the famous log-rank conjecture in communication complexity. This is by far the most challenging open question in two-party communication complexity.

The most significant progress so far on this conjecture has been made by Nisan and Wigderson [68]. They show that (i) to resolve the log-rank conjecture, it suffices to show that *every* 0–1 matrix A of rank r contains a "large" submatrix of rank at most $(1-\epsilon)r$ for some constant ϵ; here, large means having a fraction at least $\exp(-(\log \operatorname{rank} A)^{O(1)})$ of entries of A, and (ii) there is an f such that $\mathrm{cc}(f) \geq (\log \operatorname{rank}(M_f))^{\log_3 6}$. We will discuss these results below.

In what follows, $|M|$ denotes the *total* number of entries of M (including zero entries).

Theorem 5.1. If,

> for every 0–1 matrix M of rank r and for some $\delta \geq \exp(-\log^c r)$ for some constant $c > 0$, there exists a submatrix S with at least $\delta|M|$ entries such that $\operatorname{rank}(S) \leq 0.25r$,

then M has a protocol tree with at most $\exp(O(\log^{c+1} r))$ leaves. In particular, M can be covered by at most $\exp(O(\log^{c+1} r))$ disjoint monochromatic rectangles.

Remark 5.1.

- Note that instead of 0.25 in the assumption above, any constant < 1 would do since by repeatedly taking submatrices a constant number of times, we can still get a submatrix of the claimed size (with a different c) of rank $\leq 0.25r$.
- If $|M| \leq \exp(\log^{c+1} r)$, then the trivial protocol already gives a communication complexity of $O(\log|M|) = O(\log^{c+1} r)$. Hence, we can assume w.l.o.g., that $|M| \geq \exp(\log^{c+1} r)$. This guarantees that a submatrix with $\delta|M|$ entries is not too small.

Proof. Let $L(m,r)$ denote the maximum, over all 0–1 matrices M with m entries and rank r, of the minimum number of leaves in a two-party protocol tree for M. Clearly, $L(m,1) \leq 2$ for all m. We wish to prove that $L(m,r) \leq \exp(\log^{c+1} r)$ for all m assuming the hypothesis of the theorem.

Let M be indexed by $X \times Y$, i.e., rows by $x \in X$ and columns by $y \in Y$, and let S be indexed by $X_0 \times Y_0$ for $X_0 \subseteq X$ and $Y_0 \subseteq Y$. Consider the partition of $X \times Y$ into the four parts $X_\alpha \times Y_\beta$ for $\alpha, \beta \in \{0,1\}$, where $X_1 = X \backslash X_0$ and $Y_1 = Y \backslash Y_0$. The corresponding submatrices of M are denoted $S_{\alpha\beta}$, where $S = S_{00}$:

$$M = \left[\begin{array}{c|c} S_{00} & S_{01} \\ \hline S_{10} & S_{11} \end{array}\right].$$

Observe that

$$\operatorname{rank} S_{01} + \operatorname{rank} S_{10} \leq \operatorname{rank} M + \operatorname{rank} S_{00}.$$

W.l.o.g., $\operatorname{rank} S_{01} \leq \operatorname{rank} S_{10}$. Then $\operatorname{rank} S_{01} \leq (\operatorname{rank} M + \operatorname{rank} S_{00})/2 \leq \frac{5}{8} \operatorname{rank} M$ since $\operatorname{rank} S_{00} \leq \frac{1}{4} \operatorname{rank} M$. It follows that

$$\operatorname{rank}[S_{00} \,|\, S_{01}] \leq \frac{1}{4} \operatorname{rank} M + \frac{5}{8} \operatorname{rank} M = \frac{7}{8} \operatorname{rank} M.$$

By definition, then, the top part $[S_{00} \,|\, S_{01}]$ has a protocol tree with most $L(m, 7r/8)$ leaves, while the bottom part $[S_{10} \,|\, S_{11}]$ has a protocol tree with at most $L((1-\delta)m, r)$ leaves. The row player will send one bit to the column player to indicate whether his input $x \in X$ is in the top part or the bottom part. (If $\operatorname{rank} S_{01} > \operatorname{rank} S_{10}$, the players would consider the partition into the left and the right parts and then the column player would send a bit.) They will then proceed recursively on either part of the matrix.

Thus, we have the recurrence relation

$$L(m,r) \leq L\left(m, \frac{7r}{8}\right) + L((1-\delta)m, r). \tag{5.1}$$

The above recurrence relation has the solution $L(m,r) = \exp(O(\log^{c+1} r))$. □

Corollary 5.2. If, for every 0–1 matrix M of rank r and for some $\delta \geq \exp(-\log^c r)$ with some constant $c > 0$, there exists a submatrix S with at least $\delta|M|$ entries such that $\operatorname{rank} S \leq 0.25r$, then for any Boolean function $f : \{0,1\}^t \times \{0,1\}^t \to \{0,1\}$, $\operatorname{cc}(f) \leq (\log \operatorname{rank}(M_f))^{c+1}$. In other words, the log-rank conjecture holds under the hypothesis stated in Theorem 5.1.

Proof. It is known that if the $\{0,1\}$-matrix M_f has a communication protocol with L leaves, then f has communication complexity at most $O(\log L)$ (see [48, Lemma 2.8].). □

To show the best known gap between the rank and communication complexity of a matrix, we need to review the notions of sensitivity and degree of a Boolean function. We will construct a function with maximum sensitivity and relatively low degree as a polynomial over $\mathbb{R}$. A communication matrix defined in terms of this function will have low rank (owning to the low degree) and high communication complexity (owing to the maximum sensitivity).

Definition 5.1. For a Boolean function $g : \{0,1\}^t \to \{0,1\}$, define the *sensitivity of* g, $\operatorname{sens}(g)$, as

$$\operatorname{sens}(g) := \max_{x \in \{0,1\}^t} |\{y \in \{0,1\}^t \; : \; g(y) \neq g(x), \operatorname{dist}(x,y) = 1\}|,$$

where $\operatorname{dist}(x,y)$ denotes the Hamming distance between x and y. In other words, define the sensitivity of g *at input* x to be the number of Hamming neighbors of x where g takes a value different from $g(x)$ itself. Now, sensitivity of g is defined as the maximum sensitivity of g at any input.

For example, the OR, the AND, and the PARITY functions on n variables all have sensitivity n. The MAJORITY function, on the other hand, has sensitivity $(n+1)/2$ for odd n.

Definition 5.2. For a Boolean function $g : \{0,1\}^t \to \{0,1\}$, the *degree of* g, $\deg(g)$ is defined as the degree of the unique polynomial over $\mathbb{R}$

that represents g. Note that the $\deg(g)$ is also the largest $|S|$, $S \subseteq [t]$, such that the Fourier coefficient $\hat{g}(S) \neq 0$.

The exact relation between degree and sensitivity of Boolean functions is an intriguing open question in Boolean function complexity. What is currently known is summarized in the next two theorems by Nisan and Szegedy [66] and Kushilevitz (stated in [66]).

Theorem 5.3. For every Boolean function g, $\deg(g) \geq c\sqrt{\mathrm{sens}(g)}$, for some absolute constant $c > 0$.

Theorem 5.4. There exists a t-variable Boolean function g such that $\mathrm{sens}(g) = t$, but $\deg(g) \leq t^{\log_6 3} \approx t^{0.6131}$.

Proof. Let g_0 be the following 6-variable function (defined as a real polynomial of degree 3):

$$g_0(x_1,\ldots,x_6) = \sum_{i=1}^{6} x_i - \sum_{i \neq j} x_i x_j + \sum_{\{i,j,k\} \in \mathcal{F}} x_i x_j x_k,$$

where $\mathcal{F}$ is the following set system on $\{1,\ldots,6\}$:

$$\{\{1,2,5\},\{1,2,6\},\{1,3,4\},\{1,3,5\},\{1,4,6\},\{2,3,4\},$$
$$\{2,3,6\},\{2,4,5\},\{3,5,6\},\{4,5,6\}\}.$$

It's easy to verify that in $\mathcal{F}$,

- Every point is contained in exactly five sets.
- Every pair of points is contained in exactly two sets.

Using these properties it is easy to verify that g_0 is indeed a Boolean function on $\{0,1\}^6$. Its degree is clearly 3. It is also clear that $g(\underline{0}) = 0$ and that $g(x) = 1$ for any x with exactly one 1. Thus, g has sensitivity 6.

We now recursively define g_k on $t := 6^k$ variables as follows:

$$g_k(x_1,\ldots,x_t) = g_0(g_{k-1}(x_1,\ldots,x_{t/6}),\ldots,g_{k-1}(x_{5t/6+1},\ldots,x_t)),$$

i.e, g_0 is applied on 6 copies of g_{k-1} on 6^{k-1} variables each.

It is easy to see by induction on k that $\mathrm{sens}(g_k) = t$ and that $\deg(g_{k-1}) = 3^k = t^{\log_6 3}$. □

Definition 5.3. For a Boolean function $g : \{0,1\}^t \to \{0,1\}$, define the function $g_\wedge : \{0,1\}^t \times \{0,1\}^t \to \{0,1\}$ as follows:

$$g_\wedge(x,y) := g(x_1 \wedge y_1, \ldots, x_t \wedge y_t).$$

Lemma 5.5. If $g : \{0,1\}^t \to \{0,1\}$ has sensitivity t, then $\mathrm{cc}(g_\wedge) = \Omega(t)$.

Proof. W.l.o.g. (by shifting g if necessary), we can assume that $g(0) = 0$ and $\forall i$, $g(e_i) = 1$, where e_i has zeros in all coordinates except the i'th, where it has a 1. Thinking of x and y as characteristic vectors of sets $X, Y \subseteq [t]$, note that $g_\wedge(x,y) = 0$ if $X \cap Y = \emptyset$ and that $g_\wedge(x,y) = 1$ if $|X \cap Y| = 1$.

Now, a well-known lower bound [48, Section 4.6] on the *set disjointness problem* shows that even the randomized communication complexity of the problem of deciding whether the intersection between two sets is the empty set or a singleton is linear. For a more recent proof of this result, see [37].

This immediately gives the required reduction. □

Lemma 5.6. If $g : \{0,1\}^t \to \{0,1\}$ has degree d over $\mathbb{R}$, then the $2^t \times 2^t$ 0–1 matrix $M(g_\wedge)$ corresponding to $g_\wedge$ has rank at most $\sum_{i=0}^{d} \binom{t}{d} = O(t^d)$.

Proof. Consider the polynomial representation of g:

$$g(z_1, \ldots, z_t) = \sum_{S \subseteq [t], |S| \le d} \alpha_s \prod_{i \in S} z_i,$$

and note that $g_\wedge$ is similarly expressed as a polynomial:

$$g_\wedge(x_1, \ldots, x_t, y_1, \ldots, y_t) = \sum_{S \subseteq [t], |S| \le d} \alpha_s \prod_{i \in S} x_i \prod_{j \in S} y_j.$$

Now, to construct the matrix $M(g_\wedge)$, it suffices to construct matrices corresponding to each monomial and add them up (with the corresponding coefficients α_S) entry-wise. On the other hand, the 0–1 matrix corresponding to each monomial $\prod_{i\in S} x_i \prod_{j\in S} y_j$ is easily seen to be a rank-1 matrix (it's a rectangle of 1's given by the rows and columns corresponding to the subcubes defined by the products $\prod_{i\in S} x_i$ and $\prod_{j\in S} y_j$, respectively). It follows that the rank of $M(g_\wedge)$ is bounded by the number of monomials in the polynomial representing the function g and the proof is complete. □

Theorem 5.7. There is a Boolean function f such that $\mathrm{cc}(f) \geq (\log \operatorname{rank} M_f)^c$, where $c = \log_3 6 - o(1) \approx 1.631$.

Proof. Let f be $g_\wedge$ where g is as in Theorem 5.4. Then $\mathrm{cc}(f) = \Omega(t)$ from Lemma 5.5 since g is fully sensitive. On the other hand, Lemma 5.6 combined with the degree bound from Theorem 5.4 shows that $\operatorname{rank}(M_f) \leq O(t^d)$, where $d \leq t^{\log_6 3}$. Hence,

$$\log \operatorname{rank} M_f = (d \log t) + O(1) = t^{\log_6 3 + o(1)} \leq \mathrm{cc}(f)^{\log_6 3 + o(1)}. \qquad \square$$

Note that the method we used to prove the above theorem cannot yield more than a quadratic gap between log-rank and communication complexity, in view of Theorem 5.3.

5.2 Discrepancy and Randomized and Quantum Communication Complexity

Just as rank of a sign matrix is a natural lower bound on the deterministic communication complexity of the corresponding Boolean function, discrepancy turns out to be a natural lower bound on both the randomized and quantum communication complexity. In this subsection, we present some of these results.

Recall the definition of discrepancy (cf. Definition 4.7) $\mathrm{disc}(M)$ of a sign matrix M. Let $\mathrm{disc}_U(M)$ denote the discrepancy of M under the uniform distribution, i.e., $\mathrm{disc}_U(M) = \max_R ||R^+| - |R^-||/|M|$, where the maximum is taken over all combinatorial rectangles R contained in M. Nisan and Wigderson [68] also prove[1] that for a rank-r matrix M, $\mathrm{disc}_U(M) \geq \epsilon r^{-3/2}$ for some constant $\epsilon > 0$. However, from the results of [53], we get an even stronger lower bound on general discrepancy itself, i.e., $\mathrm{disc}(M)$ instead of $\mathrm{disc}_U(M)$: $\mathrm{disc}(M) \geq \frac{1}{8} r^{-1/2}$. To prove this lower bound and for use in later results, we use a generalization of the γ_2-norm of a matrix, defined in Definition 4.4 and introduced by Linial et al. [53].

Definition 5.4. For a sign matrix $M = (a_{ij})$, i.e., $M \in \{-1, +1\}^{m \times n}$, and a real number $\alpha \geq 1$,

$$\gamma_2^{\alpha}(M) := \min\{\gamma_2(B) : 1 \leq a_{ij} b_{ij} \leq \alpha \ \ \forall\, i, j\}.$$

Clearly, if $\alpha \leq \beta$, then $\gamma_2^{\alpha}(M) \geq \gamma_2^{\beta}(M)$ for any sign matrix M. In particular, $\gamma_2^{\infty}(M) \leq \gamma_2(M)$. Now, use Corollary 4.3 and Theorem 4.18 to conclude

$$8\,\mathrm{disc}(M) \geq \mathrm{margin}(M) = (\gamma_2^{\infty}(M))^{-1} \geq (\gamma_2(M))^{-1} \geq (\mathrm{rank}(M))^{-1/2}.$$

For a sign matrix M, let $R_\epsilon(M)$ be the randomized communication complexity of the corresponding Boolean function f_M with success probability $\geq 1/2 + \epsilon$. Similarly, let Q_ϵ be the quantum communication complexity (with prior entanglement) of f_M. It is well-known that both these communication complexity measures are lower bounded by $\Omega(\log((1-2\epsilon)/\mathrm{disc}(M)))$. Linial and Shraibman [54] prove more general lower bounds in terms of γ_2^{α}-norms.

Theorem 5.8. With the above notation,

$$R_\epsilon(M) \geq 2\log\left(\gamma_2^{1/2\epsilon}(M)\right) + 2\log 2\epsilon, \quad \text{and}$$

$$Q_\epsilon(M) \geq \log\left(\gamma_2^{1/2\epsilon}(M)\right) + \log 2\epsilon.$$

[1] Their result is more generally expressible in terms of the spectral norm: $\mathrm{disc}_U(M) \geq \epsilon\,(\|M\|/n)^3$.

Proof. By definition, a randomized communication protocol for f_M is a probability distribution μ on deterministic protocols for f_M such that, for each input, the probability mass on the protocols computing the correct answer is at least $1/2+\epsilon$. Let D_π be the ± 1-matrix corresponding to a deterministic protocol π (making some errors) for f_M and let $\mu(\pi)$ be the probability with which π is chosen. Recall that D_π can be partitioned into at most 2^c monochromatic rectangles, where $c := R_\epsilon(M)$ is an upper bound on the communication complexity of D_π. Hence, $\operatorname{rank}(D_\pi) \le 2^c$. By Lemma 4.16, $\gamma_2(D_\pi) \le \sqrt{\operatorname{rank}(D_\pi)} \le 2^{c/2}$. Consider the matrix

$$N := \frac{1}{2\epsilon}\sum_\pi \mu(\pi) D_\pi.$$

Clearly, $|N_{ij}| \le 1/2\epsilon$ for all i,j. When $M_{ij} = +1$ (-1), $\sum_\pi \mu(\pi)(D_\pi)_{ij} \ge 2\epsilon$ ($\le -2\epsilon$, respectively). Hence, $1 \le M_{ij}N_{ij} \le 1/2\epsilon$. Finally, since γ_2 is a norm,

$$\gamma_2(N) \le \frac{1}{2\epsilon}\sum_\pi \mu(\pi)\gamma_2(D_\pi) \le \frac{1}{2\epsilon}2^{c/2}.$$

It follows that $\gamma_2^{1/2\epsilon}(M) \le \frac{1}{2\epsilon}2^{c/2}$. This proves the first part of the theorem.

The second part of the theorem is based on the following lemma. A proof of this lemma can be found in [54].

Lemma 5.9. For a sign-matrix M, let $P := (p_{ij})$ be the acceptance probabilities of a quantum communication protocol with prior entanglement for f_M and complexity c. Then, $P = XY$, where the matrices X and Y satisfy $\|X\|_{2\to\infty}, \|Y\|_{1\to 2} \le 2^{c/2}$.

If prior entanglement is not used, the matrices X and Y can be taken to have rank at most 2^{2c}.

Given this lemma, the proof of the second part of the theorem is similar to the first part by taking N to be $\frac{1}{2\epsilon}(2P - J)$, where J is the all 1's matrix. □

5.3 Relating Probabilistic Communication Complexity to Sign-Rank and Margin Complexity

We say that a probabilistic communication protocol computes the function $f : \{0,1\}^t \times \{0,1\}^t \to \{-1,1\}$ with *unbounded error* if for all inputs $(x,y) \in \{0,1\}^t \times \{0,1\}^t$ the correct output is calculated with probability greater than $1/2$. Since the slightest advantage over random guessing is already sufficient, this model is called the *unbounded error* model. Let $n := 2^t$. The length of a communication protocol is $\lceil \log_2 K \rceil$, where K is the number of distinct message sequences that can occur in the protocol. The *unbounded error probabilistic communication complexity* [71] UPP-CC(f) of a function $f : \{0,1\}^t \times \{0,1\}^t \to \{-1,1\}$ is the smallest length of a communication protocol that computes f with unbounded error (in the sense defined above).

We briefly note that Paturi and Simon [71] have shown that, in the unbounded error model, any probabilistic two-way protocol of length k can be converted into a one-way protocol with length at most $k+1$.

Furthermore, the minimal dimension or the sign-rank of a realization (cf. Definitions 4.1 and 4.2) of the sign-matrix A_f of a Boolean function $f(x,y)$ is closely related to its probabilistic communication complexity UPP-CC(f). Paturi and Simon [71] have proved the following relation:

$$\lceil \log \text{sign-rank}(A_f) \rceil \leq \text{UPP-CC}(f) \leq \lceil \log \text{sign-rank}(A_f) \rceil + 1 \quad (5.2)$$

Using Theorem 4.2, we conclude the following.

Theorem 5.10. For any function $f : \{0,1\}^t \times \{0,1\}^t \to \{-1,1\}$, UPP-CC$(f) \geq n/\|A_f\|$.

Example 5.11. Let $\mathrm{ip}_t(x,y) := (-1)^{x^\top y}$, where $x,y \in \mathbb{Z}_2^t$, be the `inner product mod-2` function. It is well known that the corresponding matrix H such that $H_{x,y} = \mathrm{ip}_n(x,y)$ is an Hadamard matrix and easy to see that H has operator norm $2^{t/2}$. Hence, from Theorem 5.10, UPP-CC$(\mathrm{ip}_t) \geq t/2$.

Given a probabilistic communication protocol, let $\epsilon(x,y)$ denote the difference between the probability that (x,y) is accepted and the probability that (x,y) is rejected by the protocol. Thus, $f(x,y) = \text{sign}(\epsilon(x,y))$ for any probabilistic protocol that correctly computes f. The *error bound* of the protocol is defined as $\min_{x\in X,y\in Y}|\epsilon(x,y)|$. Hence, we only require that $|\epsilon(x,y)| > 0$ for all (x,y) in the Paturi–Simon model. We say a function family $f = (f_t)$ is in the communication complexity class UPP if each f_t has an unbounded error probabilistic protocol with $2^{\text{polylog}(t)}$ messages.

A weakening of the above definition gives a communication complexity analog of the class PP in the Turning Machine world. Here, we define that a family $f = (f_t)$ of Boolean functions $f_t : \{0,1\}^t \times \{0,1\}^t \to \{-1,+1\}$ belongs to PP^{cc} if there exists a probabilistic one-way protocol that transmits at most polylog(t) bits (uses at most $2^{\text{polylog}(t)}$ messages) and achieves error bound $2^{-\text{polylog}(t)}$.

Halstenberg and Reischuk [36] have shown that the class PP^{cc} does not change if we allow only *one-way* protocols. The class PP^{cc} is one of the main complexity classes in the *bounded-error* model for probabilistic communication complexity.

Interestingly, it is possible to express membership in PP^{cc} in terms of only one parameter: the *maximal margin* (cf. Sections 4.3 and 4.5.2). Here is the connection.

Theorem 5.12. $(f_t) \in \text{PP}^{\text{cc}} \Leftrightarrow \text{margin}(A_f) \geq 2^{-\text{polylog}(t)}$.

Combining this theorem with Theorem 4.18, we also get a characterization of PP^{cc} in terms of discrepancy. This characterization was originally discovered by Klauck [43] using Fourier transform techniques.

Corollary 5.13. $(f_t) \in \text{PP}^{\text{cc}} \Leftrightarrow \text{disc}(A_f) \geq 2^{-\text{polylog}(t)}$.

Theorem 5.12 will follow from Lemmas 5.14, 5.15, and 5.16 below. Lemma 5.15 makes use of a random projection technique from [3]. Lemmas 5.14 and 5.16 are implicitly proven in [71] on probabilistic communication complexity with unbounded error.

Lemma 5.14. Each probabilistic one-way protocol for f that uses at most K messages and achieves error-bound ϵ can be converted into a K-dimensional linear arrangement that realizes f with margin $\mu \geq \epsilon/\sqrt{K}$.

Proof. Let $p_i(x)$ be the probability that Player 1 sends the i'th message on input x. Let $q_i(y)$ be the probability that Player 2 outputs a 1 upon receiving the i'th message on input y. It is clear that $\epsilon(x,y) = \sum_{i=1}^{K} p_i(x)(2q_i(y) - 1)$. Define

$$u_x := (p_i(x))_{i=1}^{K}, \quad v_y = (2q_i(y) - 1)_{i=1}^{K}.$$

It is easy to see that $\{u_x/\|u_x\| : x \in \{0,1\}^t\}$ and $\{v_y/\|v_y\| : y \in \{0,1\}^t\}$ define a realization of the matrix of f in K-dimensions. Furthermore, $\|u_x\| \leq \|u_x\|_1 = 1$ since it is a vector of probabilities and $\|v_y\| \leq \sqrt{K}$ since each of its entries is at most 1 in absolute value. It follows that the margin of this realization is

$$\mu \geq \min_{x,y} |\langle u_x/\|u_x\|, v_y/\|v_y\|\rangle| \geq \frac{1}{\sqrt{K}} \min_{x,y} |\epsilon(x,y)| = \frac{\epsilon}{\sqrt{K}}. \qquad \square$$

This proves one direction of Theorem 5.12. The other direction follows by combining the next two lemmas.

Lemma 5.15. Each linear arrangement (of arbitrarily high dimension) that realizes f with margin μ can be converted into an $O(t/\mu^2)$-dimensional linear arrangement that realizes f with margin $\mu/2$.

Proof. (sketch) The following result (whose proof can be found in [11]), is based on the technique of random projections using the Johnson–Lindenstrauss lemma (see [3]):

> Let $w, x \in \mathbb{R}^r$ be arbitrary but fixed. Let $R = (R_{i,j})$ be a random $(k \times r)$-matrix such that the entries $R_{i,j}$ are i.i.d. according to the normal distribution $N(0,1)$.

Consider the random projection $u_R := \frac{1}{\sqrt{k}}(Ru) \in \mathbb{R}^k$ for all $u \in \mathbb{R}^r$. Then, the following holds for every $\mu > 0$:

$$\Pr_R\left[|\langle w_R, x_R\rangle - \langle w, x\rangle| \geq \frac{\mu}{4}\left(\|w\|^2 + \|x\|^2\right)\right] \leq 4\mathrm{e}^{-\mu^2 k/32}.$$

This result can be used in an obvious way to guarantee the existence of a random projection that maps an r-dimensional linear arrangement that realizes f with margin μ to a k-dimensional linear arrangement that realizes f with margin $\mu/2$ by choosing $k := ct/\mu^2$ for a suitable positive constant c. □

Lemma 5.16. Each K-dimensional linear arrangement that realizes f with margin μ can be converted into a probabilistic one-way protocol that uses at most $2K$ messages and achieves error-bound $\epsilon \geq \mu/\sqrt{K}$.

Proof. (sketch) Let u_x and v_y be unit vectors in $\mathbb{R}^K$ for $x, y \in \{0,1\}^t$ that realize the matrix A_f with margin μ. Player 1 sends a message $(i, \mathrm{sign}(u_x(i))$, where $u_x(i)$ denotes the i'th coordinate of u_x, on input x with probability:

$$p_i(x) := \frac{|u_x(i)|}{\|u_x\|_1}.$$

Clearly, the number of messages is $2K$. Now, Player 2 accepts on input y and receiving the i'th message from Player 1 with probability:

$$q_i(y) := \frac{\mathrm{sign}(u_x(i)) \cdot v_y(i) + 1}{2}.$$

By a calculation similar to that in the proof of Lemma 5.14, we can see that

$$\epsilon(x,y) = \frac{\langle u_x, v_y\rangle}{\|u_x\|_1} \geq \frac{\mu}{\sqrt{K}}.$$

Here, the inequality follows by the assumed bound on the margin given by u_x and v_y and by the inequality $\|u_x\|_1 \leq \sqrt{K}\|u_x\|_2 = \sqrt{K}$. □

5.4 Matrix Rigidity and Quantifier Alternation in Communication Complexity

Taking a complexity theoretic view of Yao's [104] model of two-party communication complexity, Babai et al. [5] defined analogs of various Turing Machine complexity classes. For example, they defined $\mathrm{PP}^{\mathrm{cc}}$,$\mathrm{PH}^{\mathrm{cc}}$, and $\mathrm{PSPACE}^{\mathrm{cc}}$. A class unique to communication complexity, namely, UPP, arising from the Paturi–Simon model was discussed in Section 5.3. Separating the complexity classes in the two-party communication model does not appear to be nearly as challenging as separating the corresponding classes in the Turing Machine model. Specifically, we know that P^{cc}, $\mathrm{NP}^{\mathrm{cc}}$, and $\mathrm{BPP}^{\mathrm{cc}}$ are all different from each other and that $\mathrm{NP}^{\mathrm{cc}} \cap \mathrm{Co\text{–}NP}^{\mathrm{cc}} = \mathrm{P}^{\mathrm{cc}}$ [48, Section 4.5]. We will also see later that UPP and Σ_2^{cc} are separated in a recent result by Razborov and Sherstov [88], resolving a long-standing open question in this area, and improving the result that $\mathrm{PSPACE}^{\mathrm{cc}} \not\subseteq \mathrm{UPP}$ [28]. We note that both these separation results follow from lower bounds on sign-rank we presented in Section 4. On the other hand, it is still a major challenge to separate $\mathrm{PH}^{\mathrm{cc}}$ and $\mathrm{PSPACE}^{\mathrm{cc}}$. We now relate this question to the notion $\mathcal{R}_A(r,\theta)$ of matrix rigidity (cf. Definition 3.2).

First, we review some of the definitions.

To define communication complexity classes, we consider languages consisting of pairs of strings (x,y) such that $|x| = |y|$. Denote by Σ^{2*} the universe $\{(x,y) : x,y \in \{0,1\}^* \text{ and } |x| = |y|\}$. For a language $L \subseteq \Sigma^{2*}$, we denote its characteristic function on pairs of strings of length m by L_n, where $n := 2^m$. L_n is naturally represented as an $n \times n$ matrix with 0–1 or ± 1 entries (with -1 for *true* and $+1$ for *false*).

Conversely, if $A = \{A_n\}$ is an infinite sequence of ± 1-matrices (where A_n is $n \times n$), then we can associate a language L_A with A and talk about its communication complexity. L_A is not necessarily unique (since the n's may be different from powers of two), but for the purposes of lower bounds we will fix one such language and refer to it as *the* language L_A corresponding to A.

We recall the following definitions from [5].

Definition 5.5. Let nonnegative integers $l_1(m),\ldots,l_k(m)$ be such that $l(m) := \sum_{i=1}^{k} l_i(m) \leq (\log m)^c$ for a fixed constant $c \geq 0$.

A language L is in Σ_k^{cc} if there exist $l_1(m),\ldots,l_k(m)$ as above and Boolean functions $\varphi,\psi : \{0,1\}^{m+l(m)} \to \{0,1\}$ such that $(x,y) \in L_n$ if and only if

$$\exists u_1 \forall u_2 \cdots Q_k u_k \; (\varphi(x,u) \diamondsuit \psi(y,u)),$$

where $|u_i| = l_i(m), u = u_1 \cdots u_k$, Q_k is $\forall$ for k even and is $\exists$ for k odd, and, $\diamondsuit$ stands for $\vee$ if k is even and for $\wedge$ if k is odd.

Definition 5.6.

- By allowing a bounded number of alternating quantifiers in Definition 5.5, we get an analog of the polynomial time hierarchy: $\mathrm{PH}^{\mathrm{cc}} \; = \bigcup_{k \geq 0} \Sigma_k^{\mathrm{cc}}$.
- We get an analog of PSPACE, by allowing an unbounded, but no more than polylog(m), number of alternating quantifiers in Definition 5.5: $\mathrm{PSPACE}^{\mathrm{cc}} = \bigcup_{c>0} \bigcup_{k \leq (\log m)^c} \Sigma_k^{\mathrm{cc}}$.

Theorem 5.17. Let L be a language in $\mathrm{PH}^{\mathrm{cc}}$ and A_n be its $n \times n$ ± 1-matrix, where $n := 2^m$. Then, for all constants $c \geq 0$, there exist constants $c_1, c_2 \geq 0$ such that

$$\mathcal{R}_{A_n}(2^{(\log\log n)^{c_1}}, 2^{(\log\log n)^{c_2}}) \leq \frac{n^2}{2^{(\log\log n)^c}}.$$

Corollary 5.18. Let $A = \{A_n\}$ be an infinite sequence of ± 1-matrices and L_A be the associated language. For some constant $c > 0$ and any r and θ in $2^{(\log\log n)^{\omega(1)}}$, if $\mathcal{R}_A(r,\theta) \geq n^2/2^{(\log\log n)^c}$, then $L_A \notin \mathrm{PH}^{\mathrm{cc}}$.

In particular, if A can be chosen such that $L_A \in \mathrm{PSPACE}^{\mathrm{cc}}$, then $\mathrm{PH}^{\mathrm{cc}} \; \subsetneq \mathrm{PSPACE}^{\mathrm{cc}}$.

Proof. (of Theorem 5.17) This theorem is proved using Tarui's [97] low-degree polynomials (over integers) that approximate AC^0-circuits.

The theorem will follow from the following: for all $c > 0$, there exist $c_1, c_2 > 0$, and integer matrices $\{B_n\}$, where B_n is $n \times n$, such that

(1) $\mathrm{rank}(B_n) \leq 2^{(\log\log n)^{c_1}}$.
(2) $\forall (x,y)$, $1 \leq |B_n(x,y)| \leq 2^{(\log\log n)^{c_2}}$.
(3) $\mathrm{wt}(A_n - B_n) \leq n^2/2^{(\log\log n)^c}$.

For simplicity of notation, let $L \in \Sigma_k^{\mathrm{cc}}$, where k is odd. In Definition 5.5 of Σ_k^{cc}, for any fixed sequence of moves $u = u_1,\ldots,u_k$, φ is a function of x and ψ is a function of y. Define $f_u(\cdot) \equiv \varphi(\cdot,u)$ and similarly $g_u(\cdot) \equiv \psi(\cdot,u)$. Replacing $\exists$ move by an OR-gate and $\forall$ move by an AND-gate, we see that L has a Σ_k^{cc} protocol iff it can be expressed as the output of an $\{AND, OR\}$ circuit C of depth k and size $2^{\mathrm{polylog}(m)}$, where the inputs of C are $f_u(x) \wedge g_u(y)$ for $1 \leq u \leq 2^{\mathrm{polylog}(m)}$. Hence, for all $(x,y) \in \{0,1\}^m \times \{0,1\}^m$,

$$L(x,y) = C(f_1(x) \wedge g_1(y), \ldots, f_t(y) \wedge g_t(y)), \tag{5.3}$$

where $t \leq 2^{\mathrm{polylog}(m)}$ is the number of possible u's.

Considering f_i as the characteristic function of a subset U_i of rows and g_i as that of a subset V_i of columns of the $\{0,1\}^m \times \{0,1\}^m$ matrix, we observe that $f_i(x) \wedge g_i(y)$ is a "rectangle" $U_i \times V_i$ in the matrix. We will denote this rectangle R_i and identify it with the corresponding $n \times n$ $\{0,1\}$-matrix of rank-1:

$$R_i(x,y) = \begin{cases} 1, & \text{if } f_i(x) \wedge g_i(y) = 1, \\ 0, & \text{otherwise.} \end{cases}$$

From Equation (5.3), it follows that L is in Σ_k^{cc} iff its matrix is expressible by an AC^0-circuit (of quasi-polynomial size) acting on a set of rank-1 matrices.

We now use the fact that an AC^0-circuit is well-approximated by a low-degree polynomial over $\mathbb{Z}$. Tarui [97] constructs such polynomials.

Theorem 5.19 (Tarui). Let C be an AC^0-circuit of size $2^{\mathrm{polylog}(t)}$ and $\phi_1,\ldots,\phi_t : \{0,1\}^s \to \{0,1\}$ be *arbitrary* Boolean functions. Fix

$0 < \delta = 2^{-(\log t)^{c'}}$, for some constant $c' \geq 0$. Then, there exist constants $c'_1, c'_2 \geq 0$, and a polynomial $\Phi(\phi_1,\ldots,\phi_t)$ such that

- *Low degree*: The degree of Φ in $\phi_1,\ldots,\phi_t$ is at most $(\log t)^{c'_1}$.
- *Small error*: The fraction of inputs $x \in \{0,1\}^s$, where $C(\phi_1,\ldots,\phi_t)(x) \neq \Phi(\phi_1,\ldots,\phi_t)(x)$ is at most δ.
- *Small norm*: The sum of the absolute values of the coefficients of Φ is at most $2^{(\log t)^{c'_2}}$.
- *Boolean guarantee*: When Φ differs from C, the value of $\Phi(\phi_1,\ldots,\phi_t)(x)$ is ≥ 2.

Let L_n be the $\{0,1\}$-matrix of L , i.e., $L_n = (J_n - A_n)/2$, where J_n is the $n \times n$ all 1's matrix.

From Equation (5.3), L_n is computed by an AC^0-circuit $C(z_1,\ldots,z_t)$ of size $2^{\text{polylog}(m)}$, where $z_i = f_i(x) \wedge g_i(y) = f_i(x)g_i(y)$ since f_i, g_i are $\{0,1\}$-functions. Using Theorem 5.19 for C, there is polynomial Φ of degree $d \leq \text{polylog}(t)$ such that $L(x,y) = \Phi(x,y)$, except for a $\delta = 2^{-(\log m)^c}$ (choose c' such that $(\log m)^c = (\log t)^{c'}$) fraction of $(x,y) \in \{0,1\}^m \times \{0,1\}^m$. We can write Φ as follows

$$\begin{aligned}\Phi(x,y) &= \sum_{S\subseteq[t],|S|\leq d} \alpha_S \prod_{i\in S} z_i \\ &= \sum_{S\subseteq[t],|S|\leq d} \alpha_S \prod_{i\in S} f_i(x)g_i(y) \\ &= \sum_{S\subseteq[t],|S|\leq d} \alpha_S\, f_S(x)g_S(y).\end{aligned}$$

Here, $f_S(x) = \prod_{i\in S} f_i(x)$ and similarly g_S.

Returning to our matrix interpretation, $f_S(x)g_S(y)$ is a $\{0,1\}$-matrix R_S of rank-1, and then, as a matrix, Φ is of rank at most $\sum_{i\leq d}\binom{t}{i} \leq 2^{\text{polylog}(t)}$. L and Φ agree on all but an ϵ fraction of the entries. Furthermore, by Theorem 5.19, the entries of Φ are all non-negative integers, and, > 1 if $L(x,y) \neq \Phi(x,y)$. Let us now define a matrix B_n:

$$B_n := J_n - 2\Phi = J_n - 2 \ \cdot \sum_{S\subseteq[t],|S|\leq d} \alpha_S R_S.$$

Clearly,

$$\operatorname{rank}(B_n) \leq 1 + \operatorname{rank}(\Phi) \leq 2^{\operatorname{polylog}(t)} \leq 2^{\operatorname{polylog}(m)},$$

thus proving (1); the value of constant c_1' implies constant c_1 since $\operatorname{polylog}(t) = \operatorname{polylog}(m)$ with suitable exponents. Entries of B_n are bounded in absolute value by $2^{\operatorname{polylog}(m)}$ and hence (2) is true; again c_2' implies c_2. Moreover, B_n differs from A_n in at most a $\delta = 2^{-(\log m)^c}$-fraction of entries. Thus (3) follows. (In fact, since Φ is at least 2 on the error points, B_n can only switch the signs of $+1$'s in A_n.) □

Recently, Linial and Shraibman [55] relate the question of finding an explicit language outside $\mathrm{PH}^{\mathrm{cc}}$ to the question of *soft margin complexity* or mc-rigidity. This notion measures the Hamming distance of a sign matrix to a matrix of small margin complexity.

Definition 5.7. The *mc-rigidity* function of a sign-matrix A at Hamming distance d is defined as follows

$$\text{mc-rigidity}(A, d) := \min\{\operatorname{mc}(B) : \operatorname{wt}(A - B) \leq d\},$$

where margin complexity $\operatorname{mc}(B)$ of a sign-matrix B is as defined at the beginning of Section 4.3.

Linial and Shraibman then show that an infinite family of sign matrices with sufficiently high margin complexity would yield a language outside $\mathrm{PH}^{\mathrm{cc}}$. Since the quantitative bounds and the proof techniques (mainly Theorem 5.19) are similar to those of Theorem 5.17, we refer the reader to [55] for details.

We conclude this section with a remarkable separation result in two-party communication complexity recently proved by Razborov and Sherstov [88].

Theorem 5.20. $\Sigma_2^{\mathrm{cc}} \not\subseteq \mathrm{UPP}$ and $\Pi_2^{\mathrm{cc}} \not\subseteq \mathrm{UPP}$.

Proof. By definition, the function f on t^3 bits given by (4.4) is computed by a depth-2 linear size circuit with rectangles at the inputs.

Hence, $f \in \Pi_2^{\mathrm{cc}}$. On the other hand, by Theorem 4.22, sign-rank$(A_f) = \exp(\Omega(t))$. Characterization (5.2) of UPP communication complexity implies that f requires at least $\Omega(t)$ bits of communication in the Paturi–Simon model. Hence, $f \notin \mathrm{UPP}$. Since UPP is closed under complement, the claim follows. □

6

Graph Complexity and Projective and Affine Dimensions of Graphs

Just as we associate to a Boolean function $f : \{0,1\}^\ell \times \{0,1\}^\ell \to \{0,1\}$ a $2^\ell \times 2^\ell$ 0–1 matrix M_f in the previous sections, we can also naturally associate to f a bipartite graph G_f and try to relate properties of G_f to the complexity of f. One can also think of models of computation on graphs and relate complexity of graphs in those models to the complexity of the corresponding Boolean functions. This approach was initiated by Pudlák et al. [79]. These models generalize well-known models of Boolean complexity such as circuits, branching programs, and two-party communication complexity. Measures of graph complexity such as affine dimension and projective dimension of graphs have been proposed and studied in [76, 78, 85] as criteria for lower bounds on formula size and branching program size of Boolean functions. Separation questions about classes of two-party communication complexity [5, 104] can be re-formulated as lower bound questions about bipartite graph complexity as noted in [79]. In this section, we introduce graph complexity and motivate lower bounds on affine and projective dimensions of graphs by connecting them to lower bounds on Boolean functions. We present some reductions that connect graph complexity to linear algebraic parameters and prove lower bounds on depth-3 graph formula size. Some of the techniques in this section are used in the next section on span programs.

6.1 Graph Complexity

The complexity of a graph G measures the difficulty of constructing G using a given collection of primitive graphs, called *generators*, and a given basis of operations on sets of edges. All the graphs involved are assumed to have the same set of vertices, typically $X = \{1,\ldots,n\}$. A set operation on graphs refers to the operation on the corresponding edge sets. For instance, the result of $G_1 \cup G_2$ on graphs $G_1 = (V,E_1)$ and $G_2 = (V,E_2)$ is the graph $G = (V,E_1 \cup E_2)$. Models of graph complexity are defined analogous to the standard models of circuits and formulas where the generator graphs play the role of input variables and the set operations play the role of gates.

As usual, we can imagine a circuit to be a directed acyclic graph (DAG) with the input nodes (of in-degree 0) labeled by the generator graphs and the internal nodes (gates) labeled by set operations from the basis. The target graph appears at the output gate (of out-degree 0) of this DAG. The *depth of the circuit* is defined to be the length of a longest path in the DAG. Its *size* is the number of nodes in it.

A *graph formula* is a graph circuit in which the out-degree of each gate is at most one. Thus, a graph formula can be represented as a tree with the leaves labeled by generator graphs and the internal nodes labeled by operations from the basis. The size of a formula is the number of leaves in its tree. The *formula complexity* of a graph is the smallest size of a formula that computes the graph (with respect to a fixed set of generators and a basis).

We can also define natural restricted models such as *constant depth* graph circuits and formulas. In these models, we allow unbounded fan-in and assume that the operations from the basis are naturally extendable to an unbounded number of operands (e.g., union, intersection, and symmetric difference of sets have this property). We can similarly generalize other models of Boolean complexity such as decision trees and branching programs to graph complexity.

We will consider graph complexity with the set operations of $\cup$ (UNION) and $\cap$ (INTERSECTION) only. We will naturally want the sets of generators to be *complete* in the sense that every graph should be constructible from these generators using $\cap$ and $\cup$ operators in a circuit or a formula.

We are especially interested in the complexity of bipartite graphs because of their direct relevance to lower bounds on Boolean circuits and communication complexity.

Definition 6.1. Fix the color classes X and Y. Let $\mathcal{B}$ denote the following set of complete bipartite graphs:

$$\mathcal{B} = \{A \times Y : A \subseteq X\} \ \cup \ \{X \times B : B \subseteq Y\}.$$

For a bipartite graph $G \subseteq X \times Y$, the *bipartite-formula complexity* of G is the smallest size of a graph formula computing G using $\mathcal{B}$ as the set of generators and $\cup$ and $\cap$ as the basis. Bipartite-formula complexity of G is denoted by $L_{\mathcal{B}}(G)$.

Definition 6.2. Let f be a Boolean function on 2ℓ variables, written as $f : \{0,1\}^\ell \times \{0,1\}^\ell \to \{0,1\}$. Let $n := 2^\ell$. The $n \times n$ bipartite graph $G_f \subseteq \{0,1\}^\ell \times \{0,1\}^\ell$ is defined by including the edge (x,y) in G_f iff $f(x,y) = 1$, where $x,y \in \{0,1\}^\ell$.

Note that the AND and OR operations on Boolean functions correspond to UNION and INTERSECTION operations on the edge sets of their corresponding graphs. In other words, $G_{f_1 \wedge f_2} = G_{f_1} \cap G_{f_2}$ and $G_{f_1 \vee f_2} = G_{f_1} \cup G_{f_2}$. This suggests a syntactic transformation of a Boolean formula (assuming all negations are pushed to the leaves) into a graph formula. What about the input literals of the Boolean formula? The literals are simply the projection functions and the graphs corresponding to projection functions are complete bipartite graphs isomorphic to $K_{n/2,n}$ and $K_{n,n/2}$. For instance, G_{x_i} is the complete bipartite graph $\{x \in \{0,1\}^\ell : x_i = 1\} \times \{0,1\}^\ell$. Thus each literal can be translated into a generator in $\mathcal{B}$. With this transformation of a Boolean formula for f into a bipartite-formula for G_f, it follows that

$$L_{\mathcal{B}}(G_f) \leq L(f), \tag{6.1}$$

where $L(f)$ is the minimum size of a formula (with tight negations) computing f.

Given an $n \times n$ bipartite graph G, where $n = 2^{\ell}$, we can clearly define a function f on 2ℓ variables such that $G_f = G$. Thus, we get the following criterion for lower bounds on Boolean formula size, proved in [79].

Lemma 6.1. A lower bound of $L_{\mathcal{B}}(G) \geq \psi(\log n)$ for an explicit $n \times n$ bipartite graph G, where $n = 2^{\ell}$, would yield an explicit function f on 2ℓ variables with formula size lower bound $L(f) \geq \psi(\ell)$.

Since the proof of this proposition is essentially syntactic, similar relations hold for other models such as circuits, decision trees, and branching programs as well.

Note, however, that graph complexity of G_f could be much smaller than the Boolean complexity of f. This is because in a bipartite-formula we have access to an exponential (in n) number of generators $\mathcal{B}$, whereas the transformation above uses only the $4\log n$ "canonical" generators corresponding to the projection functions. In fact, the generators in $\mathcal{B}$, with $X = Y = \{0,1\}^{\ell}$, can be viewed as defining *arbitrary* Boolean functions of either the first ℓ or the last ℓ variables. This interpretation captures the connection between two-party communication complexity and graph complexity.

The connections below indicate the generality and usefulness of graph complexity. In the following remarks, we assume the generators are from $\mathcal{B}$:

- When the model of computation is a decision tree, we get the model of Yao's two-party communication complexity [104].
- When the model is a Boolean formula of constant depth (polylog(ℓ) depth) and quasi-poly(ℓ) size, we get the definitions (cf. Definitions 5.6) of "polynomial hierarchy" $\mathrm{PH}^{\mathrm{cc}}$ ($\mathrm{PSPACE}^{\mathrm{cc}}$, respectively) in the two-party communication complexity model as defined in [5].

- In [76], graph complexity in the model of branching programs was used to derive criteria for the branching program size of the corresponding Boolean function. In particular, they define the notion of projective dimension of graphs and show that lower bounds on projective dimension of graphs imply lower bounds on branching program size of Boolean functions. We will see this notion in more detail in the next subsection.
- Formula complexity of graphs was used by Razborov [85] to derive criteria for lower bounds on formula size of the corresponding Boolean function. He defines the notion of affine dimension of graphs and shows that lower bounds on affine dimension of graphs imply lower bounds on formula size of Boolean functions. He also shows the relations between affine and projective dimensions of graphs in various cases. We will review these results as well in the next subsection.
- Using the translation of graph complexity to Boolean function complexity as an intermediate step, Lokam [58] shows that sufficiently strong lower bounds on the *monotone* complexity of the very special class of *2-slice functions* imply lower bounds on the complexity over a complete basis of certain Boolean functions.

6.2 Projective and Affine Dimensions of Graphs

Although the definitions of projective and affine dimensions make sense for any graph, we restrict our attention here to bipartite graphs only because of the motivating applications to Boolean function complexity. It is easy to generalize the following for nonbipartite graphs.

Definition 6.3. Let $G \subseteq X \times Y$ be a bipartite graph and V be a vector space of dimension d over a field $\mathbb{F}$.

- A *projective representation* of G in V associates a subspace U_z with every vertex $z \in X \cup Y$ such that $(x,y) \in G$ if and only if $U_x \cap U_y \neq \{0\}$. The *projective dimension* of G over $\mathbb{F}$ is defined to be the minimum d such that G has a d-dimensional representation and is denoted $\mathrm{pdim}_{\mathbb{F}}(G)$.

- The affine dimension of G $\mathrm{adim}_{\mathbb{F}}(G)$ is defined similarly using affine subspaces U_x, where $(x, y) \in G$ if and only if $U_x \cap U_y \neq \emptyset$.

Remark 6.1. Note that in the above definitions, we do not care about intersections of subspaces among vertices within the same color class X or Y.

Rónyai et al. [89] prove the following upper bounds on the projective dimension of any (even nonbipartite) graph.

Theorem 6.2. For any graph $G = (X, E)$,

- $\mathrm{pdim}_{\mathbb{F}}(G) \leq |E|$ over any field $\mathbb{F}$.
- $\mathrm{pdim}_{\mathbb{F}}(G) \leq 2\Delta$ if $|\mathbb{F}| \geq |E|$, where Δ is the maximum degree of a vertex in G.

Proof. Let $m := |E|$. For the first part, assign to each edge $e \in E$ a vector v_e in a basis $\{v_e \ : \ e \in E\}$ of $\mathbb{F}^m$. To each vertex $x \in E$ assign the subspace $U_x := \mathrm{span}\{v_e \ : \ e$ contains/incident with $x\}$. For the second part, pick m vectors $\{v_e : e \in E\}$ in general position in the space $\mathbb{F}^{2\Delta}$, e.g., $v_e = (1, \alpha_e, \alpha_e^2, \ldots, \alpha_e^{2\Delta-1})$, where α_e are distinct elements of $\mathbb{F}$. Let U_x be as above. □

The main motivation for studying the affine dimension of graphs comes from the following theorem [76].

Theorem 6.3. Let $f : \{0,1\}^\ell \times \{0,1\}^\ell \to \{0,1\}$ be a Boolean function and G_f be the associated $n \times n$ bipartite graph, where $n = 2^\ell$ as in Definition 6.2. Suppose f is computed by a branching program of size β. Then, for any field $\mathbb{F}$, $\mathrm{pdim}_{\mathbb{F}}(G_f) \leq \beta + 2$.

The following theorem about $\mathrm{pdim}_{\mathbb{F}}(G)$ for almost all bipartite graphs implies that most Boolean functions require branching programs of size $\Omega(2^{\ell/2})$. This leads to the main challenge about $\mathrm{pdim}_{\mathbb{F}}$: prove lower bounds, even of the form $\omega(\log n)$, for explicit bipartite graphs.

Theorem 6.4. For almost all $n \times n$ bipartite graphs G

$$\mathrm{pdim}_{\mathbb{F}}(G) = \begin{cases} \Omega(\sqrt{n}), & \text{if } \mathbb{F} \text{ is finite,} \\ \Omega\big(\sqrt{n/\log n}\big), & \text{if } \mathbb{F} \text{ is infinite.} \end{cases}$$

For finite $\mathbb{F}$, the proof follows from a simple counting argument. For infinite (or sufficiently large) $\mathbb{F}$, the proof [89] (generalizing the proof in [76] for $\mathbb{F} = \mathbb{R}$) uses upper bounds on the number of *zero patterns* of a sequence of polynomials.

The following lemma is proved in [76] (and generalized in [89]) to prove the second part of the above theorem and is often useful.

Lemma 6.5. Suppose, $\mathrm{pdim}_{\mathbb{F}}(G) \leq d$ and that $\mathbb{F}$ is sufficiently large. Then, there is a projective representation of G in $\mathbb{F}^{2d}$ in which, for every vertex $x \in X$, the associated subspace U_x is of dimension exactly d.

The only explicit graph we know of for which a lower bound on the projective dimension is known is the following.

Lemma 6.6. Let $G \subseteq X \times Y$ be the complement of a matching, e.g., $X = Y = \{1,2,\ldots,n\}$ and $(x,y) \in G$ if and only if $x \neq y$. Then, $\mathrm{pdim}_{\mathbb{F}}(G) = \Omega(\log n)$ for any sufficiently large field $\mathbb{F}$.

Proof. Suppose $\mathrm{pdim}_{\mathbb{F}}(G) = d$. By Lemma 6.5, we can assume that G has a projective representation in $V := \mathbb{F}^{2d}$ such that for each $x \in X$ ($y \in Y$) $\dim U_x = d$ ($\dim U_y = d$). We will consider the dth exterior power $W := \bigwedge^d V$ of V. Corresponding to a d-dimensional subspace $U \subseteq V$, we have a vector $u \in W$ defined by the wedge product $u = u_1 \wedge \cdots \wedge u_d$, where $u_1, \ldots, u_d$ is some basis of U. Thus, we have vectors u_x and u_y for each $x \in X$, $y \in Y$. Recall that a wedge product $w = w_1 \wedge \cdots \wedge$

$w_k = 0$ (the zero vector in $\bigwedge^k V$) if and only if the vectors $w_1, \ldots, w_k$ are linearly dependent in V. It follows that $u_x \wedge u_y = 0$ in $\bigwedge^{2d} V$ if and only if $(x,y) \in G$. By the proof of Theorem 6.12 [4, pp. 132–133], we conclude that $\{u_x : x \in X\}$ is a linearly independent set of vectors in W and hence $n \leq \dim W = \binom{2d}{d}$. Hence, $d = \Omega(\log n)$. □

Next, we consider the motivation for the affine dimension of graphs. Razborov [85] shows that for any bipartite graph G, $L_{\mathcal{B}}(G) \geq \mathrm{adim}_{\mathbb{F}}(G)$ (Theorem 6.12 below).

Before, we prove this theorem, we review some notions about *rectangle covers*. This machinery is again used in Section 7.

6.2.1 Rectangle Covers

Let P and Q be (for now) arbitrary finite sets. A *rectangle* in $P \times Q$ is a set $S \times T$, where $S \subseteq P$ and $T \subseteq Q$. A set of rectangles $\mathcal{C}$ is called a *cover* (or a rectangle cover) of $P \times Q$ if $\cup_{R \in \mathcal{C}} R = P \times Q$. A rectangle cover is a *disjoint cover* if the rectangles in it are mutually disjoint. A set of rectangles $\mathcal{C}'$ is *embedded* in the set of rectangles $\mathcal{C}$ if every $R' \in \mathcal{C}'$ is contained in some $R \in \mathcal{C}$. For a rectangle cover $\mathcal{C}$ of $P \times Q$, define

$$\alpha(\mathcal{C}) := \min\{|\mathcal{C}'| \,:\, \mathcal{C}' \text{ is a disjoint cover embedded in } \mathcal{C}\}.$$

Finding lower bounds on $\alpha(\mathcal{C})$ is reduced next to finding lower bounds on certain rank-like functions. Let A be a matrix over $P \times Q$ (i.e., rows are indexed by elements of P and columns by elements of Q). For a rectangle $R \subseteq P \times Q$, let A_R denote the matrix with entries in R equal to those of A and zero outside of R.

Lemma 6.7. For a rectangle cover $\mathcal{C}$ of $P \times Q$ and any matrix A (over any field $\mathbb{F}$) over $P \times Q$,

$$\alpha(\mathcal{C}) \geq \frac{\mathrm{rank}(A)}{\max_{R \in \mathcal{C}} \mathrm{rank}(A_R)}.$$

In particular, if A is such that A_R is monochromatic (i.e., all nonzero entries of A_R are equal) for each $R \in \mathcal{C}$, then $\alpha(\mathcal{C}) \geq \mathrm{rank}(A)$.

Proof. Let $\mathcal{C}'$ be a disjoint covering embedded in $\mathcal{C}$ such that $\alpha(\mathcal{C}) = |\mathcal{C}'|$. Note that for any $R' \in \mathcal{C}'$, there is an $R \in \mathcal{C}$ such that $R' \subseteq R$. Thus, $\operatorname{rank}(A_{R'}) \leq \operatorname{rank}(A_R)$. Now, $A = \sum_{R' \in \mathcal{C}'} A_{R'}$ and hence

$$\begin{aligned} \operatorname{rank}(A) &\leq \sum_{R' \in \mathcal{C}'} \operatorname{rank}(A_{R'}) \\ &\leq |\mathcal{C}'| \max_{R' \in \mathcal{C}'} \operatorname{rank}(A_{R'}) \leq \alpha(\mathcal{C}) \max_{R \in \mathcal{C}} \operatorname{rank}(A_R). \qquad \square \end{aligned}$$

Here, another rank-like function is useful in proving lower bounds on $\alpha(\mathcal{C})$. In a *partial matrix* A over a field $\mathbb{F}$, some entries may be left unspecified (denoted by $*$). The rank of a partial matrix is defined to be the minimum rank of a full matrix obtained from A, over all possible ways of filling the unspecified entries from $\mathbb{F}$. Suppose, $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$. Let $a_1 \neq a_2$ be two nonzero entries over $\mathbb{F}$. An element $(p,q) \in P \times Q$ is covered by $\mathcal{C}_i$ if it is contained in some rectangle $R' \in \mathcal{C}_i$. Define the partial matrix A by

$$A_{pq} = \begin{cases} a_1, & \text{if } (p,q) \text{ is } \textit{not} \text{ covered by } \mathcal{C}_1, \\ a_2, & \text{if } (p,q) \text{ is } \textit{not} \text{ covered by } \mathcal{C}_2, \\ *, & \text{otherwise.} \end{cases}$$

Lemma 6.8. For any cover $\mathcal{C}$ of $P \times Q$ such that $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$ and any partial matrix A defined as above, $\alpha(\mathcal{C}) \geq \operatorname{rank}(A)$.

Proof. We will define a matrix B such that $\operatorname{rank}(B) = \operatorname{rank}(A)$ for the partial matrix A. Let $\mathcal{C}'$ be a disjoint covering embedded in $\mathcal{C}$ such that $\alpha(\mathcal{C}) = |\mathcal{C}'|$. Partition $\mathcal{C}'$ as $\mathcal{C}' = \mathcal{C}'_1 \cup \mathcal{C}'_2$ ($\mathcal{C}'_1 \cap \mathcal{C}'_2 = \emptyset$) where[1] $\operatorname{supp}(\mathcal{C}'_1) \subseteq \operatorname{supp}(\mathcal{C}_1)$ and $\operatorname{supp}(\mathcal{C}'_2) \subseteq \operatorname{supp}(\mathcal{C}_2)$. Let

$$B := a_2 \sum_{R' \in \mathcal{C}'_1} J_{R'} + a_1 \sum_{R' \in \mathcal{C}'_2} J_{R'},$$

where $J_{R'}$ is a matrix with all 1's inside R' and 0's elsewhere. Clearly, B extends A and hence $\operatorname{rank}(A) \leq \operatorname{rank}(B)$. Moreover $\operatorname{rank}(B) \leq |\mathcal{C}'_1| + |\mathcal{C}'_2| \leq |\mathcal{C}'| = \alpha(\mathcal{C})$. $\square$

[1] Define $\operatorname{supp}(\mathcal{C})$ to be the union of all the rectangles in $\mathcal{C}$.

Let $f:\{0,1\}^t \to \{0,1\}$ be a Boolean function and let, now, $P \subseteq f^{-1}(1)$ and $Q \subseteq f^{-1}(0)$. Then, there is a *canonical cover* of $P \times Q$ consisting of the $2t$ rectangles

$$R_{i\epsilon} := \{(x,y) \in P \times Q : x_i = \epsilon, y_i = 1-\epsilon\},$$

for $1 \leq i \leq t$ and $\epsilon \in \{0,1\}$. This canonical cover is denoted by $\mathcal{C}_{\text{can}}(P,Q)$. If $P = f^{-1}(1)$ and $Q = f^{-1}(0)$, it is denoted by $\mathcal{C}_{\text{can}}(f)$. If f is a monotone Boolean function, then for any x with $f(x) = 1$ and y with $f(y) = 0$, we can find an index i such that $x_i = 1$ and $y_i = 0$. Hence, for monotone f, it suffices to consider the t canonical rectangles with $\epsilon = 1$. Such monotone canonical covers are denoted $\mathcal{C}_{\text{mon}}(P,Q)$ and $\mathcal{C}_{\text{mon}}(f)$ analogous to the above notation.

The fundamental application of rectangle covers to lower bounds is given by the following result from [85].

Theorem 6.9. For any Boolean function $f:\{0,1\}^t \to \{0,1\}$ and any $P \subseteq f^{-1}(1)$, $Q \subseteq f^{-1}(0)$, the formula size complexity $L(f)$ (in the AND, OR, NOT basis) of f is lower bounded as follows

$$L(f) \geq \alpha(\mathcal{C}_{\text{can}}(f)) \geq \alpha(\mathcal{C}_{\text{can}}(P,Q)).$$

For monotone f, the monotone formula complexity $L_{mon}(f)$ in the AND-OR basis is lower bounded as

$$L_{\text{mon}}(f) \geq \alpha(\mathcal{C}_{\text{mon}}(f)) \geq \alpha(\mathcal{C}_{\text{mon}}(P,Q)).$$

We remark that the basic intuition for this connection goes back to Khrapchenko and Rychkov (see, [85]). A similar approach was taken by Karchmer and Wigderson [40] in their characterization of circuit depth in terms of communication complexity (recall that circuit depth is logarithmic in formula size).

The ideas in the proof of Theorem 6.9 generalize to the model of graph complexity. For this, let $G \subseteq X \times Y$ be a bipartite graph. Recall that $L_{\mathcal{B}}(G)$ denotes the *graph-formula complexity* of G using operations $\cup$ and $\cap$ and complete bipartite graphs ($\mathcal{B}$) as generators. Let us look at the framework of rectangle covers in this context. To begin with, let

$P \subseteq G$ be a subset of edges of G and let $Q \subseteq (X \times Y) \backslash G$ be a subset of nonedges of G. We can then consider the rectangle covers of $P \times Q$ (or more generally those of $P' \times Q'$ for any disjoint $P', Q' \subseteq X \times Y$). For any $X_0 \subseteq X$, let $B(X_0)$ denote the complete bipartite graph $X_0 \times Y$; similarly for any $Y_0 \subseteq Y$. The *canonical cover* of $P \times Q$, denoted $\mathcal{C}_{\mathrm{gr}}(P,Q)$, is defined by the sets of rectangles:

$$\begin{aligned}
\mathcal{C}_{\mathrm{gr},X} &:= \{P \cap B(X_0) \times Q \cap B(\overline{X_0}) \ : \ X_0 \subseteq X\}, \\
\mathcal{C}_{\mathrm{gr},Y} &:= \{P \cap B(Y_0) \times Q \cap B(\overline{Y_0}) \ : \ Y_0 \subseteq Y\}.
\end{aligned}$$

When $P = G$ and $Q = (X \times Y) \backslash G$, we denote this canonical cover by $\mathcal{C}_{\mathrm{gr}}(G)$.

The generalization of Theorem 6.9 to graph complexity is as follows.

Theorem 6.10. For any bipartite graph $G \subseteq X \times Y$ and $P \subseteq G$ and $Q \subseteq (X \times Y) \backslash G$, $L_{\mathcal{B}}(G) \geq \alpha(\mathcal{C}_{\mathrm{gr}}(P,Q))$.

Consider the partial matrix $A := A(\mathbb{F}, a_1, a_2, G)$ over $\mathbb{F}$, where $a_1 \neq a_2$ are nonzero elements of $\mathbb{F}$ for the cover $\mathcal{C}_{\mathrm{gr}}(G)$.

Given an edge $p \in G$ and a non-edge $q \in (X \times Y) \setminus G$, we have

$$A_{pq} = \begin{cases} a_1 & \text{if } p \text{ and } q \text{ share a vertex in } X, \\ a_2 & \text{if } p \text{ and } q \text{ share a vertex in } Y, \\ * & \text{otherwise.} \end{cases}$$

(Note that an (edge, nonedge) pair covered by a rectangle in $\mathcal{C}_{\mathrm{gr},X}$ cannot share a common vertex in X and similarly $\mathcal{C}_{\mathrm{gr},Y}$.)

Using Lemma 6.8 for the cover $\mathcal{C}_{\mathrm{gr}}(G)$ with this partial matrix and Theorem 6.10, we get the following.

Lemma 6.11. For any bipartite graph $G \subseteq X \times Y$, $L_{\mathcal{B}}(G) \geq \mathrm{rank}(A(\mathbb{F}, a_1, a_2, G))$.

6.2.2 Affine Dimension

We can now return to the main motivation for studying the affine dimension of graphs [85].

Theorem 6.12. For any graph $G \subseteq X \times Y$ and any field $\mathbb{F}$, $L_{\mathcal{B}}(G) \geq \mathrm{adim}_{\mathbb{F}}(G)$.

Proof. We will give an affine representation for G using Lemma 6.11 in dimension $d = \mathrm{rank}(A(\mathbb{F}, a_1, a_2, G))$. Let B be a full matrix of rank d over $\mathbb{F}$ such that (usual) $\mathrm{rank}(B) = \mathrm{rank}(A)$. Let the row space of B be the ambient space ($\cong \mathbb{F}^d$). Recall that the rows of B are indexed by edges of G. With each edge $e \in G$, associate the corresponding row vector v_e. Now, represent each vertex $z \in X \cup Y$ by affine subspace U_z given by affine span of edges incident to z:

$$U_z := \mathrm{affspan}\langle v_e \ : \ e \text{ is incident to } z \rangle.$$

Clearly, if x and y are adjacent in G, the affine subspaces U_x and U_y have the vector v_{xy} in them. On the other hand, suppose x and y are not adjacent. For any edge e incident to x, e and xy share the vertex $x \in X$. Hence, for every such edge the corresponding row has a_1 in the column corresponding to the nonedge xy. Thus, all affine linear combinations of these edges will also have a_1 in that coordinate, i.e., every vector in U_x has a_1 in the coordinate indexed by the column xy. Similarly, every vector in the affine subspace U_y has a_2 in that coordinate since the nonedge xy and all edges incident to y share a vertex in Y. This show that $U_x \cap U_y = \emptyset$. □

The following theorem from [85] states the known the relations between $\mathrm{adim}_{\mathbb{F}}$ and $\mathrm{pdim}_{\mathbb{F}}$.

Theorem 6.13. For any graph G,

- over any field $\mathbb{F}$, $\mathrm{adim}_{\mathbb{F}}(G) \leq (\mathrm{pdim}_{\mathbb{F}}(G))^2$.
- over any *finite* field $\mathbb{F}$, $\mathrm{pdim}_{\mathbb{F}}(G) \leq (\mathrm{adim}_{\mathbb{F}}(G))^{O(\mathrm{adim}_{\mathbb{F}}(G))}$.

It is easy to see that, when G is the complement of a matching, $\mathrm{adim}_{\mathbb{R}}(G) = 2$. Combined with Lemma 6.6, this shows that no general relation between $\mathrm{adim}_{\mathbb{R}}$ and $\mathrm{pdim}_{\mathbb{R}}$ can exist.

6.3 Lower Bounds on Depth-3 Graph Complexity

In this subsection, we consider a *restricted model* of graph complexity, namely, depth-3 formulas. We will derive a nontrivial lower bound of $\Omega(\log^3 n/(\log\log n)^5)$ on the depth-3 complexity of Paley-type bipartite graphs [58]. The proof uses the lower bound on ℓ_2-rigidity (Lemma 3.5) and approximating polynomial representations of the OR function. Improving this to $n^{\Omega(1)}$ is a big challenge — such a bound would give superlinear lower bounds on *log-depth* Boolean circuits and a language outside the class Σ_2^{cc} in two-party communication complexity — these are two long-standing open questions.

Suppose a depth-3 bipartite-formula computes a bipartite graph $G \subseteq U \times V$, $|U| = |V| = n$. Recall that the leaves of the formula are graphs from $\mathcal{B} = \{A \times V : A \subseteq U\} \cup \{U \times B : B \subseteq V\}$. Let us first observe that the bottom gates of a bipartite-formula need not have fan-in more than 2. Indeed, an $\cap$ gate at the bottom computes a complete bipartite graph $A \times B$ and a $\cup$ bottom gate computes the complement of a complete bipartite graph $\overline{A \times B}$, where $A \subseteq U$ and $B \subseteq V$. These can be written as intersection and union, respectively, of at most two graphs from $\mathcal{B}$.

Without loss of generality, we consider $\cup\cap\cup$ formulas. By the remark above, we can write such a formula as $G = \cup_i \cap_j G_{ij}$, where G_{ij} is the complement of a complete bipartite graph, i.e., $\overline{A_{ij} \times B_{ij}}$ for some $A_{ij} \subseteq U$ and $B_{ij} \subseteq V$.

One ingredient of our proof is sign-representing polynomials for DNF's. Nisan and Szegedy [66] give the following construction of ϵ-approximating polynomials for the OR-function. They assume a constant ϵ. The refined analysis to bring out the dependence on ϵ is due to Hayes and Kutin (private communication). For a proof, see [58].

Lemma 6.14. The OR-function of n Boolean variables can be ϵ-approximated by a real polynomial of degree at most $O(\sqrt{n}\log(2/\epsilon))$. More precisely, for every $0 < \epsilon < 1/2$, there is a real polynomial p of degree at most $O(\sqrt{n}\log(2/\epsilon))$ such that for every $x \in \{0,1\}^n$, $|\mathrm{OR}(x) - p(x)| \leq \epsilon$.

For a bipartite graph $G \subseteq U \times V$, we will let $G(x,y) = 1$ if $(x,y) \in G$ and $G(x,y) = 0$ if $(x,y) \notin G$.

Lemma 6.15. Suppose an $n \times n$ bipartite graph $H \subseteq U \times V$ is written as a union of d complete bipartite graphs:

$$H = \bigcup_{i=1}^{d} (A_i \times B_i), \quad \text{where } A_i \subseteq U,\ B_i \subseteq V.$$

Then, for every ϵ, where $0 < \epsilon < 1/2$, there is a real matrix M_H such that

- For all $(x,y) \in U \times V$, $|M_H(x,y) - H(x,y)| \leq \epsilon$,
- $\mathrm{rank}(M_H) \leq \exp(O(\sqrt{d}\log(2/\epsilon)\log d))$.

Proof. Let R be the incidence matrix of H, and similarly let R_i be the incidence matrices of the complete bipartite graphs $A_i \times B_i$, $1 \leq i \leq d$, covering H. Note that R is simply the entry-wise OR of the R_i. Furthermore, each R_i is of rank one as a real matrix. We obtain M_H from R using the approximating polynomials for the OR-function given by Lemma 6.14.

Suppose $p(z_1,\ldots,z_d)$ is an ϵ-approximating polynomial of degree $k := c \cdot \sqrt{d}\log(2/\epsilon)$ for $\mathrm{OR}(z_1,\ldots,z_d)$. Syntactically substitute the matrix R_i for z_i in this polynomial, but interpret the product as entry-wise product of matrices, i.e., a monomial $z_i z_j$ is replaced by $R_i \circ R_j$, where for matrices A and B, $(A \circ B)(x,y) := A(x,y)B(x,y)$. Note that if A and B are rank-1 matrices, then $A \circ B$ is also a rank-1 matrix. Thus, a monomial $z_{i_1} \cdots z_{i_t}$ is replaced by the rank-1 0–1 matrix $R_{i_1} \circ \cdots \circ R_{i_t}$. The matrix obtained by computing the polynomial $p(R_1,\ldots,R_d)$ in this way gives us the desired matrix M_H.

It is clear that $M_H(x,y) = p(R_1(x,y),\ldots,R_d(x,y))$. From the properties of p, it is easy to see that for all x,y, $|M_H(x,y) - H(x,y)| \leq \epsilon$. Since M_H is a linear combination of rank-1 matrices, one for each monomial, it follows that rank of M_H is at most the number of monomials in p which is bounded by $\sum_{j=0}^{k} \binom{d}{j} \leq \exp(O(k\log d))$. □

Lemma 6.16. Let $G \subseteq U \times V$ be a bipartite graph. If G is realized by a depth-3 bipartite-formula:

$$G = \bigcup_{i=1}^{t} \bigcap_{j=1}^{d_i} \overline{(A_{ij} \times B_{ij})}, \quad \text{where } A_{ij} \subseteq U,\ B_{ij} \subseteq V,$$

then there exists a matrix M such that

(1) If $G(x,y) = 0$, then $M(x,y) \leq -1/6$.
(2) If $G(x,y) = 1$, then $M(x,y) \geq +1/6$.
(3) $\operatorname{rank}(M) \leq \exp(O(\sqrt{D} \log t \log D))$, where $D = \max_{i=1}^{t} d_i$.

Proof. Let $G_1,\ldots,G_t$ be the input graphs to the top gate so that $G = \cup_{i=1}^{t} G_i$. Since each G_i is an intersection of complements of complete bipartite graphs, its complement, $H_i := \overline{G_i}$ is computed by a union of complete bipartite graphs. Thus, we can apply Lemma 6.15 to the H_i. Let M_i be the real matrix given by Lemma 6.15 that ϵ_i-approximates H_i, where $\epsilon_i := 1/3t$. Since, as a matrix, $G_i = J - H_i$, where J is the $n \times n$ all 1's matrix, the matrix $N_i := J - M_i$ is an ϵ_i-approximation of G_i. Furthermore, since $\operatorname{rank}(N_i) \leq \operatorname{rank}(M_i) + 1$, it follows from Lemma 6.15 that $\operatorname{rank}(N_i) \leq \exp(O(\sqrt{d_i} \log(1/\epsilon_i) \log d_i)) \leq \exp(O(\sqrt{D} \log t \log D))$.

Let $M := N_1 + \cdots + N_t - \frac{1}{2} \cdot J$. Let us see the relation between M and G:

(1) If $G(x,y) = 0$, then $\forall i\ G_i(x,y) = 0$, and hence $\forall i |N_i(x,y)| \leq \epsilon_i$. It follows that $M(x,y) \leq \sum_{i=1}^{t} \epsilon_i - 1/2 \leq -1/6$.
(2) If $G(x,y) = 1$, then $\exists i\ G_i(x,y) = 1$ and for this i, $1 - \epsilon_i \leq N_i(x,y) \leq 1 + \epsilon_i$. For $j \neq i$, $-\epsilon_j \leq N_j(x,y) \leq 1 + \epsilon_j$. Hence, we have $1 - \epsilon_i - \sum_{j \neq i} \epsilon_j - 1/2 \leq M(x,y) \leq \sum_{j=1}^{t} (1 + \epsilon_j) - 1/2$. So, in this case, $2/3 - 1/2 \leq M(x,y) \leq t + 1/3 - 1/2$, and hence $M(x,y) \geq 1/6$.
(3) Moreover, $\operatorname{rank}(M) \leq \sum_{i=1}^{t} \operatorname{rank}(N_i) + 1 \leq t \exp(O(\sqrt{D} \log D \log t)) + 1 \leq \exp(O(\sqrt{D} \log D \log t))$. □

Lemma 6.17. Let G be an $n \times n$ bipartite graph $G \subseteq U \times V$. If G is realized by a depth-3 bipartite-formula:

$$G = \bigcup_{i=1}^{t} \bigcap_{j=1}^{d_i} \overline{(A_{ij} \times B_{ij})}, \quad \text{where } A_{ij} \subseteq U,\ B_{ij} \subseteq V,$$

then there exists a matrix M such that

(1) If $G(x,y) = 0$, then $M(x,y) \geq 1$.
(2) If $G(x,y) = 1$, then $M(x,y) = 0$.
(3) $\text{rank}(M) \leq \prod_{i=1}^{t} d_i$.

Proof. It will be convenient to consider the complement graph $\overline{G}$ of G. Clearly,

$$\overline{G} = \bigcap_{i=1}^{t} \underbrace{\bigcup_{j=1}^{d_i} \underbrace{(A_{ij} \times B_{ij})}_{G_{ij}}}_{G_i}.$$

Let R_{ij} be the incidence matrix of the complete bipartite graph G_{ij}. Define $M_i = \sum_{j=1}^{d_i} R_{ij}$. Finally, let $M = M_1 \circ M_2 \circ \cdots \circ M_t$. Hence, we have

$$M(x,y) = \prod_{i=1}^{t} \sum_{j=1}^{d_i} R_{ij}(x,y).$$

Note that if $(x,y) \in G_i$ then $M_i(x,y) = \sum_{j=1}^{d_i} R_{ij}(x,y) \geq 1$ and if $(x,y) \notin G_i$, then $M_i(x,y) = 0$. Hence, we have

$$(x,y) \in \overline{G} \Rightarrow M(x,y) \geq 1,$$
$$(x,y) \notin \overline{G} \Rightarrow M(x,y) = 0.$$

From this (1) and (2) follow.

To see the bound on rank(M), note that rank is sub-multiplicative under $\circ$: $\text{rank}(A \circ B) \leq \text{rank}(A) \cdot \text{rank}(B)$. Since $\text{rank}(R_{ij}) = 1$, we have $\text{rank}(M_i) \leq d_i$. Hence, $\text{rank}(M) \leq \prod_{i=1}^{t} \text{rank}(M_i) \leq \prod_{i=1}^{t} d_i$. This gives (3). □

Theorem 6.18. Let G be an $n \times n$ bipartite graph $G \subseteq U \times V$. If G is realized by a depth-3 bipartite-formula:

$$G = \bigcup_{i=1}^{t} \bigcap_{j=1}^{d_i} \overline{(A_{ij} \times B_{ij})}, \quad \text{where } A_{ij} \subseteq U,\ B_{ij} \subseteq V,$$

then there exists a matrix M such that

(1) If $G(x,y) = 0$, then $M(x,y) \leq -1/12$.
(2) If $G(x,y) = 1$, then $M(x,y) \geq +1/12$.
(3) $\mathrm{rank}(M) \leq \exp(O(L^{1/3} \log^{5/3} L))$, where $L = \sum_{i=1}^{t} d_i$.

Proof. Let D^* be a parameter to be fixed later. We will separate the formula into two parts depending on whether the middle fan-in d_i is smaller or larger than D^*. This idea of splitting the formula based on middle fan-in was used earlier in [44].

Specifically, let G_s be the subgraph of G realized by the subformula given by union of middle gates of fan-in at most D^* and let G_l be the subgraph of G similarly given by middle gates of fan-in larger than D^*:

$$\begin{aligned} G_s &= \bigcup_{d_i \leq D^*} \bigcap_{j=1}^{d_i} \overline{(A_{ij} \times B_{ij})}, \\ G_l &= \bigcup_{d_i > D^*} \bigcap_{j=1}^{d_i} \overline{(A_{ij} \times B_{ij})}. \end{aligned}$$

We apply Lemma 6.16 to G_s to get a matrix M_s such that $M_s(x,y) \geq 1/6$ if $(x,y) \in G_s$ and $M_s(x,y) \leq -1/6$ if $(x,y) \notin G_s$. Furthermore, we have $\mathrm{rank}(M_s) \leq \exp(O(\sqrt{D^*} \log D^* \log t))$ since all middle fan-in's of the formula for G_s are at most D^* and the top fan-in is at most t.

We apply Lemma 6.17 to G_l and get a matrix M_l such that $M_l(x,y) \geq 1$ if $(x,y) \notin G_l$ and $M_l(x,y) = 0$ if $(x,y) \in G_l$. Since all middle fan-in's of the formula for G_l are at least D^*, the top fan-in t_l of this formula is at most $t_l \leq \sum_{i=1}^{t} d_i / D^* \leq L/D^*$. We bound rank of M_l

as follows:

$$\begin{aligned}
\mathrm{rank}(M_l) &\leq \prod_{d_i > D^*,\, 1 \leq i \leq t} d_i \\
&\leq \left(\sum_{d_i > D^*} \frac{d_i}{t_l} \right)^{t_l} \\
&\leq \exp(O(t_l \log D)), \quad \text{where } D = \sum_{d_i > D^*} \frac{d_i}{t_l} \leq L \\
&\leq \exp\left(O\left(\frac{L}{D^*} \log D \right) \right).
\end{aligned}$$

Define $M = M_s \circ M_l + (1/12)J$. If $(x,y) \in G_l$, then $M(x,y) = 1/12$, and if $(x,y) \in G_s \backslash G_l$, then $M(x,y) \geq 1/6 + 1/12$. If $(x,y) \notin G_s \cup G_l$, then $M(x,y) \leq -1/6 + 1/12 \leq -1/12$. Hence, (1) and (2) are satisfied by M.

By sub-multiplicativity of rank under $\circ$, we get the upper bound on rank of M:

$$\mathrm{rank}(M) \leq \exp\left(O(\sqrt{D^*} \log D^* \log t) + O\left(\frac{L}{D^*} \log D \right) \right) + 1.$$

We set $D^* = \Theta((L/\log L)^{2/3})$. Using the trivial upper bound of $O(\log L)$ on $\log D^*$, $\log D$, and $\log t$, we get that $\mathrm{rank}(M) = \exp(O(L^{1/3}(\log L)^{5/3}))$, verifying (3). □

We now use Forster's Theorem 4.2 to show that for some "interesting" graphs G any matrix satisfying (1) and (2) of Theorem 6.18 must have a large rank and hence conclude a lower bound on the depth-3 complexity of G using (3).

Theorem 6.19. Let G be an $n \times n$ bipartite graph and A_G be its ± 1 incidence matrix, i.e., $A_G(x,y) = 1$ if (x,y) is an edge of G and $A_G(x,y) = -1$ if (x,y) is not an edge of G. Then any depth-3 bipartite-formula for G must have size at least

$$\Omega\left(\frac{\log^3(n/\|A_G\|)}{\log\log^5(n/\|A_G\|)} \right).$$

Proof. Given a depth-3 formula for G of size L, let M be the matrix given by Theorem 6.18. Note that if $(x,y) \notin G$, then $M(x,y) \leq -1/12$, and if $(x,y) \in G$, then $M(x,y) \geq 1/12$. Hence, $12M$ is a sign-preserving variation of A_G and we can apply Theorem 4.2 to conclude that $\operatorname{rank}(M) = \operatorname{rank}(12M) = \Omega(n/\|A_G\|)$. On the other hand, from Theorem 6.18, (3), we get that $\operatorname{rank}(M) \leq \exp(O(L^{1/3}\log^{5/3} L))$. Combining the two estimates on $\operatorname{rank}(M)$, we get

$$\exp(O(L^{1/3}\log^{5/3} L)) = \frac{n}{\|A_G\|}.$$

Solving for L proves the theorem. □

Corollary 6.20. For any graph G such that A_G is an Hadamard matrix, the depth-3 bipartite-formula complexity of G is at least $\Omega((\log n)^3/(\log\log n)^5)$. An example of such a graph is the Paley-type bipartite graph.

7

Span Programs: A Linear Algebraic Model of Computation

The model of span programs was introduced by Karchmer and Wigderson [41]. A span program over a field $\mathbb{F}$ computes a Boolean function $f : \{0,1\}^n \to \{0,1\}$ as follows. With each of the $2n$ literals x_i ($\overline{x_i}$, respectively), a span program associates a subspace U_i^1 (U_i^0) of an ambient space V over $\mathbb{F}$. There is no condition on the dimensions of these subspaces. Fix a nonzero vector $z \in V$. Now, the condition is that

$$f(x) = 1 \quad \text{if and only if } z \in \operatorname{span}\{U_1^{x_1}, U_2^{x_2}, \ldots, U_n^{x_n}\},$$

i.e., an input assignment "picks" either U_i^1 or U_i^0 depending on whether it sets x_i to 1 or 0, respectively, and the function evaluates to 1 on that assignment if and only if the special vector z falls in the linear span of the picked subspaces. W.l.o.g, we let z to be the all 1's vector $\mathbf{1}$ in what follows. The *dimension* of the span program is the dimension of the ambient space V. Define now $\operatorname{sdim}_{\mathbb{F}}(f)$ to be the smallest dimension d of an ambient vector space V in which a span program as described above can be realized for computing the function f. The *size* of the span program is the sum of the dimensions of the $2n$ subspaces associated to the literals: $\sum_{i,\epsilon} \dim U_i^{\epsilon}$. The span program size $\mathrm{SP}_{\mathbb{F}}(f)$ of a function f is the minimum size of a span program computing the function f.

In matrix language, a span program for a Boolean function f is a matrix S whose rows span the ambient space V. Each row is labeled by a unique literal (the same vector could repeat as multiple rows). Each subspace U_i^ϵ is then specified by a subset of rows and the number of rows in S is the size of the span program. For a given input assignment $x \in \{0,1\}^n$, let S_x denote the submatrix given by the rows "activated" by x, i.e., for each i, take the set of rows corresponding to $U_i^{x_i}$. Then, $f(x) = 1$ if and only if $\mathbf{1}$ is in the row space of S_x, i.e., there exists a vector c_x such that $c_x^T S_x = \mathbf{1}^T$.

A *monotone* span program has subspaces associated only to the n positive literals x_i. The criterion for when a monotone span program computes a monotone Boolean function is defined similarly as above. Let $\mathrm{mSP}_{\mathbb{F}}(f)$ denote the size of a *monotone* span program that computes the monotone function f.

Using Warren's inequality [102], Babai et al. [6] prove that for *almost all* f, $\mathrm{sdim}_{\mathbb{F}}(f)$ is at least $\Omega(2^{n^{1/3}}/(n\log n)^{1/3})$. Proving superpolynomial lower bounds for explicit f is a major challenge. Span programs provide a "static" model of computation for a Boolean function and this model promises to be a vehicle for realizing the approach of the *fusion method* [103] for circuit lower bounds. Span programs are well-related to other models of complexity such as Boolean formulas, symmetric branching programs, algebraic proof systems, and secret sharing schemes.

A major achievement is a sequence of results [6, 9, 32] culminating in Gál's proof of an $n^{\Omega(\log n)}$ lower bound on *monotone* span program size of several explicit Boolean functions. These results are highly nontrivial applications of combinatorial and algebraic techniques. Gál [32] actually provides a *characterization* of span program size and uses a *gap* between rectangle cover numbers (cf. Section 6.2.1) and ranks to get the lower bound.

7.1 Characterizing Span Program Size

Recall the notion of a rectangle cover of a cartesian product $P \times Q$ from Section 6.2.1. Here, we will use a parameter that generalizes $\alpha(\mathcal{C})$.

Definition 7.1. Let a rectangle cover $\mathcal{C}$ of $P \times Q$ and a field $\mathbb{F}$ be given. A set $\mathcal{K}$ of rank-1 matrices (with entries indexed by $P \times Q$) over $\mathbb{F}$ is said to be *embedded* in $\mathcal{C}$ if for every $K \in \mathcal{K}$, the set of all nonzero entries of K are contained in some rectangle $R \in \mathcal{C}$. We then define

$$\alpha_{\mathbb{F}}(\mathcal{C}) := \min \left\{ |\mathcal{K}| \ : \ \sum_{K \in \mathcal{K}} K = J \text{ and } \mathcal{K} \text{ is a set of rank-1 matrices} \right. \\ \left. \text{over } \mathbb{F} \text{ embedded in } \mathcal{C} \right\}.$$

Here, J denotes the all 1's matrix over $\mathbb{F}$ indexed by $P \times Q$.

Clearly, $\alpha(\mathcal{C}) \geq \alpha_{\mathbb{F}}(\mathcal{C})$: for a disjoint rectangle cover $\mathcal{C}'$ embedded in $\mathcal{C}$, simply take for each $R \in \mathcal{C}'$ a rank-1 matrix whose entries take the value 1 inside R and 0 outside R. An analog of Theorem 6.9 characterizes the span program size completely.

Theorem 7.1. Fix a field $\mathbb{F}$. For every Boolean function f, $\mathrm{SP}_{\mathbb{F}}(f) = \alpha_{\mathbb{F}}(\mathcal{C}_{\mathrm{can}}(f))$. For every monotone Boolean function f, $\mathrm{mSP}_{\mathbb{F}}(f) = \alpha_{\mathbb{F}}(\mathcal{C}_{\mathrm{mon}}(f))$.

To prove this theorem, we make use of the following lemma.

Lemma 7.2. A Boolean function $f : \{0,1\}^n \to \{0,1\}$ has a span program of size s over $\mathbb{F}$ if and only if there exist sets of vectors

$$B = \left\{ b_x \in \mathbb{F}^s \ : \ x \in f^{-1}(0) \right\}, \quad \text{and} \quad C = \left\{ c_y \in \mathbb{F}^s \ : \ y \in f^{-1}(1) \right\},$$

such that

(1) The set $[s]$ of coordinates can be partitioned into subsets $s_{i\epsilon}$ corresponding to the $2n$ literals, $1 \leq i \leq n$, $\epsilon \in \{0,1\}$, such that the following holds: Every $b_x \in B$ can partitioned into $2n$ "segments" (or subvectors) $b_x^{i\epsilon}$, where $b_x^{i\epsilon}$ is the restriction of b_x to $s_{i\epsilon}$, such that the segments activated by x are identically zero vectors, i.e., $b_x^{ix_i} = \mathbf{0}$ for all $i, 1 \leq i \leq n$. Similarly, every

$c_y \in C$ can be partitioned into $c_y^{i\epsilon}$ such that, w.o.l.g., we can assume that the segments *not* activated by y are identically zero, i.e, $c_y^{i,1-y_i} = \mathbf{0}$ for all $i, 1 \leq i \leq n$.

(2) For every $b \in B$, $c \in C$, $c^T b = 1$.

Proof. The if-direction is trivial: let S be the matrix with vectors from B as columns. It is easy to verify that S is a valid span program for f and its size is clearly s. Consider the submatrix of S given by the segments $b_x^{i\epsilon}$ for all $x \in f^{-1}(0)$. The rows of this submatrix generate U_i^ϵ of dimension $|s_{i\epsilon}| = d_{i\epsilon}$.

For the other direction, let S be any span program computing f and of size s. Then, [41, Theorem 6] shows that S can be turned into a *canonical* span program S' of the same size. In a canonical span program, there is a one-to-one correspondence between the zeros of the function and the columns of the span program and for every zero of the function, the corresponding column has zeros in all rows activated by that input. Let B be the set of column vectors of S'. Since S' is canonical they satisfy (1). By definition, for every $y \in f^{-1}(1)$, there exists a vector c_y such that $c_y^T S'_y = \mathbf{1}$. Taking C to be the set of vectors c_y, we have (2). Since the only rows of S' that participate to generate $\mathbf{1}$ on an input $y \in f^{-1}(1)$ are those activated by y, we can assume, w.l.o.g., that the restriction of c_y to the remaining coordinates is zero. Thus, c_y can also be taken to satisfy (1). □

Proof. (of Theorem 7.1) By the lemma, it suffices to show that $\alpha := \alpha_{\mathbb{F}}(\mathcal{C}_{\text{can}}(f))$ is the smallest s such that sets of vectors B and C as in the lemma exist.

First, we show that $\alpha \leq s$: Fix i, $1 \leq i \leq n$, and $\epsilon \in \{0,1\}$. Let $B_{i\epsilon}$ be the $d_{i\epsilon} \times |f^{-1}(0)|$ matrix with $b_x^{i\epsilon}$ as columns. Let $C_{i\epsilon}$ be the $|f^{-1}(1)| \times d_{i\epsilon}$ matrix with rows $c_y^{i\epsilon}$. Now, define $K_{i\epsilon} := C_{i\epsilon} B_{i\epsilon}$. Note that $\text{rank}(K_{i\epsilon}) \leq d_{i\epsilon}$. We can write $K_{i\epsilon}$ as a sum of at most $d_{i\epsilon}$ rank-1 matrices over $\mathbb{F}$. Define now $\mathcal{K}$ to be the set of all such rank-1 matrices for all i and ϵ. Clearly, $|\mathcal{K}| \leq \sum_{i,\epsilon} \text{rank}(K_{i\epsilon}) \leq \sum_{i\epsilon} d_{i\epsilon} = s$. We now argue that $\mathcal{K}$ is indeed a set of rank-1 matrices embedded in $\mathcal{C}_{\text{can}}(f)$ and that $\sum_{K \in \mathcal{K}} K = J$. This will show that $\alpha \leq s$. We claim that all nonzero entries of $K_{i\epsilon}$ are contained in the rectangle $R_{i,\epsilon} \in \mathcal{C}_{\mathrm{can}}(f)$. By property (1) in Lemma 7.2 for $b_x^{i\epsilon}$, we see that for any x such that $x_i = \epsilon$, the corresponding columns in $B_{i\epsilon}$ are the all-0 columns. Hence, all columns in $K_{i\epsilon}$ indexed by $x \in f^{-1}(0)$ with $x_i = \epsilon$ are all-0 columns. Also by (1) applied to c_y for $y \in f^{-1}(1)$, we see that the rows of $C_{i\epsilon}$ indexed by y with $y_i = 1 - \epsilon$ are all-0 rows. Thus, any nonzero entry $K_{i\epsilon}(y,x)$ must have $y_i = \epsilon$ and $x_i = 1 - \epsilon$, i.e., such an entry is in $R_{i,\epsilon}$. Finally, the (y,x) entry of $\sum_{K \in \mathcal{K}} K$ is given by $\sum_{i\epsilon} K_{i\epsilon}(x,y) = \sum_{i,\epsilon} (c_y^{i\epsilon})^T b_x^{i\epsilon} = c_y^T b_x = 1$ by property (2) in the lemma. This shows that $\sum_{K \in \mathcal{K}} K = J$.

Next, we show that $s \leq \alpha$. Suppose $\mathcal{K}$ is a set of rank-1 matrices embedded in $\mathcal{C}_{\mathrm{can}}(f)$ such that $|\mathcal{K}| = \alpha$. Now, collect together the rank-1 matrices contained in $R_{i\epsilon} \in \mathcal{C}_{\mathrm{can}}$ (resolve ambiguities arbitrarily, but uniquely) and add them up to form $K_{i\epsilon}$. Clearly, $\mathrm{rank}(K_{i\epsilon})$ is at most the number of these rank-1 matrices. Write $K_{i\epsilon} = C_{i\epsilon} B_{i\epsilon}$, where $C_{i\epsilon}$ ($B_{i\epsilon}$) has $\mathrm{rank}(K_{i\epsilon})$ columns (rows). Define $b_x^{i\epsilon}$ ($c_y^{i\epsilon}$) to be the columns (rows) of $B_{i\epsilon}$ ($C_{i\epsilon}$). Now, define b_x (c_y) as the concatenation of $b_x^{i\epsilon}$ ($c_y^{i\epsilon}$) for all i and ϵ. Since nonzero entries of $K_{i\epsilon}$ are contained in $R_{i\epsilon}$, it is easy to ensure (1) in the lemma for $b_x^{i,\epsilon}$ and $c_y^{i,\epsilon}$. Since $\sum_{K \in \mathcal{K}} K = J$ implies that $\sum_{i\epsilon} K_{i\epsilon} = J$ and hence $\sum_{i\epsilon} (c_y^{i\epsilon})^T b_x^{i\epsilon} = 1$ for all $(y,x) \in f^{-1}(1) \times f^{-1}(0)$. It follows that $c_y^T b_x = 1$ as required in (2). We thus have vectors b_x and c_y satisfying (1) and (2) in dimension at most $\sum_{i\epsilon} \mathrm{rank}(K_{i\epsilon}) \leq |\mathcal{K}|$ and hence $s \leq \alpha$.

The claim for monotone Boolean functions is proved similarly by considering $\mathcal{C}_{\mathrm{mon}}$ and observing that an analog of Lemma 7.2 holds for a characterization of monotone span programs computing monotone Boolean functions. □

7.2 Explicit Lower Bounds on Monotone Span Program Size

Recall that for a matrix A over (indexed by) $P \times Q$ and a rectangle $R \subseteq P \times Q$, A_R denotes the restriction of A to the entries indexed by R. Analogous to Lemma 6.7, we have the following lower bound for $\alpha_{\mathbb{F}}$ in terms of ranks.

Lemma 7.3. For a rectangle cover $\mathcal{C}$ and any matrix A over $P \times Q$ over the field $\mathbb{F}$,

$$\alpha_{\mathbb{F}}(\mathcal{C}) \geq \frac{\operatorname{rank}(A)}{\max_{R \in \mathcal{C}} \operatorname{rank}(A_R)}.$$

In particular, if A is such that A_R is monochromatic (i.e., all nonzero entries of A_R are equal) for each $R \in \mathcal{C}$, then $\alpha_{\mathbb{F}}(\mathcal{C}) \geq \operatorname{rank}(A)$.

The proof of this lemma is easy and similar to that of Lemma 6.7.

Thus, if we have a matrix A indexed by $f^{-1}(0) \times f^{-1}(1)$ (or more generally by $P \times Q$, where $P \subseteq f^{-1}(0)$ and $Q \subseteq f^{-1}(1)$) such that $\operatorname{rank}_{\mathbb{F}}(A)$ is superpolynomial in the number of variables and every submatrix A_R for each rectangle $R \in \mathcal{C}_{\text{can}}(f)$ is monochromatic, we will have a superpolynomial lower bound on $\mathrm{SP}_{\mathbb{F}}(f)$. Analogous statement holds for $\mathrm{mSP}_{\mathbb{F}}(f)$ for a monotone f by considering $\mathcal{C}_{\text{mon}}(f)$. We will exploit an interesting connection in the *opposite* direction. That is, *if there is an explicit matrix A over $\mathbb{F}$ such that* $\operatorname{rank}(A)$ *is superpolynomially large compared to the minimum number of monochromatic submatrices of A needed to cover all entries of A, then we can get an explicit function f with* $\mathrm{mSP}_{\mathbb{F}}(f)$ *superpolynomial.* This reverse connection was already observed by Razborov [85] in the context of lower bounds on formula size. The connection also exists for general, i.e., not necessarily monotone, lower bounds by considering the so-called "self-complementary" covers. But, so far, only the monotone case was successfully exploited for explicit lower bounds. Razborov used this connection to obtain a superpolynomial lower bound on the monotone formula size of an explicit Boolean function from a Boolean function whose nondeterministic and co-nondeterministic communication complexity is almost quadratically smaller compared to its deterministic communication complexity. Gál used the same approach to obtain a monotone Boolean function with a superpolynomial lower bound on its monotone span program size.

Let us make the above connection more precise. Suppose A is a matrix with entries indexed by $P \times Q$ and with a monochromatic rectangle cover of size t, i.e., $A_1, \ldots, A_t$ are monochromatic submatrices of

A such that every entry $(p,q) \in P \times Q$ is in some A_i. Let $R_1,\dots,R_t$ be the rectangles in $P \times Q$ defining the nonzero entries of $A_1,\dots,A_t$. We define a monotone Boolean function $f : \{0,1\}^t \to \{0,1\}$ such that $\alpha_{\mathbb{F}}(\mathcal{C}_{\text{mon}}(f)) \geq \text{rank}(A)$. Associate a variable x_i with each rectangle R_i for $1 \leq i \leq t$. Each column q (an element of $Q \subseteq f^{-1}(1)$) will be a *minterm* and each row p (an element of $P \subseteq f^{-1}(0)$) will be a *maxterm* of f. Recall that a minterm of a monotone Boolean function f is a minimal subset of variables such that fixing each of them to 1 will ensure that the function value will be 1 no matter what we assign to the other variables. Similarly, a maxterm of a monotone Boolean function f is a minimal subset of variables such that fixing each of them to 0 will ensure that the function value will be 0 no matter what we assign to the other variables. (Geometrically, a minterm (maxterm) is a maximal subcube of the cube $\{0,1\}^t$ on which the function is constantly 1 (0).) Note that a monotone Boolean function is completely specified by its set of minterms or maxterms. Now, a minterm q is the set of all variables corresponding to the rectangles that intersect column q. Similarly, a maxterm p is the set of all variables corresponding to the rectangles that intersect row p. Next, extend each minterm q to an assignment, also denoted q, in $\{0,1\}^t$ by assigning 1 to all variables in the minterm and 0 to all the others. Similarly, extend each maxterm p to an assignment $p \in \{0,1\}^t$ by assigning 0 to all variables in the maxterm and 1 to others. This way, we get $P \subseteq f^{-1}(0)$ and $Q \in f^{-1}(1)$. It is now easy to verify that $R_1,\dots,R_t$ form the monotone canonical cover $\mathcal{C}_{\text{mon}}(P,Q)$ (cf. Theorem 6.9 and the discussion preceding it). Furthermore, matrix A over $P \times Q$ is monochromatic restricted to each R_i. Hence, by Theorem 7.1 and Lemma 7.3, we have $\text{mSP}_{\mathbb{F}}(f) \geq \alpha_{\mathbb{F}}(P,Q) \geq \text{rank}(A)$.

We now construct the explicit matrix A that accomplishes the above goal. The matrix A will be the 0–1 disjointness matrix over sets of size at most $s = \Theta(\log m)$ over a universe of size m: the rows and columns of A are indexed by $P = Q =$ subsets of $[m]$ of cardinality $\leq s$ and $A(p,q) = 1$ if and only $p \cap q = \emptyset$. It is well-known [85] that A is full-rank, i.e., $\text{rank}(A) = \sum_{i=0}^{s} \binom{m}{i}$, over any field $\mathbb{F}$. We next need to show that there is set of $O(m)$ monochromatic rectangles that cover all the entries of A. For this, we use an important pseudo-random property of Paley graphs.

A Paley graph is defined over a field K such that $|K| \equiv 1 \pmod 4$ where the vertex set is K and $i \sim j$ if and only if $i - j$ is a quadratic residue in K. A graph G is said to have *Property* $\widetilde{\mathsf{P}}_k$ if for any two disjoint sets S and T of vertices such that $|S| + |T| \leq k$, there is a vertex v in G such that v is adjacent to every vertex in S and not adjacent to any vertex in T. Theorem 13.11 from [18] shows that the Paley graph has property $\widetilde{\mathsf{P}}_k$ for $k \leq \epsilon \log|K|$ for some constant ϵ.

Let $m := |K|$ and $s = (\epsilon/2)\log m$. We identify the vertices of the Paley graph G on K with the universe $[m]$. For the matrix A, we obtain a rectangle cover by associating one rectangle with each vertex of G. By Property $\widetilde{\mathsf{P}}_{2s}$ of G, for any two disjoint subsets p and q of vertices of size at most s, there exists a vertex v of G such that every vertex in p is adjacent to v and no vertex in q is adjacent to x. It follows that the matrix A has a rectangle cover of size m while it has a rank of $\sum_{i=0}^{s} \binom{m}{i} = m^{\Theta(\log m)}$. We can use A to define an explicit Boolean function on m variables with monotone span program size $m^{\Theta(\log m)}$.

Definition of f: Given a Paley graph G on m vertices, we let $f : \{0,1\}^m \to \{0.1\}$ as follows. The m variables of f are identified with the vertices of G. For each subset p of at most s vertices f has a minterm and is given by all vertices of G (variables of f) that are adjacent to all vertices of p. Similarly, for each subset q of at most s vertices f has a maxterm and is given by all vertices of G (variables of f) that are not adjacent to any vertex of q.

The following theorem follows from the foregoing arguments.

Theorem 7.4. For the function f defined above and any field $\mathbb{F}$, $\mathrm{mSP}_{\mathbb{F}}(f) = m^{\Omega(\log m)}$.

It is known [10] that span programs become exponentially weak if we restrict them to be monotone. On the other hand, there exist functions with linear size monotone span programs but requiring exponential size monotone Boolean circuits [6]. Another interesting result in [10] is the exponential separation of monotone span programs over fields of different characteristics.

The characterization of Gál of span program size in terms of gaps between combinatorial cover numbers and ranks seems to be a very promising criterion for lower bounds. It has already been extensively used in [10]. These results also bring together the theme that communication complexity techniques provide very powerful approaches to problems in lower bounds.

8

Conclusions and Open Problems

In this last section, we make some concluding remarks and mention several open problems.

Rigid matrices over small number fields: Proving strong superlinear lower bounds on the rigidity of explicit matrices (Open Question 2.1) remains an intriguing open problem. On the one hand, intuitive criteria, such as having all submatrices nonsingular, are insufficient as shown in Theorem 2.12. On the other hand, using algebraic independence of very high-degree among entries of matrices such as $(\sqrt{p_{ij}})$, we were able to prove quadratic lower bounds on their rigidity (Corollary 2.19 and Theorem 2.21). These lower bounds are not satisfactory since the entries of those matrices live in number fields of exponentially large dimensions. It would be interesting to prove $\mathcal{R}_A(\epsilon n) \geq n^{1+\delta}$ for $A \in \mathbb{C}^{n\times n}$, where entries a_{ij} of A span a number field of poly(n) dimension, i.e., $\dim(\mathbb{Q}(a_{ij}) : \mathbb{Q}) \leq \text{poly}(n)$. Note that the Fourier transform matrix $F = (\omega^{ij})_{i,j=0}^{n-1}$ defines a number field of dimension $\varphi(n) \leq n - 1$ [50, Ch. VI, Theorem 3.1].

Rigidity of Vandermonde matrices: Note also that the Fourier transform matrix is a Vandermonde matrix. However, we do not know strong lower bounds on the rigidity of even a generic Vandermonde matrix, i.e., $V = (x_i^{j-1})_{1 \leq i,j \leq n}$, where the x_i are *algebraically independent over* $\mathbb{Q}$. Can we prove (or disprove) that $\mathcal{R}_V(\epsilon n) \geq n^{1+\delta}$? Theorem 2.17 gives an $\Omega(n^2)$ lower bound, but only when $r = O(\sqrt{n})$. For an arbitrary Vandermonde matrix, we only know a lower bound of $\Omega(n^2/r)$ (Theorem 2.10) and only a slightly better bound of $\Omega(\frac{n^2}{r} \log \frac{n}{r})$ for specific Vandermonde matrices, e.g., when the x_i are distinct positive integers and when they are powers of a primitive root of unity of a prime order (Theorem 2.8 and the discussion following it).

Rigidity of character tables: The Fourier transform matrix and the Sylvester-type Hadamard matrix (defined by (2.3)) are conjectured to have high-rigidity. They are the character tables of the groups $\mathbb{Z}_n$ and $\mathbb{Z}_2^t$, respectively. We pose the following more general question: *Given a finite abelian group G of order n, let $A_G \in \mathbb{C}^{n \times n}$ be the character table of G. Prove that $\mathcal{R}_{A_G}(\epsilon n) \geq n^{1+\delta}$ for some group G.* For the character table of any finite abelian group, a lower bound of $\Omega(n^2/r)$ follows from Theorem 3.8 using spectral techniques.

Paturi–Pudlák dimensions over infinite fields: The techniques from Section 2.6 currently apply to finite fields. Proving nontrivial bounds on the inner and outer dimensions over infinite fields may be an interesting extension to [70].

Spectral techniques: In Section 3, we looked at several rank robustness functions expressed in terms of ℓ_2-norm. Spectral techniques have been helpful in proving lower bounds on such normed variants of rigidity. We have also seen that such variants, and spectral techniques in general, have been very successful in proving lower bounds on *bounded coefficient* complexity of linear and bilinear circuits. Notable among these are the $\Omega(n \log n)$ lower bound for the Fourier transform and the $\Omega(n^2 \log n)$ lower bound for matrix multiplication.

Sign-rank versus margin complexity: Spectral techniques are also used in Section 4 to prove lower bounds on sign-rank and margin complexity; note that both lower bounds (Theorems 4.2 and 4.7) are $\sqrt{mn}/\|A\|$. While the lower bound on margin complexity of an Hadamard matrix is tight, we do not know if the lower bound on its sign-rank is tight.

From Lemma 4.17, sign-rank can be at most quadratically (ignoring log-factors) larger than margin complexity. Recall that [2] show that for almost all sign matrices, sign-rank is $\Theta(n)$. Hence, a *random* matrix shows that this gap is tight apart from log-factors. It would thus be interesting to find an explicit sign matrix with a sign-rank of $\Omega(n)$.

Connections between complexity measures of sign matrices and communication complexity: Consider the sequence of inequalities[1]

$$\sqrt{\text{sign-rank}(A)} \lesssim \text{mc}(A) \leq \gamma_2^{1/2\epsilon}(A) \leq \sqrt{\text{rank}(A)}, \tag{8.1}$$

valid for any matrix $A \in \{-1,+1\}^{n\times n}$ and compare it to the hierarchy of communication complexity classes:

$$\text{UPP} \supseteq \text{PP}^{\text{cc}} \supseteq \text{BPP}^{\text{cc}} \supseteq \text{P}^{\text{cc}}. \tag{8.2}$$

Recall that sign-rank(A) and mc(A) (and hence disc(A)) *characterize* the communication complexity classes UPP and PP$^{\text{cc}}$, respectively. More specifically, if these functions are bounded above by exp(polylog log(n)) for an infinite family of matrices $\{A_n\}$, then the language defined by $\{A_n\}$ is in the corresponding communication complexity classes. The next two parameters $\gamma_2^{1/2\epsilon}(A)$ and rank(A) are only known to give *lower bounds* on bounded error probabilistic communication complexity (Theorem 5.8) and deterministic communication complexity, respectively. It would be very interesting to prove (or disprove) that the parameters $\gamma_2^{1/2\epsilon}$ and rank can also be used to characterize the corresponding communication complexity classes. It would also be interesting to investigate if γ_2-norm can be used to characterize quantum communication complexity. Note that proving that polylog(rank(A)) is an upper bound on deterministic communication complexity is precisely the log-rank conjecture.

Gaps between log-rank and communication complexity: The log-rank conjecture is one of the basic open questions in complexity theory. It would be interesting to improve the gap between log-rank and communication complexity given by Theorem 5.7, for instance, to quadratic by constructing an n-variable Boolean function of degree

[1] We use $\lesssim$ instead of $\leq$ to ignore factors up to $\log n$.

$O(\sqrt{n})$ and sensitivity $\Omega(n)$. More generally, is it possible to construct Boolean matrices of small rank r (say, $r = \exp((\log n)^{O(1)})$) in which every monochromatic rectangle is small (say, containing at most an $\exp(-(\log r)^{\omega(1)})$-fraction of entries)? Note that if, for some constant $c > 0$, every rank-r Boolean matrix contains a monochromatic submatrix with at least an $\exp(-(\log r)^c)$-fraction of entries, then the log-rank conjecture holds [68].

Lower bounds for higher classes in communication complexity: Regarding the "higher" complexity classes in the two-party communication model, separating $\mathrm{PH}^{\mathrm{cc}}$ from $\mathrm{PSPACE}^{\mathrm{cc}}$ is a long-standing open question [5]. The approaches via matrix rigidity or margin complexity rigidity from Section 5.4 seem to demand extremely strong lower bounds on those parameters. However, it is also a challenging open question to find an explicit language outside Σ_2^{cc}; we know explicit languages outside several classes smaller than Σ_2^{cc} such as $\mathrm{P}^{\mathrm{cc}}, \mathrm{NP}^{\mathrm{cc}}, \mathrm{Co\text{–}NP}^{\mathrm{cc}}$, and $\mathrm{BPP}^{\mathrm{cc}}$. We also know explicit languages, e.g., inner product mod-2, outside $\mathrm{PP}^{\mathrm{cc}}$ and UPP, thanks to their characterizations in terms of margin complexity and sign-rank. Recall also the recent breakthrough result of [88] that $\Sigma_2^{\mathrm{cc}} \not\subseteq \mathrm{UPP}$ (Theorem 5.20).

Lower bounds on dimensions of graphs: Proving strong lower bounds, i.e., at least $2^{\log\log n^{\omega(1)}}$, on depth-3 graph complexity for explicit $n \times n$ bipartite graphs is one approach to finding explicit languages outside Σ_2^{cc}. An $n^{\Omega(1)}$ lower bound on depth-3 graph complexity will have an even more amazing consequence; it would imply a superlinear lower bound on log-depth Boolean circuits. Unfortunately, we currently know only $\tilde{\Omega}(\log^3 n)$ lower bounds (Theorem 6.19).

Proving superlogarithmic lower bounds on the affine and projective dimensions of explicit graphs is a challenging open problem. For instance, what are the projective and affine dimensions of graphs such as the Paley graph or explicit Ramsey-like graphs? It is interesting that depth-3 graph complexity can also be related to certain geometric representations of graphs. Indeed, Theorem 6.19 shows that a depth-3 lower bound can be derived from a lower bound on the dimension d for the following kind of geometric representation: associate vectors $u_x, v_y \in \mathbb{R}^d$ such that $u_x^T v_y \geq \delta$ if (x,y) is an edge and $u_x^T v_y \leq -\delta$ if

(x,y) is not an edge, where $\delta > 0$ is a constant. As seen in the proof of Theorem 6.19, a lower bound of $d = \Omega(n)$ for the Paley graph follows from a lower bound on the sign-rank of an Hadamard matrix.

Span program complexity: The parameters $\alpha(\mathcal{C})$ (Section 6.2.1) and $\alpha_{\mathbb{F}}(\mathcal{C})$ (Definition 7.1) have lower bounds in terms of certain rank-like functions of (partial) matrices (Lemmas 6.7 and 7.3). When applying this connection to prove lower bounds on monotone span program size, we used an explicit matrix A such that $\mathrm{rank}(A)$ is quasi-polynomially large compared to the size of a monochromatic rectangle cover of A (see, Section 7.2). In contrast, recall that, to prove the log-rank conjecture in two-party communication complexity, we need to show that *every* Boolean matrix can be covered by quasi-polynomially many monochromatic rectangles compared to its rank.

Computational complexity of rank robustness functions: We observe here a correlation between the complexity of computing various rank robustness functions and the difficulty of proving lower bounds on those functions for explicit matrices. This is reminiscent of the theory of *natural proofs* [87] in the context of lower bounds for explicit Boolean functions. For concreteness, let us consider the three functions: (i) Valiant's general rigidity $\mathcal{R}_A(r)$ in Definition 2.1, (ii) Raz's geometric rigidity in Definition 3.4, and (iii) ℓ_2-rigidity in Definition 3.2. As we will explain below, functions (i)–(iii) are apparently of decreasing computational complexity and, interestingly, proving explicit lower bounds on these functions also appears to be of decreasing difficulty. Moreover, the corresponding circuit lower bounds that can be derived also appear to be of decreasing strength and generality as we go from (i) to (iii).

(1) Given a matrix A (say, over a finite field) and an integer r, computing $\mathcal{R}_A(r)$ appears to be an intractable question. In fact, Deshpande (private communication) showed the following problem to be NP-complete: Given a matrix $A \in {\mathbb{F}_q}^{m\times n}$, and integers $0 \le w \le n^2$ and $0 \le r \le n$, does there exist a matrix $C \in {\mathbb{F}_q}^{m\times n}$ with no more than w nonzero entries, such that $\mathrm{rank}(A + C) \le r$? The reduction is from the Nearest Codeword Problem. Alekhnovich [1] shows an even stronger

hardness result based on a conjecture about the average hardness of *maximum satisfying linear subsystem* in a system of linear equations over $GF(2)$: it is infeasible to distinguish between matrices M with $\mathcal{R}_M(\epsilon n) < n^{1+\delta}$ and random M (hence of quadratic rigidity). It follows that having high rigidity is unlikely to be an *easy* property of a matrix. This may partly explain why coming up with explicit matrices of high rigidity remains a challenging open problem.

(2) A notion essentially equivalent to Raz's geometric rigidity (Definition 3.4) is studied in computational geometry under the name *outer radius* of a set of points (see, e.g., [100]). In [100] and other related works, the complexity of computing outer radius has been investigated. From these results, we obtain the following: (a) For any $r \leq n$, $\mathrm{Rig}_r(A)$ can be approximated in polynomial time within a factor of $O(\sqrt{\log m})$. When r is a *constant*, $\mathrm{Rig}_r(A)$ can be approximated within a factor of $(1+\epsilon)$ for any constant $\epsilon > 0$. (b) For any $r \leq n^\epsilon$, where $0 < \epsilon < 1$, $\mathrm{Rig}_r(A)$ cannot be approximated within a factor of $(\log m)^\delta$, where $\delta > 0$ is a fixed constant, unless $\mathrm{NP} \subseteq \tilde{\mathrm{P}}$. Furthermore, it is NP-hard to approximate $\mathrm{Rig}_{n-1}(A)$ within any constant factor. It follows that $\mathrm{Rig}_r(A)$ is in general hard to compute but seems to be of varying difficulty to approximate. We also saw an explicit matrix, namely, a generalized Hadamard matrix H, that has $\mathrm{Rig}_r(H) \geq \sqrt{n-r}$ (Corollary 3.12). However, for Raz's lower bound on matrix multiplication, we only need that a random matrix Y has $\mathrm{Rig}_{\epsilon n}(Y) = \Omega(\sqrt{n})$ (Lemma 3.20(2)).

(3) Measuring the distance of a matrix from the set of low rank matrices in ℓ_2-distance as opposed to Hamming distance seems to make the problem considerably easier. This is illustrated by the complete characterization of ℓ_2-$\mathrm{Rig}_r(A)$ in terms of singular values of A (Lemma 3.5). As a result of this characterization, we have both a polynomial time algorithm to compute ℓ_2-$\mathrm{Rig}_r(A)$ and explicit matrices with strong lower bounds, e.g., Hadamard matrix, on this rank robustness function.

Understanding the computational complexity of other rank robustness functions is an interesting and emerging area of research. In particular, it would be interesting to prove hardness results (or find efficient algorithms) for Paturi-Pudlák dimensions.

Linial et al. [53], formulate margin and γ_2-norm in terms of semidefinite programs. Thus, they are computable in polynomial time. However, the computational complexity of the sign-rank of a matrix is still unknown. Very recently, Kushilevitz and Weinreb [49] studied the complexity of computing the communication complexity of a Boolean matrix and showed that, under certain cryptographic assumptions, it is intractable to compute, or even to approximate well, the communication complexity.

Acknowledgments

I am grateful to the following people for their careful reading of the early versions of this paper and valuable feedback: Amit Deshpande, Seny Kamara, Neeraj Kayal, Meena Mahajan, Jayalal Sharma, Madhu Sudan, and Ramarathnam Venkatesan.

References

[1] M. Alekhnovich, "More on average case vs approximation complexity," in *FOCS*, pp. 298–307, IEEE Computer Society, 2003.

[2] N. Alon, P. Frankl, and V. Rödl, "Geometric realizations of set systems and probabilistic communication complexity," *Proceedings of the 26th IEEE Symposium on Foundations of Computer Science*, pp. 277–280, 1985.

[3] R. I. Arriaga and S. Vempala, "An algorithmic theory of learning: Robust concepts and random projection," *IEEE Symposium on Foundations of Computer Science*, pp. 616–623, 1999.

[4] L. Babai and P. Frankl, *Linear Algebra Methods in Combinatorics, Preliminary version 2*. Department of Computer Science, University of Chicago, 1992.

[5] L. Babai, P. Frankl, and J. Simon, "Complexity classes in communication complexity theory," *Proceedings of the 27th IEEE Symposium on Foundations of Computer Science*, pp. 337–347, 1986.

[6] L. Babai, A. Gál, and A. Wigderson, "Superpolynomial lower bounds for monotone span programs," *Combinatorica*, vol. 19, no. 3, pp. 301–319, 1999.

[7] W. Baur, "Simplified lower bounds for polynomials with algebraic coefficients," *Journal of Complexity*, vol. 13, pp. 38–41, 1997.

[8] W. Baur and V. Strassen, "The complexity of partial derivatives," *Theoretical Computer Science*, vol. 22, pp. 317–330, 1983.

[9] A. Beimel, A. Gál, and M. Paterson, "Lower bounds for monotone span programs," *Computational Complexity*, vol. 6, no. 1, pp. 29–45, 1996/97.

[10] A. Beimel and E. Weinreb, "Separting the power of monotone span programs over different fields," *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, pp. 428–437, 2003.

[11] S. Ben-David, N. Eiron, and H. U. Simon, "Limitations of learning via embeddings in Euclidean half spaces," *Journal of Machine Learning Research*, vol. 3, pp. 441–461, 2003.

[12] A. S. Besicovitch, "On the linear independence of fractional powers of integers," *Journal of the London Mathematical Society*, vol. 15, pp. 3–6, 1940.

[13] R. Bhatia, *Matrix Analysis.* Vol. 169 of *Graduate Texts in Mathematics*, New York, NY: Springer-Verlag, 1997.

[14] R. Blei, "An elementary proof of the grothendieck inequality," *Proceedings of the American Mathematical Society*, vol. 100, no. 1, pp. 58–60, 1987.

[15] B. Bollig, M. Sauerhoff, D. Sieling, and I. Wegener, "On the power of different types of restricted branching programs," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 1, no. 26, 1994.

[16] B. Bollobás, *Modern Graph Theory.* Vol. 184 of *Graduate Texts in Mathematics*, New York, NY: Springer-Verlag, 1998.

[17] B. Bollobás, *Linear Analysis.* Cambrdige Mathematical Textbooks. Cambridge Universty Press, Second ed., 1999.

[18] B. Bollobás, *Random Graphs.* Vol. 73 of *Cambrdige Studies in Advanced Mathematics*, Cambridge, United Kingdom: Cambridge University Press, Second ed., 2001.

[19] J. Bruck and R. Smolensky, "Polynomial threshold functions, AC^0 functions and spectral norms," *Proceedings of the 31st Symposium on Foundations of Computer Science*, pp. 632–641, 1990.

[20] P. Bürgisser, M. Clausen, and M. A. Shokhrollahi, *Algebraic Complexity Theory.* Springer-Verlag, 1997.

[21] P. Bürgisser and M. Lotz, "Lower bounds on the bounded coefficient complexity of bilinear maps," *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, pp. 659–668, 2002.

[22] P. Bürgisser and M. Lotz, "Lower bounds on the bounded coefficient complexity of bilinear maps," *Journal of the ACM*, vol. 51, no. 3, pp. 464–482, 2004.

[23] J.-Y. Cai and E. Bach, "On testing for zero polynomials by a set of points with bounded precision," *Theoretical Computer Science*, vol. 296, no. 1, pp. 15–25, 2003.

[24] B. Chazelle, "A spectral approach to lower bounds," *Proceedings of the 35th IEEE Symposium on Foundations of Computer Science*, pp. 674–682, 1994.

[25] Z. Chen and M. Kao, "Reducing randomness via irrational numbers," *Proceedings of the 29th Symposium Theory of Computing (STOC)*, pp. 200–209, 1997.

[26] R. de Wolf, "Lower bounds on matrix rigidity via a quantum argument," in *ICALP (1)*, (M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds.), pp. 62–71, Vol. 4051 of *Lecture Notes in Computer Science*, Springer, 2006.

[27] P. Erdős, R. Graham, and E. Szemerédi, "On sparse graphs with dense long paths," *Computers and Mathematics with Applications*, pp. 365–369, 1976.

[28] J. Forster, "A linear lower bound on the unbounded error probabilistic communication complexity," *Journal of Computer and System Sciences*, vol. 65, no. 4, pp. 612–625, Special issue on Complexity, (Chicago, IL, 2001), 2002.

[29] J. Forster, M. Krause, S. V. Lokam, R. Mubarakzjanov, N. Schmitt, and H. U. Simon, "Relations between communication complexity, linear arrangements, and computational complexity," in *FST TCS 2001: Foundations of Software Technology and Theoretical Computer Science (Bangalore)*, pp. 171–182, Vol. 2245 of *Lecture Notes in Computer Science*, Berlin: Springer, 2001.

[30] J. Forster and H.-U. Simon, "On the smallest possible dimension and the largest possible margin of linear arrangements representing given concept classes," *Theoretical Computer Science*, vol. 350, no. 1, pp. 40–48, 2006.

[31] J. Friedman, "A note on matrix rigidity," *Combinatorica*, vol. 13, no. 2, pp. 235–239, 1993.

[32] A. Gál, "A characterization of span program size and improved lower bounds for monotone span programs," *Computational Complexity*, vol. 10, no. 4, pp. 277–296, 2001.

[33] G. H. Golub and C. F. Van Loan, *Matrix Computations.* The Johns Hopkins University Press, Third ed., 1996.

[34] D. Grigoriev, "Using the notions of separability and indepndence for proving lower bounds on the circuit complexity," *Notes of the Leningrad Branch of the Steklov Mathematical Institute*, 1976.

[35] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turan, "Threshold functions of bounded depth," *Journal of Computer and System Sciences*, vol. 46, pp. 129–154, 1993.

[36] B. Halstenberg and R. Reischuk, "Relations between communication complexity classes," *Journal of Computer and System Sciences*, vol. 41, pp. 402–429, 1990.

[37] J. Håstad and A. Wigderson, "The randomized communication complexity of set disjointness," *Theory of Computing*, vol. 3, no. 1, pp. 211–219, 2007.

[38] A. J. Hoffman and H. W. Wielandt, "The variation of the spectrum of a normal matrix," *Duke Mathematical Journal*, vol. 20, pp. 37–39, 1953.

[39] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge University Press, February 1990.

[40] M. Karchmer and A. Wigderson, "Monotone circuits for connetivity require super-logarithmic depth," *SIAM Journal on Discrete Mathematics*, vol. 3, no. 2, pp. 255–265, 1990.

[41] M. Karchmer and A. Wigderson, "On span programs," in *Proceedings of the 8th Annual Structure in Complexity Theory Conference (San Diego, CA, 1993)*, pp. 102–111, Los Alamitos, CA, 1993. IEEE Computer Society Press.

[42] B. S. Kashin and A. A. Razborov, "Improved lower bounds on the rigidity of hadamard matrices," *Matematicheskie Zametki*, vol. 63, no. 4, pp. 535–540, English translation at http://www.mi.ras.ru/ razborov/hadamard.ps, 1998.

[43] H. Klauck, "Lower bounds for quantum communication complexity," *SIAM Journal on Computing*, vol. 37, no. 1, pp. 20–46, 2007.

[44] A. R. Klivans and R. A. Servedio, "Learning DNF in time $2^{\tilde{O}(n^{1/3})}$," *Journal of Computer and System Sciences*, vol. 68, no. 2, pp. 303–318, 2004.

[45] A. Kotlov, L. Lovász, and S. Vempala, "The Colin de Verdière number and sphere representations of a graph," *Combinatorica*, vol. 17, pp. 483–521, 1997.

[46] M. Krause, "Geometric arguments yield better bounds for threshold circuits and distributed computing," *Theoretical Computer Science*, vol. 156, nos. 1&2, pp. 99–117, 1996.

[47] M. Krause and P. Pudlák, "On the computational power of depth 2 circuits with threshold and modulo gates," *STOC*, pp. 48–57, 1994.

[48] E. Kushilevitz and N. Nisan, *Communication Complexity.* Cambridge University Press, 1997.

[49] E. Kushilevitz and E. Weinreb, "On the complexity of communication complexity," *STOC 2009*, 2009.

[50] S. Lang, *Algebra.* Addison-Wesley Publishing Company, Third ed., 1993.

[51] D. Lewin and S. Vadhan, "Checking polynomial identities over any field: Towards a derandomization?," *Proceedings of the 30th Symposium on Theory of Computing (STOC)*, pp. 438–447, 1998.

[52] T. Lickteig, *Ein elementarer Beweis für eine geometrische Gradschranke für die Zahl der Operationen bei der Berechnung von Polynomen.* Diplomarbeit, Universität Konstanz, 1980.

[53] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman, "Complexity measures of sign matrices," *Combinatorica*, 2006.

[54] N. Linial and A. Shraibman, "Lower bounds in communication complexity based on factorization norms," in *STOC '07: Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pp. 699–708, New York, NY, USA: ACM, 2007.

[55] N. Linial and A. Shraibman, "Learning complexity vs communication complexity," *CCC '08: Conference on Computational Complexity*, 2008.

[56] S. V. Lokam, "On the rigidity of Vandermonde matrices," *Theoretical Computer Science*, vol. 237, nos. 1–2, pp. 477–483, Presented at the DIMACS-DIMATIA workshop on *Arithmetic Circuits and Algebraic Methods, June, 1999*, 2000.

[57] S. V. Lokam, "Spectral methods for matrix rigidity with applications to size-depth trade-offs and communication complexity," *Journal of Computer and System Sciences*, vol. 63, no. 3, pp. 449–473, 2001.

[58] S. V. Lokam, "Graph complexity and slice functions," *Theory of Computing Systems*, vol. 36, no. 1, pp. 71–88, 2003.

[59] S. V. Lokam, "Quadratic lower bounds on matrix rigidity," *Proceedings of Theory and Applications of Models of Computation (TAMC)*, vol. 3959, pp. 295–307, *Lecture Notes in Computer Science*, 2006.

[60] L. Lovász, "On the ratio of optimal integral and fractional covers," *Discrete Mathematics*, vol. 13, no. 4, pp. 383–390, 1975.

[61] L. Lovász, "On the shannon capacity of a graph," *IEEE Transactions on Information Theory*, vol. 25, pp. 1–7, 1979.

[62] L. Lovász and M. Saks, "Communication complexity and combinatorial lattice theory," *Journal of Computer and System Sciences*, vol. 47, no. 2, pp. 322–349, *29th Annual IEEE Symposium on Foundations of Computer Science.* White Plains, NY, 1988, 1993.

[63] K. Mehlhorn and E. M. Schmidt, "Las Vegas is better than determinism in VLSI and distributed computing (Extended Abstract)," in *STOC*, pp. 330–337, ACM, 1982.

[64] G. Midrijanis, “Three lines proof of the lower bound for the matrix rigidity,” *CoRR*, vol. abs/cs/0506081, 2005.

[65] J. Morgenstern, “Note on a lower bound of the linear complexity of the fast Fourier transform,” *Journal of the ACM*, vol. 20, no. 2, pp. 305–306, 1973.

[66] N. Nisan and M. Szegedy, “On the degree of boolean functions as real polynomials,” *Computational Complexity*, vol. 4, pp. 301–313, 1994.

[67] N. Nisan and A. Wigderson, “Lower bounds on arithmetic circuits via partial derivatives,” *Proceedings of the 36th IEEE Symposium on Foundations of Computer Science*, pp. 16–25, 1995.

[68] N. Nisan and A. Wigderson, “On rank vs communication complexity,” *Combinatorica*, vol. 15, no. 4, pp. 557–565, 1995.

[69] N. Nisan and A. Wigderson, “On the complexity of bilinear forms,” *Proceedings of the 27th ACM Symposium on the Theory of Computing*, pp. 723–732, 1995.

[70] R. Paturi and P. Pudlák, “Circuit lower bounds and linear codes,” in *Teoria Slozhnosti Vychislenij IX*, (E. A. Hirsch, ed.), Vol. 316, pp. 188–204, Notes of Mathematical Seminars of St. Petersburg Department of Steklov Institute of Mathematics, 2004.

[71] R. Paturi and J. Simon, “Probabilistic communication complexity,” *Journal of Computer and System Sciences*, vol. 33, no. 1, pp. 106–123, *25th Annual Symposium on Foundations of Computer Science* (Singer Island, Florida, 1984), 1986.

[72] N. Pippenger, “Superconcentrators,” *SIAM Journal on Computing*, vol. 6, no. 2, pp. 298–304, 1977.

[73] G. Pólya and G. Szegö, *Problems and Theorems in Analysis, Volume II.* New York, NY: Springer-Verlag, 1976.

[74] P. Pudlák, “Large communication in constant depth circuits,” *Combinatorica*, vol. 14, no. 2, pp. 203–216, 1994.

[75] P. Pudlák, “A note on the use of determinant for proving lower bounds on the size of linear circuits,” *Electronic Colloquium on Computational Complexity (ECCC)*, 1998.

[76] P. Pudlák and V. Rödl, “A combinatorial approach to complexity,” *Combinatorica*, vol. 12, no. 2, pp. 221–226, 1992.

[77] P. Pudlák and V. Rödl, “Modified ranks of tensors and the size of circuits,” *Proceedings of the 25th ACM Symposium on the Theory of Computing*, pp. 523–531, 1993.

[78] P. Pudlák and V. Rödl, “Some combinatorial-algebraic problems from complexity theory,” *Discrete Mathematics*, vol. 136, nos. 1–3, pp. 253–279, Trends in discrete mathematics, 1994.

[79] P. Pudlák, V. Rödl, and P. Savický, “Graph complexity,” *Acta Informatica*, vol. 25, no. 5, pp. 515–535, 1988.

[80] P. Pudlák, V. Rödl, and J. Sgall, “Boolean circuits, tensor ranks and communication complexity,” *Manuscript*, March 1994.

[81] P. Pudlák and Z. Vavřín, “Computation of rigidity of order n^2/r for one simple matrix,” *Commentationes Mathematicae Universitatis Carolinae*, vol. 32, no. 2, pp. 213–218, 1991.

[82] J. Radhakrishnan and A. Ta-Shma, “Bounds for dispersers, extractors, and depth-two superconcentrators,” *SIAM Journal on Discrete Mathematics*, vol. 13, no. 1, pp. 2–24 (electronic), 2000.

[83] R. Raz, “On the complexity of matrix product,” in *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pp. 144–151, ACM Press, 2002.

[84] A. A. Razborov, “On rigid matrices,” *Manuscript, (Russian)*, 1989.

[85] A. A. Razborov, “Applications of matrix methods to the theory of lower bounds in computational complexity,” *Combinatorica*, vol. 10, no. 1, pp. 81–93, 1990.

[86] A. A. Razborov, “A note on the use of determinant for proving lower bounds on the size of linear circuits,” *Electronic Colloquium on Computational Complexity (ECCC)*, 1998.

[87] A. A. Razborov and S. Rudich, “Natural proofs,” *Journal of Computer and System Sciences*, vol. 55, no. 1, part 1, pp. 24–35, *26th Annual ACM Symposium on the Theory of Computing (STOC ’94)* (Montreal, PQ, 1994), 1997.

[88] A. A. Razborov and A. A. Sherstov, “The sign-rank of AC^0,” *FOCS 2008, Proceedings of Symposium on Foundations of Computer Science*, 2008.

[89] L. Rónyai, L. Babai, and M. K. Ganapathy, “On the number of zero-patterns of a sequence of polynomials,” *Journal of the American Mathematical Society*, vol. 14, no. 3, pp. 717–735 (electronic), 2001.

[90] A. A. Sherstov, “The pattern matrix method for lower bounds on quantum communication,” in *STOC*, (R. E. Ladner and C. Dwork, eds.), pp. 85–94, ACM, 2008.

[91] V. Shoup and R. Smolensky, “Lower bounds for polynomial evaluation and interpolation,” *Proceedings of the 32nd IEEE Symposium on Foundations of Computer Science*, pp. 378–383, 1991.

[92] J. W. Silverstein, “The smallest eigenvalue of a large dimensional wishart matrix,” *The Annals of Probability*, vol. 13, no. 4, pp. 1364–1368, 1985.

[93] D. A. Spielman, V. Stemann, and M. A. Shokhrollahi, “A remark on matrix rigidity,” *Information Processing Letters*, vol. 64, no. 6, pp. 283–285, 1997.

[94] P. Stevenhagen and H. W. Lenstr Jr., “Chebotarëv and his density theorem,” *Mathematical Intelligencer*, vol. 18, no. 2, pp. 26–37, 1996.

[95] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory.* Academic Press, 1990.

[96] T. Tao, “An uncertainty principle for cyclic groups of prime order,” *Mathematical Research Letters*, vol. 12, pp. 121–127, 2005.

[97] J. Tarui, “Randomized polynomials, threshold circuits and polynomial hierarchy,” *Theoretical Computer Science*, vol. 113, pp. 167–183, 1993.

[98] L. Valiant, “Graph–theoretic arguments in low-level complexity,” in *Proceedings of the 6th Symposium on Mathematical Foundations of Computer Science*, pp. 121–127, 1977. *Lecture Notes in Computer Science.*

[99] V. Vapnik, *The Nature of Statistical Learning Theory.* Springer-Verlag, 1999. ISBN 0-387-98780-0.

[100] K. R. Varadarajan, S. Venkatesh, Y. Ye, and J. Zhang, “Approximating the radii of point sets,” *SIAM Journal on Computing*, vol. 36, no. 6, pp. 1764–1776, 2007.

[101] J. von zur Gathen and J. Gerhard, *Modern Computer Algebra.* Cambridge University Press, 1999.

[102] H. E. Warren, "Lower bounds for approximation by nonlinear manifolds," *Transactions of the American Mathematical Society*, vol. 133, pp. 167–178, 1968.

[103] A. Wigderson, "The fusion method for lower bounds in circuit complexity," in *Combinatorics, Paul Erdős is Eighty,* Vol. 1, pp. 453–468, Budapest: János Bolyai Mathematical Society, Bolyai Society Mathematical Studies, 1993.

[104] A. C.-C. Yao, "Some complexity questions related to distributive computing," *Proceedings of the 11th ACM Symposium on the Theory of Computing*, pp. 209–213, 1979.