ALGEBRAIC-GEOMETRIC METHODS FOR COMPLEXITY LOWER BOUNDS

By

PRANAV DANDEKAR

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2004

I dedicate this work to my parents.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Professor Meera Sitharam. Though it is a cliché to say so, this work would have been impossible without her very active involvement, constant encouragement and the inspiration I drew from her. The best analogy for what she taught me is someone teaching a toddler to first crawl, then stand, walk and run by holding his hand. Working with her has been a wonderful learning experience. She is everything a great advisor is, and then some.

I am thankful to Prof. Tim Davis and Prof. Andrew Vince for serving on my supervisory committee, and for posing some thought-provoking questions about my work. I would also like to thank my friends here at UF, specially Srijit Kamath and Andrew Lomonosov, for helping me write this thesis and feigning interest in my work during idle discussions.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS . . . iv

ABSTRACT . . . vii

CHAPTER

1 INTRODUCTION . . . 1

2 MATHEMATICAL PRELIMINARIES . . . 3

- 2.1 Matrices . . . 3
- 2.2 Inner Products and Norms . . . 4
- 2.3 Geometric Concepts . . . 5

3 APPLICATION IN COMPUTER SCIENCE . . . 7

- 3.1 Communication Complexity . . . 7
- 3.2 Threshold Circuit Complexity . . . 8
- 3.3 Maximal Margin Classifiers . . . 8
- 3.4 Geometric Embeddings . . . 9

4 REALIZATIONS AND REALIZABILITY . . . 12

- 4.1 Half Space Realization . . . 12
  - 4.1.1 Known Bounds . . . 12
  - 4.1.2 Generalized Lower Bound . . . 14
- 4.2 A New Notion of Realization . . . 15
  - 4.2.1 Constructing B from Forster's Realization . . . 15
  - 4.2.2 Niceness Properties of Realization Subspaces . . . 16
  - 4.2.3 Rephrasing Forster in Terms of B . . . 17
  - 4.2.4 Rephrasing the Generalized Lower Bound in Terms of B . . . 20
  - 4.2.5 Intuitive Justification . . . 21

5 NICE REALIZATIONS . . . 22

- 5.1 New Definition of Niceness . . . 22
  - 5.1.1 Improved Bounds for $M \in \{-1, +1\}^{n \times m}$ . . . 22
  - 5.1.2 Improved Bounds for $M \in \mathbb{R}^{n \times m}$ . . . 25
  - 5.1.3 Why Is This Bound Better? . . . 28
- 5.2 Nice Properties of Specific Constructions . . . 28
  - 5.2.1 Forster's Construction . . . 28
  - 5.2.2 Forster's construction has property 5.1 . . . 29
  - 5.2.3 Belcher-Hicks-Sitharam Construction . . . 31
  - 5.2.4 Belcher-Hicks-Sitharam construction has property 5.1 . . . 31

6 THE TOPOLOGY OF BALLS 34

6.1 The Generalized Almost-Spherical Property 34
6.2 Observations About The Generalized Almost-spherical Property 35
6.3 Observations on the Figure 39

7 CONCLUSION AND FUTURE WORK 42

REFERENCES 45

BIOGRAPHICAL SKETCH 46

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

ALGEBRAIC-GEOMETRIC METHODS FOR COMPLEXITY LOWER BOUNDS

By

Pranav Dandekar

December 2004

Chair: Meera Sitharam
Major Department: Computer and Information Science and Engineering

Complexity lower bounds is the study of the *intrinsic complexity* of computational problems. Lower bound arguments prove the (in)solvability of a problem in a certain time/space assuming a certain model of computation for *all* valid problem instances. In this work we study a geometric lower bound problem–to find the minimal dimension of a subspace that intersects a given set of cells (orthants) in $\mathbb{R}^m$. This problem underlies many problems in diverse areas of Computer Science such as Communication Complexity, Learning Theory, Geometric Embeddings, etc. We start by rephrasing the proof of a long-standing open problem in Communication Complexity. Rephrasing the problem in terms of *realization subspaces* leads us to some interesting properties of these subspaces. We relate these properties and use them to improve the known result. We also pose some questions about these subspaces and their relation to some results in functional analysis. Finally, we identify the various directions to proceed in using this geometric lower bound problem as a starting point.

CHAPTER 1
INTRODUCTION

Complexity lower bounds is the area of theoretical Computer Science that studies the *intrinsic complexity* of computational problems. In other words, lower bound problems require us to show that a certain computational problem needs *at least* a certain amount of resources assuming a certain model of computation. Lower bound arguments are in general hard to make because each such argument is a sweeping statement about the (in)solvability of the problem in a certain time/space for *all* valid problem instances. For example, there are a number of well-known sorting methods such as Bubble Sort, Quick Sort etc. but the argument that no comparison-based sorting method can take less than $O(n \log n)$ time on a random-access machine is decidedly harder to make.

Lower bound problems arise in various contexts - there are a number of lower bound questions on circuit size/depth, many such questions on graphs (most of which are known to be *NP complete*) and many involving matrices. In particular, several matrix functions related to rank and its robustness under various transformations of the matrix have been extensively used in many complexity problems. Matrix related problems lend themselves to the use of elegant techniques from algebraic-geometry.

This work studies a number of matrix related lower bound problems arising in diverse areas. Further, it establishes the interconnections between them and uses techniques from algebraic geometry to explain the results more intuitively and to improve the known results. The basic geometric problem we seek to answer can be stated simply as: *Given a set of n orthants (an orthant is a generalization of a quadrant in 2D to m dimensions) in $\mathbb{R}^m$, what is the lowest dimensional homogeneous subspace of $\mathbb{R}^m$ that intersects these orthants.* Stated in linear algebraic terms, we ask: *Given a matrix $M \in \{-1,+1\}^{n\times m}$, what is the lowest rank matrix $M' \in \mathbb{R}^{n\times m}$ such that $sign(M'_{ij}) = M_{ij}$ for all i,j.* On one hand this is the common problem underlying many problems in diverse areas of Computer Science, and on the other hand it allows the use of algebraic-geometric concepts such as *Grassmann manifolds* and the topology of L_1, L_2 and L_∞ balls.

The thesis is organized as follows: We motivate the study of this geometric problem in Chapter 3 by mentioning its links with lower bound questions in areas such as Communication Complexity, Threshold Circuit Complexity, Learning Theory etc. We then present the original longstanding open problem in Communication Complexity and its recent solution by Jürgen Forster[1] in Chapter 4. Forster defines the problem as that of finding a low-dimension *hyperplane arrangement* that represents M and calls it a geometric *realization* of M. Forster uses the notion of a 'nice' realization - realization that has a non-trivial property that helps attain good bounds. We rephrase Forster's result and its proof in more intuitive terms and provide a justification for it. We also define a few nice properties of our own, one of which is a relaxation of Forster's property and the other helps us arrive at a new result. The study of these nice properties, their relationships, showing their existence and analysing what they buy us has been the central thread and the major contribution of this work. We define and relate these properties in Chapters 4 and 5 and show a new bound on the minimal dimension in Chapter 5. We also give an intuitive explanation of what our result means and why it is better. We pose a question about the containment of the L_1 ball in L_2 ball that stems from one of the 'nice' properties and try to answer it in Chapter 6. Finally, we identify the different directions in which one could delve deeper using this problem as a starting point in Chapter 7.

CHAPTER 2
MATHEMATICAL PRELIMINARIES

In this chapter we fix some notation for the rest of the thesis, and quote some relevant results from linear algebra and functional analysis that we will use later.

2.1 Matrices

$\mathbb{R}^{n\times m}$ is the set of all $n \times m$ matrices defined over the reals. In general if $\mathbb{F}$ denotes a field, then $\mathbb{F}^{n\times m}$ denotes the set of all $n \times m$ matrices over $\mathbb{F}$. In this work, we usually talk about matrices defined over the reals ($M \in \mathbb{R}^{n\times m}$) or ± 1 matrices ($M \in \{-1, 1\}^{n\times m}$). The entry in the ith row and jth column of M is denoted by M_{ij}. We use i to index the rows and j to index the columns of the matrix.

The rank of a matrix M is the number of linearly independent rows or columns in M. The product of two matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times p}$ is the matrix $C = A.B \in \mathbb{R}^{n\times p}$ where $C_{ik} = \sum_j A_{ij}B_{jk}, i \in \{1..n\}, j \in \{1..m\}, k \in \{1..p\}$.

The Kronecker product (or tensor product) of two matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{p\times q}$ is defined as:

$$A \bigotimes B = \begin{bmatrix} A_{1,1}B & A_{1,2}B & \dots & A_{1,n}B \\ A_{2,1}B & A_{2,2}B & \dots & A_{2,n}B \\ . & & & \\ . & & & \\ . & & & \\ A_{m,1}B & A_{m,2}B & \dots & A_{m,n}B \end{bmatrix} \in \mathbb{R}^{np\times mq}$$

such that $\text{rank}(A \bigotimes B) = \text{rank}(A).\text{rank}(B)$.

A *square matrix* is a $m \times m$ matrix. The *identity matrix* I is a square matrix such that $I_{ij} = 1$ if $i = j$ and 0 otherwise. The inverse of a square matrix A is the matrix A^{-1} such that

$$AA^{-1} = I$$

A square matrix A has an inverse *iff* $|A| \neq 0$. A matrix that has an inverse is called *non-singular* or invertible. The group of non-singular matrices $A \in \mathbb{R}^{m\times m}$ is denoted by $GL(m)$. The trace of a square matrix $A \in \mathbb{R}^{m\times m}$ is defined as

$$\text{trace}(A) = \sum_{i=1}^{m} A_{ii}$$

For $A \in \mathbb{R}^{m\times m}$, if there is a vector $\mathbf{x} \in \mathbb{R}^m \neq 0$ such that

$$A\mathbf{x} = \lambda\mathbf{x}$$

for some scalar λ, then λ is called the *eigenvalue* of A with the corresponding *eigenvector* $\mathbf{x}$.

The transpose of a matrix $A \in \mathbb{R}^{n\times m}$ is a matrix $A^{\intercal} \in \mathbb{R}^{m\times n}$ such that $A_{ij} = A^{\intercal}_{ji}$. In particular, the transpose of a column vector is a row vector. A matrix A is called *orthogonal* if $AA^{\intercal} = I$. A square matrix $A \in \mathbb{R}^{m\times m}$ is called *symmetric* if $A = A^{\intercal}$. A symmetric matrix is said to be *positive semi-definite*, denoted by $A \geq 0$, if all of its eigenvalues are non-negative or equivalently, if $\mathbf{x}^{\intercal}A\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^m$. A symmetric matrix is said to be *positive definite* if all of its eigenvalues are positive.

A special kind of matrix that we will encounter often is the *Hadamard* matrix. Hadamard matrices $H_n \in \mathbb{R}^{2^n\times 2^n}$ are examples of matrices with pairwise orthogonal rows and pairwise orthogonal columns. They are defined in the Sylvester form as

$$H_0 = 1, \quad H_{n+1} = \begin{pmatrix} H_n & H_n \\ H_n & -H_n \end{pmatrix}.$$

2.2 Inner Products and Norms

The *inner product* of two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^m$ is defined as

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{m} \mathbf{v}_i \mathbf{w}_i$$

The vectors $\mathbf{v}$,$\mathbf{w}$ are called *orthogonal* if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. For a vector $\mathbf{x} \in \mathbb{R}^m$, the vector norm L_p is defined as

$$||\mathbf{x}||_p = \left(\sum_{i=1}^{n} |\mathbf{x}_i|^p \right)^{1/p}$$

for $p = 1, 2, 3, \ldots$. The most commonly encountered norms are:

1. L_1 (or *Manhattan*) norm: $||\mathbf{x}||_1 = \sum_i |\mathbf{x}_i|$. It is used, for example, in computing the Hamming distance between two bit strings.
2. L_2 (or *Euclidean*) norm: $||\mathbf{x}||_2 = \sqrt{\sum_i \mathbf{x}_i^2}$. It is the most common measure of distance between two points.
3. L_∞ norm: $||\mathbf{x}||_\infty = \max_i |\mathbf{x}_i|$.

The vectors $\mathbf{v}$,$\mathbf{w}$ are called *orthonormal* if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ and $||\mathbf{v}||_2 = ||\mathbf{w}||_2 = 1$.

The operator norm of a matrix $A \in \mathbb{R}^{n \times m}$ is

$$||A|| = \sup_{\mathbf{x} \in \mathbb{R}^m} \frac{||A\mathbf{x}||_2}{||\mathbf{x}||_2}$$

For a matrix $A \in \{-1, 1\}^{n \times m}$, it is easy to see that $||A|| = \sqrt{m}$ if the rows of A are pairwise orthogonal and $||A|| = \sqrt{n}$ if the columns of A are pairwise orthogonal. It follows that $||H_n|| = 2^{\frac{n}{2}}$. Also if rank($A$)=1 then $||A|| = \sqrt{mn}$.

It is well known that for any matrix $A \in \mathbb{R}^{n \times m}$, $||A||^2 = ||A^\intercal A|| = ||AA^\intercal||$. For every symmetric matrix $A \in \mathbb{R}^{m \times m}$ there is an orthonormal basis $d_1, d_2, \ldots, d_m$ of $\mathbb{R}^m$ and there are scalars $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbb{R}^m$ (the eigenvalues of A) such that

$$A = \sum_{i=1}^{m} \lambda_i d_i d_i^\intercal$$

The signum function sign: $\mathbb{R} \to \mathbb{R}$, which we encounter often, is defined as

$$\text{sign}(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0 \end{cases}$$

2.3 Geometric Concepts

$\mathbb{R}^m$ is the m-dimensional real space - the set of all m-dimensional vectors defined over the reals. We use m and k for the dimension of a space. Thus a matrix $M \in \mathbb{R}^{n \times m}$ can be treated as a subspace of $\mathbb{R}^m$, i.e. a set of n vectors in $\mathbb{R}^m$. We use the matrix and subspace notion interchangeably.

The rank of a matrix M is the same as the dimension of the subspace denoted by M. A basis of a vector space V is defined as a subset of vectors in V that are linearly independent and span V, i.e. every vector in V can be written as a linear combination of the basis vectors. The number of basis vectors is equal to the dimension of V.

A *hyperplane* is a generalisation of a normal two-dimensional plane in three-dimensional space to its $(m-1)$-dimensional analogue in m-dimensional space. Each hyperplane separates the ambient space into two *half-spaces*. The terms hyperplane and half-space are used interchangeably in the literature. A hyperplane (or half-space) is called homogeneous if its boundary contains the origin.

An *orthant* is the set of all vectors that have the same signs. For example, in 2D we have 4 orthants (commonly called quadrants) and in 3D we have 8 orthants (commonly called octants). In general, there are 2^m orthants in m dimensions, each of which can be uniquely identified by a m-long sign vector.

The $(m-1)$-dimensional sphere (*unit ball* in m dimensions) is the set $\{\mathbf{x} \in \mathbb{R}^m \colon ||\mathbf{x}||_2 = 1\}$ and is denoted by S^{m-1}.

In general, $L_p = d$ ball in $\mathbb{R}^m$ is defined as the set $\{\mathbf{x} \in \mathbb{R}^m \colon ||\mathbf{x}||_p = d\}$. We also use L_p^m to mean the space $\mathbb{R}^m$ under L_p norm.

An *arrangement of Euclidean half-spaces* that *represents* $M \in \mathbb{R}^{n \times m}$ consists of points $u_i \in \mathbb{R}^k$ for every row $i \in \{1..n\}$ of M and of half-spaces $H_j \subseteq \mathbb{R}^k$ for every column $j \in \{1..m\}$ of M such that u_i lies in the half-space H_j if and only if sign(M_{ij})=+1. k is called the *dimension* of the arrangment.

A homogeneous half-space H is of the form

$$H = \{\mathbf{z} \in \mathbb{R}^k \colon \langle \mathbf{z}, \mathbf{v} \rangle \geq 0\}$$

for some vector $\mathbf{v}$ *normal* to the boundary of the half-space. Then we define an arrangement of homogeneous Euclidean half-spaces as:

Definition. An arrangement of homogeneous Euclidean half-spaces representing a matrix $M \in \mathbb{R}^{n \times m}$ with no zero entries is given by the collection of vectors $\mathbf{u}_i, \mathbf{v}_j \in \mathbb{R}^k$ for $i \in \{1..n\}, j \in \{1..m\}$ such that $\text{sign}(M_{ij}) = \text{sign}\langle \mathbf{u}_i, \mathbf{v}_j \rangle$.

Here each vector $\mathbf{v}_j$ is normal to and uniquely defines the homogeneous half-space H_j.

CHAPTER 3
APPLICATION IN COMPUTER SCIENCE

In this chapter we motivate the study of the geometric problem stated in the introduction - namely, determining the minimal dimension of the subspace that intersects a given set of orthants - by stating its applications in diverse areas of Computer Science. This chapter should also serve to hold the interest of the reader in this work who values a problem by its usefulness in the real world, and to justify the study of this problem to those who don't find its mathematical elegance adequately appealing.

3.1 Communication Complexity

Consider the two-party communication complexity problem - two parties A and B want to compute $f(x, y)$, where f is a distributed function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Both have unbounded computational power. A has input string x and B has input string y. A two-way probabilistic communication protocol is a probabilistic algorithm used by A and B to compute f by exchanging messages. A and B take turns probabilistically choosing a message to send to the other party according to the protocol. We say that a protocol computes the distributed function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with unbounded error if for all inputs $(x, y) \in \{0,1\}^n \times \{0,1\}^n$ the correct output is calculated with a probability greater than $\frac{1}{2}$. The complexity of a communication protocol is $\lceil \log_2 N \rceil$, where N is the number of distinct message sequences that can occur in computations that follow the protocol. The communication complexity $\tilde{C}_f$ of a distributed function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is the smallest complexity that a communication protocol for f can have.

To establish the connection between the problem of estimating the minimal dimension of ± 1 matrices and the communication complexity problem, we note that a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ induces a matrix $M \in \{-1,1\}^{2^n \times 2^n}$ such that $M_{ij} = 2f(x, y) - 1$ where x and y are the binary representations of integers i and j respectively. Paturi and Simon [2] showed that the minimal dimension $d(f)$ of the matrix induced by the function

f is related to its communication complexity $\tilde{C}_f$ as:

$$\lceil \log d(f) \rceil \leq \tilde{C}_f \leq \lceil \log d(f) \rceil + 1$$

Alon, Frankl and Rödl [3] showed that for *almost all* $n \times n$ matrices with ± 1 entries, the minimal dimension is at least $\Omega(n)$ implying that for most Boolean functions the unbounded error probabilistic communication complexity is asymptotically as large as it can be (that is, linear in the length of the input). However, proving even a superlogarithmic lower bound on the minimal dimension of an explicit matrix remained a difficult open question. Jürgen Forster [1] solved this long standing open question by showing a general lower bound on the minimal dimension of a matrix in terms of its operator norm. As a corollary, he derived a lower bound of $\sqrt{n}$ on the minimal dimension of an $n \times n$ Hadamard matrix. Later, Forster, Krause *et al* [4] generalized Forster's bound to real matrices (hence, to functions defined over the reals) with non-zero entries.

One of the motivations behind studying this problem was to replace the dependency on the operator norm of M by some other property that captures the orthant structure represented by the rows of M, because intuitively the minimal dimension should depend on the set of orthants represented by M.

3.2 Threshold Circuit Complexity

A linear threshold gate is a boolean gate that outputs a 1 if the weighted sum of its inputs is greater than a threshold t and 0 otherwise. Forster, Krause *et al* [4] proved a lower bound on the size of depth-2 threshold circuits where the top gate is a linear threshold gate with unrestricted weights and the bottom level gates are linear threshold gates with restricted integer weights. The result implies that (loosely speaking) boolean functions with 'high' minimal dimension cannot be computed by 'small' depth-2 circuits having the above restrictions. Meera Sitharam (personal communication) observed that the restriction on the weights of the bottom level gates is not required to prove the result.

3.3 Maximal Margin Classifiers

Computational learning theory is the study of efficient algorithms that improve their performance based on past experience. A class of problems in learning theory is classfication problems. The learning algorithm is 'trained' on some training data which is a set of data instances and their classes or labels. After the algorithm is trained, it is then presented with unlabeled data and is required to label (or classify) it accurately. The most important test of the usefulness of a learning algorithm is its performance in experiments on real-world data. Recently there has been a lot of interest in maximal margin classifiers which have shown excellent empirical performance. Maximal margin classifiers are learning algorithms that compute the hyperplane that separates a sample with the largest margin and use this hyperplane to classify new instances. Often the set of instances is mapped to some possibly high dimension before the hyperplane with the maximal margin is computed. If the norms of the instances are bounded and a hyperplane with a large margin can be found, then that gives a bound on the Vapnik-Chervonenkis (VC) dimension of the classfier. The VC-dimension is a measure of the power of the learning algorithm; a small VC-dimension means that a class can be learned with a small sample size. Knowing only the VC-dimension and the training error, it is possible to estimate the error on the future data.

The success of maximal margin classifiers raises the question which concept classes can be represented by an arrangement of Euclidean half spaces with a large margin. We say that the matrix $M \in \mathbb{R}^{n \times m}$ with no zero entries can be realized by an arrangement of homogeneous half spaces with margin γ if there are vectors $\mathbf{u}_i, \mathbf{v}_j \in S^{k-1}$ (where k can be arbitrarily large) such that $\text{sign}(M_{ij}) = \text{sign}\langle \mathbf{u}_i, \mathbf{v}_j \rangle$ and $|\langle \mathbf{u}_i, \mathbf{v}_j \rangle| \geq \gamma$ for all i, j. Interpreting $\mathbf{v}_j$ as the normal to the boundary of the jth half space, $\text{sign}(M_{ij}) = \text{sign}\langle \mathbf{u}_i, \mathbf{v}_j \rangle$ means that the vector $\mathbf{u}_i$ lies in the halfspace containing $\mathbf{v}_j$ if and only if M_{ij} is positive. The requirement $|\langle \mathbf{u}_i, \mathbf{v}_j \rangle| \geq \gamma$ means that the point $\mathbf{u}_i$ has distance at least γ from the boundary of the half space.

Forster [1] shows that the lower bound on the minimal dimension $d(M)$ is the inverse of the upper bound on the maximal margin γ. Therefore, a tighter lower bound on the minimal dimension will directly result in a tighter upper bound on the maximal margin.

3.4 Geometric Embeddings

Geometric embeddings lie at the intersection of analysis and geometry and have attracted much attention in recent times due to their algorithmic applications in data mining and information extraction from massive multidimensional data sets. Specifically, the geometric embeddings that lead to *dimensionality reduction* (which we will define shortly) are the ones of great interest for efficiently solving problems in data mining, protein matching etc.

A *metric space* M (also called a *metric*) is a pair (X, D), where X is a set of *points* and $D : X \times X \to [0, \infty)$ is a *distance function* satisfying the following properties for $p, q, r \in X$:

- $D(p, q) = 0$ iff $p = q$
- $D(p, q) = D(q, p)$
- $D(p, q) + D(q, r) \geq D(p, r)$

Formally, geometric embeddings are mappings $f : P_A \to P_B$ such that

- P_A is a set of points in the original space (which is usually a metric space and therefore has a distance function $D(\cdot, \cdot)$)
- P_B is a set of points in the host space (which is usually a normed space L_s^d)
- for any $p, q \in P_A$ we have

  $$1/c \cdot D(p, q) \leq ||f(p) - f(q)||_s \leq D(p, q)$$

  for a certain parameter c called distortion.

Any embedding $f : A \to B$ can be classified based on the types of spaces A and B. B is usually a normed space $L_{p'}^{d'}$. A can be any of the following:

1. A is a finite metric $M = (X, D)$ induced by graphs, e.g. obtained by computing all-pairs shortest paths. The main applications of such embeddings are approximation algorithms for optimization problems on graphs. Other applications include proximity-preserving labeling and proving hardness of approximation.
2. A is a subset of L_p^d. Here the intent is to map the *whole* normed space into the host space as opposed to a finite set of points. This is useful in cases when all the points in the set are not known in advance, e.g. when the points constitute a solution to an NP-hard problem, or when they are given online by the user. Here we have two scenarios:
   (a) $d' << d$ (and most often $p = p'$): such embeddings result in *dimensionality reduction*. The goal of dimensionality reduction is to map a set of points in a high-dimensional space to a space with low dimension while (approximately)

preserving important characteristics of the pointset. This leads to a speed up of algorithms whose running time depends on the dimension.

   (b) $p \neq p'$ (and most often $d' >> d$): Such embeddings allow us to switch from "difficult" norms (e.g. L_2) to "easier" norms (e.g. L_∞).

3. A is a special metric, usually more general than a norm. Examples of such metrics are the *edit metric* (defining similarity between strings of characters), *Hausdorff metric* (defining similarity between sets of points) and *Hamming metric* (defining the Hamming distance between a set of points over $\mathbb{F}_2$).

Forster formulated the problem of estimating the minimal dimension k of a matrix $M \in \mathbb{R}^{n \times m}$ as a problem to *realize* the matrix by a linear arrangement of homogeneous hyperplanes in k dimensions. Now treating each row of M as a vector in $\mathbb{R}^m$, we notice that the hyperplane arrangement induces a Hamming metric on the set of n vectors (each vector $M_i, i \in \{1..n\}$ corresponds to a point $\mathbf{p}_i \in \{-1, 1\}^m$ such that $\mathbf{p}_i = (\text{sign}\langle M_i, \mathbf{v}_j\rangle : j \in \{1..m\})$ where $\mathbf{v}_j$ is the normal to the jth hyperplane).

Thus our original problem can now be cast as a problem of embedding an m-dimensional Hamming metric induced by the linear hyperplane arrangement into a Hamming metric in k dimensions induced by the same hyperplane arrangement without any distortion. This problem has direct applications in protein matching [5].

CHAPTER 4
REALIZATIONS AND REALIZABILITY

In this chapter, we discuss the known results on the problem outlined in the introduction and rephrase some of them in more intuitive terms using techniques that lead to improved bounds. We start by defining the concept of realization of a matrix M, first in terms of half-spaces and later redefining it in terms of subspaces that intersect the same set of orthants as defined by the rows of M.

4.1 Half Space Realization

The longstanding open question on the minimal dimension of matrices was first posed by Paturi and Simon[2]. Forster [1] who solved the problem by showing a lower bound on the minimal dimension of M in terms of the operator norm of M defines the problem as realizing M in terms of a linear arrangement of homogeneous halfspaces. We restate Forster's definition of realization:

Definition. A matrix $M \in \mathbb{R}^{n\times m}$ with no zero entries can be *realized* by a k-dimensional linear arrangement of homogeneous half spaces if there are vectors $\mathbf{u}_i, \mathbf{v}_j \in \mathbb{R}^k$ for $i \in \{1..n\}, j \in \{1..m\}$ such that $\text{sign}(M_{ij}) = \text{sign}\langle \mathbf{u}_i, \mathbf{v}_j\rangle$.

The vector $\mathbf{v}_j$ is interpreted as a normal to the boundary of the homogenous half-space and hence uniquely defines it. The vector $\mathbf{u}_i$ defines a point that lies in the half space defined by $\mathbf{v}_j$ if $\text{sign}\langle \mathbf{u}_i, \mathbf{v}_j\rangle = +1$ and in the opposite half space if $\text{sign}\langle \mathbf{u}_i, \mathbf{v}_j\rangle = -1$.

4.1.1 Known Bounds

The following lower bound on the minimal dimension was proved recently by Forster [1]:

Theorem 4.1 (Forster[1]) *If a matrix $M \in \{-1,1\}^{n\times m}$ can be realized by an arrangement of homogeneous half spaces in $\mathbb{R}^k$, then*

$$k \geq \frac{\sqrt{mn}}{||M||} \tag{4.1}$$

Proof idea. Forster uses a bridging quantity to prove his result. He shows an upper bound on the bridging quantity that holds for any realization of the given matrix M. He also shows a lower bound on the bridging quantity that holds for only the 'nice' realizations of M. He then shows that given any realization, he can iteratively find one that is nice and therefore satisfies the lower bound on the bridging quantity. While rephrasing Forster's proof, we also define a few nice properties of our realizations, namely, the almost-spherical property which we show to be a relaxation of Forster's property and the projection-type nice property which helps us change Forster's upper bound on the bridging quantity. The study of these nice properties of realizations, relating them, showing their existence and finding out if they buy us better results is the central theme of this work.

Proof (by Forster). Assume that there are vectors $\mathbf{u}_i, \mathbf{v}_j \in \mathbb{R}^k$ such that $\operatorname{sign}\langle \mathbf{u}_i, \mathbf{v}_j \rangle = M_{ij}$. Forster observed that normalizing $\mathbf{u}_i$ and $\mathbf{v}_j$ will not affect the sign of the inner product $\langle \mathbf{u}_i, \mathbf{v}_j \rangle$. Therefore we may assume that $\mathbf{u}_i, \mathbf{v}_j \in S^{k-1}$. He also observed that for any nonsingular linear transformation $A \in \mathrm{GL}(k)$, replacing $\mathbf{u}_i, \mathbf{v}_j$ by $(A^\intercal)^{-1}\mathbf{u}_i, A\mathbf{v}_j$ also does not affect the sign of the inner product. He then shows that there exists a linear mapping that transforms a given set of $\mathbf{v}_j$'s into a set of "nice" $\mathbf{v}_j$'s such that $\sum_j \mathbf{v}_j {\mathbf{v}_j}^\intercal = \frac{m}{k} I_k$, i.e. the vectors $\mathbf{v}_j$ are nicely balanced in this sense. For such nice $\mathbf{v}_j$'s it holds that,

$$\sum_j |\langle \mathbf{u}_i, \mathbf{v}_j \rangle| \geq \sum_j \langle \mathbf{u}_i, \mathbf{v}_j \rangle^2 = {\mathbf{u}_i}^\intercal \left(\sum_j \mathbf{v}_j {\mathbf{v}_j}^\intercal \right) \mathbf{u}_i = \frac{m}{k} \tag{4.2}$$

Inequality (4.2) means that for all j, the absolute values of the inner products $\langle \mathbf{u}_i, \mathbf{v}_j \rangle$ are on the average at least $\frac{1}{k}$, i.e. the vectors $\mathbf{u}_i$ cannot lie arbitrarily close to the homogeneous hyperplane with normal $\mathbf{v}_j$.

Forster gives a corresponding upper bound in terms of the operator norm $||M||$ of the matrix on the absolute values of the scalar products $\langle \mathbf{u}_i, \mathbf{v}_j \rangle$ as:

$$\sum_i \left(\sum_j |\langle \mathbf{u}_i, \mathbf{v}_j \rangle| \right)^2 \leq m||M||^2 \tag{4.3}$$

From (4.2) and (4.3), it follows that

$$n\left(\frac{m}{k}\right)^2 \leq \sum_i \left(\sum_j |\langle \mathbf{u}_i, \mathbf{v}_j \rangle|\right)^2 \leq m||M||^2 \tag{4.4}$$

The lower bound directly follows. ▮

4.1.2 Generalized Lower Bound

Forster, Krause, Lokam *et al* [4] generalized the lower bound result obtained in [1] to real matrices. This result widened the applicability of Forster's bound from boolean functions to functions defined over the reals. Also the generalized result gives accurate bounds for orthogonal matrices that have a small norm which the previous result fails to achieve. The generalized lower bound result can be stated as:

Theorem 4.2 *Let $M \in \mathbb{R}^{n\times m}$ be a matrix with no zero entries. Then the lowest dimension k in which M can be realized by a linear arrangement of homogeneous halfspaces is given by:*

$$k \geq \frac{\sqrt{mn}}{||M||} \min_{i,j} |M_{ij}| \tag{4.5}$$

Proof The proof is along the lines of Forster's original result. It assumes that there are vectors $\mathbf{u}_i, \mathbf{v}_j \in \mathbb{R}^k$ such that $\text{sign}(M_{ij}) = \text{sign}\langle \mathbf{u}_i, \mathbf{v}_j \rangle$ for all i, j. Using Forster's argument, it is assumed that a given linear arrangement $\mathbf{u}_i, \mathbf{v}_j$ can be normalized such that $\mathbf{u}_i, \mathbf{v}_j \in S^{k-1}$ and

$$\sum_j \mathbf{v}_j {\mathbf{v}_j}^{\mathsf{T}} = \frac{m}{k} I_k$$

Let $\min_{i,j} |M_{ij}| = M_{min}$. Then for all j, it holds that

$$\sum_j M_{ij}\langle \mathbf{u}_i, \mathbf{v}_j \rangle = M_{min} \sum_j \frac{M_{ij}}{M_{min}} \langle \mathbf{u}_i, \mathbf{v}_j \rangle \geq M_{min} \sum_j \langle \mathbf{u}_i, \mathbf{v}_j \rangle^2 = M_{min}.{\mathbf{u}_i}^{\mathsf{T}} \left(\sum_j \mathbf{v}_j {\mathbf{v}_j}^{\mathsf{T}}\right) \mathbf{u}_i$$

or,

$$\sum_j M_{ij}\langle \mathbf{u}_i, \mathbf{v}_j \rangle \geq M_{min} \frac{m}{k} \tag{4.6}$$

[4] also proves the same upper bound on the new bridging quantity. Specifically that

$$\sum_i \left(\sum_j M_{ij}\langle \mathbf{u}_i, \mathbf{v}_j \rangle\right)^2 \leq m||M||^2 \tag{4.7}$$

It is to be noted that the upper bound does not depend upon the minimum entry in M. From (4.6) and (4.7), we get

$$M_{min}^2 n \left(\frac{m}{k}\right)^2 \leq \sum_i \left(\sum_j M_{ij}\langle \mathbf{u}_i, \mathbf{v}_j\rangle\right)^2 \leq m||M||^2 \tag{4.8}$$

The generalized lower bound result follows. ▮

4.2 A New Notion of Realization

We characterize a realization of the matrix $M \in \mathbb{R}^{n\times m}$ in terms of the k-dimensional subspace of $\mathbb{R}^m$ that cuts through the set of orthants defined by the signs of the rows in M. The motivation behind this new notion of realization is to cast Forster's result in more simplified terms as an answer to the questions posed by us in the introduction. Also, this notion of realization leads us to some interesting properties of subspaces, and to an improved result.

Definition. A matrix $M \in \mathbb{R}^{n\times m}$ with no zero entries can be *realized* by a k-dimensional subspace B if there exist vectors $\mathbf{w}_i = \mathbf{a}_i B$, where $\mathbf{a}_i \in \mathbb{R}^k$, such that $\text{sign}(\mathbf{w}_i) = \text{sign}(M_i)$ for all i.

4.2.1 Constructing B from Forster's Realization

To establish an equivalence between Forster's definition of realization and our notion of realization, we first give a construction to obtain the B matrix and the $\mathbf{w}_i$'s from Forster's $\mathbf{u}_i$ and $\mathbf{v}_j$ vectors.

Lemma 4.3 *Given $M \in \mathbb{R}^{n\times m}$ and $\mathbf{u}_i, \mathbf{v}_j \in S^{k-1}$ such that sign$\langle \mathbf{u}_i, \mathbf{v}_j\rangle$ = sign(M_{ij}) for all i, j, construct the $k \times m$ matrix B such that the $\mathbf{v}_j$'s are the columns of B and let $\mathbf{w}_i = \sqrt{k/m} \cdot \mathbf{u}_i B$. Then the rows of the matrix B form a basis for the k-dimensional subspace that realizes M.*

Proof We note that vectors $\mathbf{w}_i = \mathbf{u}_i B$ are actually the vectors $\mathbf{u}_i$ expressed in the basis B. In other words, vectors $\mathbf{w}_i$ are nothing but $\mathbf{u}_i$ living in the ambient m-dimensional space. The scaling of $\mathbf{w}_i$ by a factor of $\sqrt{k/m}$ is needed later in rephrasing Forster's proof where the $\mathbf{w}_i$'s are required to be unit vectors. Now, the jth co-ordinate of $\mathbf{w}_i$, $\mathbf{w}_i(j) = \sqrt{k/m}\langle \mathbf{u}_i, B_j\rangle$. Since the columns B_j of the matrix are actually the vectors $\mathbf{v}_j$ by construction, therefore we have that

$$\operatorname{sign}(\mathbf{w}_i(j)) = \operatorname{sign}\left(\sqrt{\frac{k}{m}}\langle \mathbf{u}_i, B_j\rangle\right) = \operatorname{sign}\left(\sqrt{\frac{k}{m}}\langle \mathbf{u}_i, \mathbf{v}_j\rangle\right) = \operatorname{sign}(M_{ij})$$

for all i, j. Hence this B satisfies the definition of a subspace realizing the matrix M. ∎
In the sequel, we will talk about realizations in terms of vectors $\mathbf{w}_i$ and the subspace B rather than vectors $\mathbf{u}_i$ and $\mathbf{v}_j$.

In order to rephrase Forster's proof of the lower bound result, we first define few niceness properties of our realization subspace B and show their relationship. This also leads us to some interesting questions about the topology of L_1^m, L_2^m and L_∞^m balls which we will examine in chapter 6.

4.2.2 Niceness Properties of Realization Subspaces

Property 4.4 *A $k \times m$ matrix B is said to be "nice" if its rows form an orthogonal basis with the basis vectors all having L_2 norm equal to $\sqrt{m/k}$ and its columns all have a L_2 norm at most 1.*

Property 4.5 *A k-dimensional realization B of M is said to be* almost-spherical *if the one-norm of all unit vectors in B is at least $\sqrt{m/k}$.*

Property 4.6 *A k-dimensional realization B of M is a nice subspace if the infinity-norm of all unit vectors in the subspace is at most $\sqrt{k/m}$.*

After having defined the niceness properties of B, we show their equivalence.

Property 4.6 is stronger than Almost-spherical Property. We prove this in the following theorem:

Theorem 4.7 *For a subspace B of $\mathbb{R}^m$, B has property 4.6 implies B is almost-spherical.*

Proof For all vectors $\mathbf{w} \in B$, if $||\mathbf{w}||_2 = 1$ and $||\mathbf{w}||_\infty \leq \sqrt{k/m}$ then $\mathbf{w}$ needs to have at least m/k non-zero elements. Of these, the $\mathbf{w}$'s having minimum one-norm will have exactly m/k non-zeros. For such $\mathbf{w}$'s the condition $||\mathbf{w}||_2 = 1$ is satisfied only when all the non-zeros have the same value which is $\sqrt{k/m}$ and hence $||\mathbf{w}||_1 = \frac{m}{k}.\sqrt{k/m} = \sqrt{m/k}$. Therefore it follows that if $||\mathbf{w}||_2 = 1$ and $||\mathbf{w}||_\infty \leq \sqrt{k/m}$ then $||\mathbf{w}||_1 \geq \sqrt{m/k}$. Hence proved. ∎

Property 4.4 $\equiv$ property 4.6. We show the equivalence of the two properties by first showing that property 4.4 implies property 4.6, and then showing that property 4.6 implies property 4.4.

Lemma 4.8 *For a subspace B of $\mathbb{R}^m$, B has property 4.4 $\Rightarrow$ B has property 4.6.*

Proof If B has property 4.4, then there exists an orthogonal basis for B with the L_2 norm of all rows equal to $\sqrt{m/k}$ and the L_2 norm of all columns at most 1. Therefore, for all unit vectors $\mathbf{w} \in B$ constructed as linear combinations of the basis vectors ($\mathbf{w} = \mathbf{a}B$, where $||\mathbf{a}||_2 = \sqrt{k/m}$), it holds that $\mathbf{w}(j) = \langle \mathbf{a}, B_j \rangle \leq ||\mathbf{a}||_2 ||B_j||_2 = \sqrt{k/m}$. Therefore all unit-vectors in B have an infinity norm at most $\sqrt{k/m}$. This implies that the subspace B has property 4.6. ■

Lemma 4.9 *For a subspace B of $\mathbb{R}^m$, B has property 4.6 $\Rightarrow$ B has property 4.4.*

Proof Consider an orthonormal basis (i.e. a $k \times m$ matrix with orthonormal rows) for B. If all unit-norm vectors in B have an infinity norm at most $\sqrt{k/m}$, then the same holds true for the basis vectors as well. Let $\mathbf{a} \in \mathbb{R}^k$ be a unit vector. Now, the vector $\mathbf{a}B$ is a unit vector in B and hence also has the above property. In other words, $\langle \mathbf{a}, B_j \rangle \leq \sqrt{k/m}$ for $j \in \{1..m\}$. Also we have that $\langle \mathbf{a}, B_j \rangle \leq ||\mathbf{a}||_2 ||B_j||_2$. Since the property holds true for *all* $\mathbf{a} \in \mathbb{R}^k$, it also holds true for the case when $\langle \mathbf{a}, B_j \rangle = ||\mathbf{a}||_2 ||B_j||_2$. For such $\mathbf{a}$, it holds that $\langle \mathbf{a}, B_j \rangle = ||\mathbf{a}||_2 ||B_j||_2 \leq \sqrt{k/m}$. Since $||\mathbf{a}||_2 = 1$, it implies that $||B_j||_2 \leq \sqrt{k/m}$ for all j. Now if we scale the matrix B by the factor $\sqrt{m/k}$ then B becomes a $k \times m$ matrix having orthogonal rows whose L_2 norm is $\sqrt{m/k}$ and columns whose L_2 norm is at most 1. ■

Theorem 4.10 *For a k-dimensional subspace B of $\mathbb{R}^m$, property 4.6 is equivalent to property 4.4.*

Proof From lemma 4.8 and lemma 4.9, it follows that for a k-dimensional subspace B of $\mathbb{R}^m$, property 4.4 and property 4.6 are equivalent.

Corollary 4.11 *For a subspace B of $\mathbb{R}^m$, property 4.4 $\equiv$ almost-spherical property.*

4.2.3 Rephrasing Forster in Terms of B

We are now ready to rephrase Forster's proof of the lower bound result in terms of the B matrix.

Lemma 4.12 *Given an orthogonal basis B for the realization subspace of $M \in \{-1, 1\}^{n \times m}$ and vectors $\mathbf{w}_i = \sqrt{k/m} \cdot \mathbf{u}_i B, i \in \{1..n\}$ such that sign$(\mathbf{w}_i) = M_i$, then*

$$\sum_i \langle \mathbf{w}_i, M_i \rangle^2 = \sum_i \|\mathbf{w}_i\|_1^2 \leq k\|M\|^2 \tag{4.9}$$

Proof Forster proved the upper bound on his bridging quantity as

$$\sum_{i=1}^{n} \left(\sum_j |\langle \mathbf{u}_i, \mathbf{v}_j \rangle| \right)^2 \leq m||M||^2$$

Since the B matrix has the $\mathbf{v}_j$'s as its columns, therefore we have that $\langle \mathbf{u}_i, \mathbf{v}_j \rangle = \mathbf{w}_i(j)/\sqrt{k/m}$. Substituting this in the above inequality, we get

$$\sum_{i=1}^{n} \left(\sum_j \frac{|\mathbf{w}_i(j)|}{\sqrt{k/m}} \right)^2 \leq m||M||^2$$

which is equivalent to

$$\frac{m}{k} \sum_{i=1}^{n} ||\mathbf{w}_i||_1^2 \leq m||M||^2$$

Since sign$(\mathbf{w}_i)$ = sign(M_i), therefore $||\mathbf{w}_i||_1 = \langle \mathbf{w}_i, M_i \rangle$. Hence

$$\sum_{i=1}^{n} \langle \mathbf{w}_i, M_i \rangle^2 = \sum_{i=1}^{n} ||\mathbf{w}_i||_1^2 \leq k||M||^2 \quad \blacksquare$$

The following lemma corresponds to Forster's observation about the existence of a "nice" realization.

Lemma 4.13 *Given a $k \times m$ matrix B constructed from the $\mathbf{v}_j$'s, there exists a non-singular linear transformation $A \in GL(k)$, such that the matrix $B' = AB$ cuts through the same set of orthants and additionally has property 4.4.*

Proof Forster proves that given $\mathbf{u}_i, \mathbf{v}_j \in S^{k-1}$, there exists a non-singular linear transformation $A \in \mathrm{GL}(k)$ such that if $\mathbf{u'}_i = A^{-1}\mathbf{u}_i$ and $\mathbf{v'}_j = A\mathbf{v}_j$, then

$$\mathrm{sign}\langle \mathbf{u}_i, \mathbf{v}_j \rangle = \mathrm{sign}\langle \mathbf{u'}_i, \mathbf{v'}_j \rangle \tag{4.10}$$

and

$$\sum_j \mathbf{v'}_j \mathbf{v'}_j^{\mathsf{T}} = \frac{m}{k} I_k \tag{4.11}$$

Furthermore, since scaling does not affect the sign of the inner product, therefore $\mathbf{u'}_i, \mathbf{v'}_j$ can be normalized so that $\mathbf{u'}_i, \mathbf{v'}_j \in S^{k-1}$.

We observe here that since the matrix B has the $\mathbf{v}_j$'s as its columns, therefore the matrix $B' = AB$ has the $\mathbf{v'}_j$ as its columns. Also the matrix $\sum_j \mathbf{v'}_j \mathbf{v'}_j^{\intercal}$ is in fact $B'B'^{\intercal}$. Using (4.11), we get

$$B'B'^{\intercal} = \frac{m}{k} I_k$$

which implies that B' is an orthogonal matrix with each row having a 2-norm of $\sqrt{m/k}$ and all columns being unit-norm.

To show that B' realizes M, we construct $\mathbf{w'}_i = \mathbf{u'}_i B', i \in \{1..n\}$. Then from (4.10), it is obvious that $\text{sign}(\mathbf{w'}_i(j)) = \text{sign}\langle \mathbf{u'}_i, \mathbf{v'}_j \rangle = \text{sign}\langle \mathbf{u}_i, \mathbf{v}_j \rangle = \text{sign}(M_{ij})$. ▮

Remark Forster's nice realization has property 4.4 and is therefore almost-spherical.

After having established the existence of a "nice" realization subspace of M, we restate Forster's left inequality in the following lemma:

Lemma 4.14 *Given a "nice" B having property 4.4, construct unit vectors* $\mathbf{w}_i = \sqrt{k/m} \cdot \mathbf{u}_i B$ *as the linear combination of the rows of B. Then,*

$$n\frac{m}{k} \leq \sum_i ||\mathbf{w}_i||_1^2 = \sum_i \langle \mathbf{w}_i, M_i \rangle^2 \tag{4.12}$$

Proof In the previous section we showed that for a realization subspace B, property 4.4 implies B is almost-spherical. This means that for all vectors $\mathbf{w} \in B, ||\mathbf{w}||_2 = 1$ implies $||\mathbf{w}||_1 \geq \sqrt{m/k}$. Now since the vectors $\mathbf{w}_i$ are unit vectors, therefore

$$||\mathbf{w}_i||_1 \geq \sqrt{\frac{m}{k}}$$

or,

$$||\mathbf{w}_i||_1^2 \geq \frac{m}{k}$$

or,

$$\sum_i ||\mathbf{w}_i||_1^2 \geq n\frac{m}{k} \quad ▮$$

Theorem 4.15 *If a matrix $M \in \{-1,1\}^{n \times m}$ can be realized by a k-dimensional subspace, then*

$$k \geq \frac{\sqrt{mn}}{\|M\|} \tag{4.13}$$

Proof From (4.9) and (4.12), we get

$$n\frac{m}{k} \leq \sum_i \|\mathbf{w}_i\|_1^2 \leq k\|M\|^2 \tag{4.14}$$

which gives,

$$n\frac{m}{k} \leq k\|M\|^2$$

or,

$$k \geq \frac{\sqrt{mn}}{\|M\|} \quad \blacksquare$$

4.2.4 Rephrasing the Generalized Lower Bound in Terms of B

The proof of the generalized lower bound in terms of the B matrix is very similar to the proof of Forster's original result. The main difference is in the interpretation of the bridging quantity used in the two results. While both use

$$\sum_{i=1}^{n} \langle \mathbf{w}_i, M_i \rangle^2$$

as the bridging quantity, Forster's original result holds only for $M \in \{-1,+1\}^{n \times m}$ and therefore the bridging quantity equals $\sum_i ||\mathbf{w}_i||_1^2$ which is not the case for the generalized lower bound. So for the generalized proof, we keep the bridging quantity in its original form and prove the upper bound on it as

$$\sum_i \langle \mathbf{w}_i, M_i \rangle^2 \leq k||M||^2$$

Then we prove the lower bound on the bridging quantity which holds for almost-spherical realization subspaces. The lower bound is

$$\sum_i \langle \mathbf{w}_i, M_i \rangle^2 \geq n\frac{m}{k}(\min_{i,j}|M_{ij}|)^2$$

The proof for both these results is a direct adaptation of the generalized lower bound proof in [4] obtained by substituting $\langle \mathbf{u}_i, \mathbf{v}_j \rangle = \mathbf{w}_i/\sqrt{k/m}$. Combining the two inequalities, we get the generalized lower bound.

4.2.5 Intuitive Justification

The right inequality in (4.14) is true for all realization subspaces of M, whereas the left inequality is true for only the "nice" subspaces that have property (4.4). Thus both inequalities simultaneously hold only for the "nice" B's.

The reason for rephrasing Forster's inequalities in this manner is that the bridging quantity in the Forster's result varied with the dimension k of the realization which appears to be counter-intuitive. In our formulation, the $\mathbf{w}_i$'s are all unit-norm vectors and hence it ensures that the bridging quantity remains the same irrespective of the realization dimension. We also introduce k into the right inequality so that the right side quantity now increases with k instead of being independent of k as it was in Forster's formulation.

CHAPTER 5
NICE REALIZATIONS

The existence of nice realizations - realizations that have some non-trivial 'nice' properties - helps in obtaining better bounds if their existence can be proven. This technique was used by Forster to obtain his lower bound result. In this section, we define another 'nice' property of our realization and relate it with the ones defined in the previous chapter. We make a conjecture about the existence of subspaces that have both the 'nice' properties which if proven results in improved bounds.

5.1 New Definition of Niceness

Property 5.1 *A k-dimensional realization B of M is nice if for all i, the projection of the vector defined by row M_i on B lies in the same orthant as M_i.*

To motivate the introduction of this new notion of 'niceness', we show improved bounds for Forster's result and the generalized lower bound result assuming the existence of nice realizations having property 5.1.

Conjecture There exists a realization B of M, that is almost-spherical and also has property 5.1.

5.1.1 Improved Bounds for $M \in \{-1, +1\}^{n \times m}$

Lemma 5.2 *Given a "nice" realization B of $M \in \{-1, +1\}^{n \times m}$ that has property 5.1, let vectors $\mathbf{w}_i$ be unit vectors in the direction of M_i projected on B. Then,*

$$||\mathbf{w}_i||_1 = \sqrt{\sum_{l=1}^{k} \langle M_i, B_l \rangle^2}$$

where B_l is the lth row of B.

Proof For B to realize M, there should exist vectors $\mathbf{w}_i, i \in \{1..n\}$ such that $\text{sign}(\mathbf{w}_i) = M_i$. Since the subspace B has property (5.1), therefore the projection of M_i on B lies in the same orthant as M_i. Hence the choice of $\mathbf{w}_i$ as being a unit vector in the direction of M_i

projected on to B is perfectly valid.

The jth co-ordinate of vector $\mathbf{w}_i$ is

$$\mathbf{w}_i(j) = \frac{\sum_{l=1}^{k} \langle M_i, B_l \rangle . B_{lj}}{\sqrt{\sum_l \langle M_i, B_l \rangle^2}}$$

Now

$$||\mathbf{w}_i||_1 = \sum_j |\mathbf{w}_i(j)| = \langle \mathbf{w}_i, M_i \rangle = \sum_{j=1}^{m} \mathbf{w}_i(j) M_{ij}$$

or,

$$||\mathbf{w}_i||_1 = \sum_{j=1}^{m} \frac{\sum_l \langle M_i, B_l \rangle . B_{lj}}{\sqrt{\sum_l \langle M_i, B_l \rangle^2}} M_{ij}$$

or,

$$||\mathbf{w}_i||_1 = \frac{\sum_l \langle M_i, B_l \rangle^2}{\sqrt{\sum_l \langle M_i, B_l \rangle^2}} = \sqrt{\sum_l \langle M_i, B_l \rangle^2}$$

Hence proved. ∎

Lemma 5.3 *For a realization B of $M \in \{-1, +1\}^{n \times m}$ that has property 5.1,*

$$\sum_i ||\mathbf{w}_i||_1{}^2 = \sum_i \sum_{1 \le l \le k} \langle M_i, B_l \rangle^2 \le nk + \frac{k}{m} . ORT(M)$$

where

$$ORT(M) = 2 \sum_{j1, j2} \langle M_{j1}, M_{j2} \rangle$$

for all $j1 < j2$, such that columns M_{j1} and M_{j2} are non-orthogonal.

Proof

$$\begin{aligned}
\sum_i \sum_{1 \le l \le k} \langle M_i, B_j \rangle^2 &= \sum_i \sum_{1 \le l \le k} (\sum_j M_{ij} . B_{lj})^2 \\
&= \sum_i \sum_{1 \le l \le k} \sum_j \sum_{j'} M_{ij} B_{lj} M_{ij'} B_{lj'} \\
&= \sum_i \sum_j \sum_{j'} M_{ij} . M_{ij'} \sum_l B_{lj} . B_{lj'} \\
&= \sum_j \sum_{j'} \sum_i M_{ij} . M_{ij'} \sum_l B_{lj} . B_{lj'}
\end{aligned}$$

Now,

$$\sum_i M_{ij} . M_{ij'} = \sum_i M^{\intercal}{}_{ji} . M_{ij'} = M^{\intercal} M(j, j')$$

and

$$\sum_{l} B_{lj}.B_{lj'} = \sum_{l} B^{\intercal}{}_{jl}.B_{lj'} = B^{\intercal}B(j,j').$$

Therefore the expression reduces to

$$\sum_{j}\sum_{j'} M^{\intercal}M(j,j')B^{\intercal}B(j,j')$$

or

$$\sum_{j}(||M_j||||B_j||)^2 + 2.\sum\langle M_{j1}, M_{j2}\rangle\langle B_{j1}, B_{j2}\rangle$$

(for all $j1 < j2$, such that columns M_{j1} and M_{j2} are non-orthogonal)

Now, since $M \in \{-1,+1\}^{n\times m}$, therefore $||M_j||_2 = \sqrt{n}$.

Also we have that

$$\sum_{j} ||B_j||^2 = \sum_{l=1}^{k} ||B_l||^2 = k$$

and therefore,

$$\langle B_{j1}, B_{j2}\rangle \leq \frac{k}{m}$$

Using these observations, we get

$$\sum_{i}\sum_{1\leq l\leq k}\langle M_i, B_l\rangle^2 \leq nk + 2\frac{k}{m}\sum\langle M_{j1}, M_{j2}\rangle$$

(for all $j1 < j2$, such that columns M_{j1} and M_{j2} are non-orthogonal)

or

$$\sum_{i}\sum_{1\leq l\leq k}\langle M_i, B_l\rangle^2 \leq nk + \frac{k}{m}\text{ORT}(M)$$

■

Corollary 5.4 *If $M \in \{-1,+1\}^{n\times m}$ has orthogonal columns, then*

$$\sum_{i} ||\mathbf{w}_i||_1{}^2 = \sum_{j}(||M_j||_2||B_j||_2)^2 = nk$$

Remark Therefore when M is a Hadamard matrix of size m, then $||M_j||_2^2 = m$ and the expression evaluates to mk.

Theorem 5.5 *If a matrix* $M \in \{-1,+1\}^{n\times m}$ *can be realized by a* k*-dimensional subspace then*

$$k \geq \sqrt{\frac{mn}{n + \frac{ORT(M)}{m}}}$$

Proof This result requires the existence of a realization subspace that is almost-spherical and also has property 5.1, which we conjectured to be the case.

Since B is almost-spherical, there exist unit vectors $\mathbf{w}_i$ such that $\text{sign}(\mathbf{w}_i) = M_i$ and

$$\sum_{i=1}^{n} ||\mathbf{w}_i||_1^2 \geq n\frac{m}{k}, \text{ from lemma 4.14.}$$

Also, in lemma 5.3 we showed that if B has property 5.1, then

$$\sum_{i=1}^{n} ||\mathbf{w}_i||_1^2 \leq nk + \frac{k}{m}.\text{ORT}(M)$$

Therefore, for a realization subspace that is almost-spherical and also has property 5.1, it holds that

$$n\frac{m}{k} \leq \sum_{i=1}^{n} ||\mathbf{w}_i||_1^2 \leq nk + \frac{k}{m}.\text{ORT}(M)$$

This gives,

$$n\frac{m}{k} \leq k\left(n + \frac{ORT(M)}{m}\right)$$

This directly gives the stated bound for k. ■

Corollary 5.6 *For a Hadamard matrix* H_n *of size* $2^n \times 2^n$*, the minimal dimension of a realization subspace is bounded as*

$$k \geq \sqrt{\frac{2^n.2^n}{2^n + 0}} = 2^{n/2}$$

5.1.2 Improved Bounds for $M \in \mathbb{R}^{n\times m}$

Lemma 5.7 *Given a "nice" realization* B *of* $M \in \mathbb{R}^{n\times m}$ *that has property 5.1, let vectors* $\mathbf{w}_i$ *be unit vectors in the direction of* M_i *projected on* B*. Then,*

$$\sum_{j=1}^{m} M_{ij}\mathbf{w}_i(j) = \langle \mathbf{w}_i, M_i\rangle = \sqrt{\sum_{l=1}^{k}\langle M_i, B_l\rangle^2}$$

where B_l *is the* l*th row of* B*.*

Proof In this case too the choice of $\mathbf{w}_i$ as the unit vectors in the direction of M_i projected onto B is perfectly valid since B has property 5.1.

Again, the jth co-ordinate of vector $\mathbf{w}_i$ is

$$\mathbf{w}_i(j) = \frac{\sum_{l=1}^{k} \langle M_i, B_l \rangle . B_{lj}}{\sqrt{\sum_l \langle M_i, B_l \rangle^2}}$$

Now

$$\langle \mathbf{w}_i, M_i \rangle = \sum_{j=1}^{m} \mathbf{w}_i(j) M_{ij} = \sum_{j=1}^{m} \frac{\sum_l \langle M_i, B_l \rangle . B_{lj}}{\sqrt{\sum_l \langle M_i, B_l \rangle^2}} M_{ij}$$

or,

$$\langle \mathbf{w}_i, M_i \rangle = \sum_l \frac{\langle M_i, B_l \rangle \sum_j M_{ij} B_{jl}}{\sqrt{\sum_l \langle M_i, B_l \rangle^2}}$$

or,

$$\langle \mathbf{w}_i, M_i \rangle = \sum_l \frac{\langle M_i, B_l \rangle^2}{\sqrt{\sum_l \langle M_i, B_l \rangle^2}} = \sqrt{\sum_l \langle M_i, B_l \rangle^2}$$

Hence proved. ■

Lemma 5.8 *For a realization B of $M \in \mathbb{R}^{n \times m}$ that has property 5.1,*

$$\sum_{i=1}^{n} \langle \mathbf{w}_i, M_i \rangle^2 = \sum_{i=1}^{n} \sum_l \langle M_i, B_l \rangle^2 \leq \frac{k}{m} \left(\sum_j (||M_j||)^2 + 2.ORT(M) \right)$$

Proof

$$\begin{aligned}
\sum_i \sum_{1 \leq l \leq k} \langle M_i, B_j \rangle^2 &= \sum_{1 \leq l \leq k} \sum_i (\sum_j M_{ij}.B_{lj})^2 \\
&= \sum_i \sum_{1 \leq l \leq k} \sum_j \sum_{j'} M_{ij} B_{lj} M_{ij'} B_{lj'} \\
&= \sum_i \sum_j \sum_{j'} M_{ij}.M_{ij'} \sum_l B_{lj}.B_{lj'} \\
&= \sum_j \sum_{j'} \sum_i M_{ij}.M_{ij'} \sum_l B_{lj}.B_{lj'}
\end{aligned}$$

Now,

$$\sum_i M_{ij}.M_{ij'} = \sum_i M^{\mathsf{T}}{}_{ji}.M_{ij'} = M^{\mathsf{T}} M(j, j')$$

and

$$\sum_l B_{lj}.B_{lj'} = \sum_l B^{\mathsf{T}}{}_{jl}.B_{lj'} = B^{\mathsf{T}} B(j, j').$$

Therefore the expression reduces to

$$\sum_j \sum_{j'} M^\intercal M(j,j') B^\intercal B(j,j')$$

or

$$\sum_j (||M_j||_2 ||B_j||_2)^2 + 2.\sum \langle M_{j1}, M_{j2}\rangle \langle B_{j1}, B_{j2}\rangle$$

(for all $j1 < j2$, such that columns M_{j1} and M_{j2} are non-orthogonal).

Now it holds that

$$||B_j||^2 = \frac{k}{m}$$

and therefore,

$$\langle B_{j1}, B_{j2}\rangle \leq \frac{k}{m}$$

This gives,

$$\sum_{1\leq l\leq k} \sum_i \langle M_i, B_l\rangle^2 \leq \frac{k}{m}\left(\sum_j ||M_j||_2^2 + ORT(M)\right) \quad \blacksquare$$

Corollary 5.9 *If $M \in \mathbb{R}^{n\times m}$ has orthogonal columns each having norm d, then*

$$\sum_i \langle \mathbf{w}_i, M_i\rangle^2 = \frac{k}{m}\sum_j ||M_j||_2^2 = d^2 k$$

Theorem 5.10 *If a matrix $M \in \mathbb{R}^{n\times m}$ can be realized by a k-dimensional subspace, then*

$$k \geq \sqrt{\frac{m^2 n}{\sum_j ||M_j||_2^2 + ORT(M)}} \min_{i,j} |M_{ij}|$$

where

$$ORT(M) = 2\sum_{j1,j2} \langle M_{j1}, M_{j2}\rangle$$

for all $j1 < j2$, such that columns M_{j1} and M_{j2} are non-orthogonal.

Proof From (4.6) and (5.8), we get

$$M_{min}^2 n \frac{m}{k} \leq \sum_{i=1}^{n} \langle \mathbf{w}_i, M_i\rangle^2 \leq \frac{k}{m}\left(\sum_j ||M_j||_2^2 + ORT(M)\right)$$

This is also based upon our conjecture that there exists a realization subspace that is almost-spherical and also has property 5.1 (since the left inequality holds for almost-spherical subspaces whereas the right inequality holds for realizations having property 5.1.

This gives,

$$M_{min}^2 n \frac{m}{k} \leq \frac{k}{m} \left(\sum_j ||M_j||_2^2 + ORT(M) \right)$$

or,

$$k^2 \geq \frac{M_{min}^2 m^2 n}{\sum_j ||M_j||_2^2 + ORT(M)}$$

The stated bound on k directly follows. $\blacksquare$

Thus we have shown that the existence of nice realizations gives us a new bound on the minimal dimension.

5.1.3 Why Is This Bound Better?

The norm of the matrix M that appears in the denominator of Forster's bound is defined as

$$||M|| = \sup_{\mathbf{x} \in \mathbb{R}^m} \frac{||M\mathbf{x}||_2}{||\mathbf{x}||_2}$$

Now if $\mathbf{x} = (1, 1, 1, \ldots, 1)$ be one of the vectors that causes the supremum to be attained, then we have

$$||M|| = \frac{||M\mathbf{x}||_2}{||\mathbf{x}||_2} = \frac{\sqrt{\sum_i (\sum_j M_{ij})^2}}{\sqrt{m}} = \frac{\sqrt{\sum_j ||M_j||_2^2 + ORT(M)}}{\sqrt{m}}$$

which is the denominator of our new bound (note that $\sum_j ||M_j||_2^2 = mn$ for the case when $M \in \{-1, 1\}^{n \times m}$). It therefore follows that our bound is as good as Forster's when the all-ones vector is one of the supremizing vectors (such as for Hadamard matrices) and in all other cases our bound is better because our denominator is lesser than $||M||$.

5.2 Nice Properties of Specific Constructions

In this section we show that the constructions given by Forster in his doctoral dissertation[6] and by Belcher *et al* in [7] have property 5.1.

5.2.1 Forster's Construction

Forster [6] gave the construction of a rank 3 matrix M' that had the same sign pattern as H_2. He used the construction to give an upper bound of $3^{\lceil \frac{n}{2} \rceil}$ on the realization dimension.

For the $2^n \times 2^n$ Hadamard matrix, the matrix M'_{2^n} constructed as:

$$M'_{2^n} = M'^{\bigotimes \frac{n}{2}} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 5 & -1 \\ 1 & 5 & -1 & -1 \\ 1 & -1 & -1 & 5 \end{pmatrix}^{\bigotimes \frac{n}{2}} \tag{5.1}$$

is of rank $3^{\frac{n}{2}}$.

5.2.2 Forster's construction has property 5.1

To show that Forster's construction has property 5.1, we will use an orthonormal basis B for M' such that the projection of M_i on B lies in the same orthant as M_i.

Lemma 5.11 *The rows of the matrix B given by*

$$\begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{pmatrix}$$

form an orthonormal basis for M'.

Using this basis, we show by induction that Forster's construction has property 5.1.

Lemma 5.12 *Base Case: For $M = H_{2^2}$, $\mathbf{w}_i = M_i|_B$ lies in same orthant as M_i.*

Proof Given,

$$B = \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{pmatrix}; M = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

Now, $\mathbf{w}_i = M_i|_B = \sum_j \langle M_i, B_j \rangle . B_j$. Therefore,

$$\begin{aligned} \mathbf{w}_1 &= (1 \quad 1 \quad 1 \quad 1) \\ \mathbf{w}_2 &= (0 \quad -2/3 \quad 4/3 \quad -2/3) \\ \mathbf{w}_3 &= (0 \quad 4/3 \quad -2/3 \quad -2/3) \\ \mathbf{w}_4 &= (0 \quad -2/3 \quad -2/3 \quad 4/3) \end{aligned}$$

It is obvious that $\mathbf{w}_i$ lies in the same orthant as M_i. ∎

Lemma 5.13 *Induction Step: If for* $M^k = H_{2^k}, \mathbf{w}_i{}^k = \sum_j \langle M_i{}^k, B_j{}^k\rangle.B_j{}^k$ *lies in the same orthant as* $M_i{}^k$*, then for* $M^{k+1} = M^k \bigotimes M, \mathbf{w}_i{}^{k+1} = \sum_j \langle M_i{}^{k+1}, B_j{}^{k+1}\rangle.B_j{}^{k+1}$ *lies in the same orthant as* $M_i{}^{k+1}$*.*

Proof The ith row of the matrix M^{k+1}, $M_i{}^{k+1}$ is either $M_i{}^k \bigotimes M_1$ or $M_i{}^k \bigotimes M_2$ or so on. Consider the case when $M_i{}^{k+1} = M_i{}^k \bigotimes M_1$. Then,

$$\mathbf{w}_i{}^{k+1} = \sum_j \langle M_i{}^{k+1}, B_j{}^{k+1}\rangle.B_j{}^{k+1}$$

or,

$$\mathbf{w}_i{}^{k+1} = \sum_j \sum_l \langle M_i{}^k \bigotimes M_1, B_j{}^k \bigotimes B_l\rangle.B_j{}^k \bigotimes B_l$$

or,

$$\mathbf{w}_i{}^{k+1} = \sum_j \sum_l \langle M_i{}^k, B_j{}^k\rangle\langle M_1, B_l\rangle.B_j{}^k \bigotimes B_l$$

or,

$$\mathbf{w}_i{}^{k+1} = \sum_j \langle M_i{}^k, B_j{}^k\rangle.B_j{}^k \bigotimes \sum_l \langle M_1, B_l\rangle.B_l$$

or,

$$\mathbf{w}_i{}^{k+1} = \mathbf{w}_i{}^k \bigotimes \mathbf{w}_1$$

Since tensor product preserves sign patterns, therefore we can say that if $\mathbf{w}_i{}^k$ has the same sign as $M_i{}^k$ and if $\mathbf{w}_1$ has the same signs as M_1, then $\mathbf{w}_i{}^{k+1} = \mathbf{w}_i{}^k \bigotimes \mathbf{w}_1$ will have the same signs as $M_i{}^{k+1} = M_i{}^k \bigotimes M_1$.

Similar argument can made about the cases when $M_i{}^{k+1}$ is either $M_i{}^k \bigotimes M_2$ or $M_i{}^k \bigotimes M_3$ or $M_i{}^k \bigotimes M_4$.

Theorem 5.14 *For all* n*, Forster's realization of* $M^n = H_{2^n}$ *has property 5.1.*

Proof The result follows from the base case (lemma 5.12) and the induction step (lemma 5.13).

5.2.3 Belcher-Hicks-Sitharam Construction

This construction gives an upper bound of $3/4 \times 2^p$ on the realization dimension of H_{2^p}. The M' that realizes H_{2^2} is given by:

$$\begin{pmatrix} 1 & 1 & 1 & 3 \\ 1 & -3 & 1 & -1 \\ 1 & 1 & -3 & -1 \\ 3 & -1 & -1 & 1 \end{pmatrix}$$

and is of rank 3. The $M_{2^{k+1}}$ is constructed from M_{2^k} as follows: The first 2^k rows of $M_{2^{k+1}}$ are constructed as $(M_{2^k}(i)\ M_{2^k}(i)), 1 \leq i \leq 2^k$. The next 2^k rows are constructed as $(M_{2^k}(i)\ -M_{2^k}(i)), 2^k + 1 \leq i \leq 2^{k+1}$.

5.2.4 Belcher-Hicks-Sitharam construction has property 5.1

Again to show that this construction has property 5.1, we will use an orthonormal basis B for M' such that the projection of M_i on B lies in the same orthant as M_i.

Lemma 5.15 *The rows of the matrix B given by*

$$\begin{pmatrix} -\sqrt{3}/6 & -\sqrt{3}/6 & -\sqrt{3}/6 & -\sqrt{3}/2 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} & 0 \end{pmatrix}$$

form an orthonormal basis for M'.

Using this basis, we show by induction that this construction has property 5.1.

Lemma 5.16 *Base Case: For $M = H_{2^2}$, $\mathbf{w}_i = M_i|_B$ lies in same orthant as M_i.*

Proof Given,

$$B = \begin{pmatrix} -\sqrt{3}/6 & -\sqrt{3}/6 & -\sqrt{3}/6 & -\sqrt{3}/2 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} & 0 \end{pmatrix}; M = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

Now, $\mathbf{w}_i = M_i|_B = \sum_j \langle M_i, B_j \rangle . B_j$. Therefore,

$$\begin{aligned}
\mathbf{w}_1 &= (3/2 \quad 3/2 \quad 3/2 \quad 9/2) \\
\mathbf{w}_2 &= (1/2 \quad -3/2 \quad 1/2 \quad -1/2) \\
\mathbf{w}_3 &= (1/2 \quad 1/2 \quad -3/2 \quad -1/2) \\
\mathbf{w}_4 &= (3/2 \quad -1/2 \quad -1/2 \quad 1/2)
\end{aligned}$$

It is obvious that $\mathbf{w}_i$ lies in the same orthant as M_i.

Lemma 5.17 *Induction step: If for* $M^k = H_{2^k}, {\mathbf{w}_i}^k = \sum_j \langle {M_i}^k, {B_j}^k \rangle . {B_j}^k$ *lies in the same orthant as* ${M_i}^k$*, then for* M^{k+1} *given by*

$$\begin{pmatrix} M^k & M^k \\ M^k & -M^k \end{pmatrix}$$

${\mathbf{w}_i}^{k+1} = \sum_j \langle {M_i}^{k+1}, {B_j}^{k+1} \rangle . {B_j}^{k+1}$ *lies in the same orthant as* ${M_i}^{k+1}$*, where* B^{k+1} *given by*

$$\begin{pmatrix} B^k & B^k \\ B^k & -B^k \end{pmatrix}$$

Proof For $1 \leq i \leq 2^k$,

$${\mathbf{w}_i}^{k+1} = \sum_j \langle {M_i}^{k+1}, {B_j}^{k+1} \rangle . {B_j}^{k+1}$$

or

$${\mathbf{w}_i}^{k+1} = 2. \sum_{j=1}^{d/2} \langle {M_i}^k, {B_j}^k \rangle . {B_j}^{k+1} + 0. \sum_{j=\frac{d}{2}+1}^{d} {B_j}^{k+1} \quad (d = 3/4 \times 2^{k+1})$$

or

$${\mathbf{w}_i}^{k+1} = 2. \sum_{j=1}^{d/2} \langle {M_i}^k, {B_j}^k \rangle . ({B_j}^k \ {B_j}^k) = ({\mathbf{w}_i}^k \ {\mathbf{w}_i}^k)$$

Since ${\mathbf{w}_i}^k$ has the same signs as ${M_i}^k$, therefore the vector $({\mathbf{w}_i}^k \ {\mathbf{w}_i}^k)$ has the same signs as $({M_i}^k \ {M_i}^k)$.

For $2^k + 1 \leq i \leq 2^{k+1}$,

$${\mathbf{w}_i}^{k+1} = \sum_j \langle {M_i}^{k+1}, {B_j}^{k+1} \rangle . {B_j}^{k+1}$$

or

$${\mathbf{w}_i}^{k+1} = 0. \sum_{j=1}^{d/2} {B_j}^{k+1} + 2. \sum_{j=\frac{d}{2}+1}^{d} \langle {M_i}^k, {B_j}^k \rangle . {B_j}^{k+1} \quad (d = 3/4 \times 2^{k+1})$$

or

$$\mathbf{w}_i{}^{k+1} = 2.\sum_{j=1}^{d/2} \langle M_i{}^k, B_j{}^k \rangle . (B_j{}^k \ - B_j{}^k) = (\mathbf{w}_i{}^k \ - \mathbf{w}_i{}^k)$$

Since $\mathbf{w}_i{}^k$ has the same signs as $M_i{}^k$, therefore the vector $(\mathbf{w}_i{}^k \ - \mathbf{w}_i{}^k)$ has the same signs as $(M_i{}^k \ - M_i{}^k)$.

Therefore the result follows. ■

Theorem 5.18 *For all n, Belcher-Hicks-Sitharam construction of $M^n = H_{2^n}$ has property 5.1.*

Proof The result follows from the base case (lemma 5.16) and the induction step (lemma 5.17).

CHAPTER 6
THE TOPOLOGY OF BALLS

In chapter 4 we defined property (4.5) - which we call the *almost-spherical* property - and property (4.6) for realization subspaces of M. These properties naturally lead to interesting questions about the topology of L_1, L_2 and L_∞ balls. To pose and answer those questions, we first formally restate the almost-spherical property in terms of L_1 and L_2 balls.

6.1 The Generalized Almost-Spherical Property

Here we give a general definition of the almost-spherical property for subspaces of $\mathbb{R}^m$ as a generalization of property (4.5).

Property 6.1 *A subspace B of $\mathbb{R}^m$ is said to be r-spherical for $r \in \mathbb{R}$ if*

$$(L_1^m = r) \cap B \subseteq (L_2^m = 1) \cap B$$

In plain English, the property says that a subspace B is r-spherical if the $L_1 = r$ ball interesected with B is fully contained in the intersection of the $L_2 = 1$ ball with B.

This statement is a generalized version of property (4.5) since it says that for all vectors in $\mathbf{w} \in B$, $||\mathbf{w}||_1 = r$ implies $||\mathbf{w}||_2 \leq 1$ or in other words, $\forall \mathbf{w} \in B, ||\mathbf{w}||_2 = 1$ implies $||\mathbf{w}||_1 \geq r$ (which is property (4.5) for $r = \sqrt{m/k}$).

The question we ask and seek to answer in this chapter is:

What is the max. k such that there exists a k-dim. subspace B of $\mathbb{R}^m$ that is r-spherical for a given r?

Alternatively, *given k, what is the max. r such that there exists a k-dim. subspace B of $\mathbb{R}^m$ that is r-spherical?*

Before trying to arrive at a closed-form expression for r in terms of m and k, we prove bounds on the value of r for specific cases. The goal is to give a range in which the limiting value of r, say r_{lim}, may lie such that for $r \leq r_{lim}$ there exists a subspace that is r-spherical and for $r > r_{lim}$ there exists no subspace that is r-spherical. Thus for every specific case that we work out, we prove the following two statements:

1. *Lower Bound Statement* (implies $r_{lim} \geq r_{lb}$):

   For $r \leq r_{lb}$, there exists a subspace B of dim. k, such that

   $$\forall \mathbf{w} \in B, (\|\mathbf{w}\|_1 = r) \Rightarrow (\|\mathbf{w}\|_2 \leq 1)$$

   We prove this statement for some extreme vectors $\mathbf{w} \in B$, which are shown to be sufficient to inspect.

2. *Upper Bound Statement* (implies $r_{lim} \leq r_{ub}$):

   For $r > r_{ub}$, for all k-dim. subspaces B of $\mathbb{R}^m$,

   $$\exists \mathbf{w} \in B, \text{ such that } (\|\mathbf{w}\|_1 = r) \text{ and } \|\mathbf{w}\|_2 > 1$$

   i.e. $(L_1^m = r) \bigcap B$ is not completely contained in $(L_2^m = 1) \bigcap B$.

After we prove the two statements, we will have narrowed down r_{lim} for that particular case to $r_{lb} \leq r_{lim} \leq r_{ub}$.

We note here that we only need to show $r_{lim} \geq \sqrt{m/k}$ for rephrasing Forster's result in terms of almost-spherical subspaces.

6.2 Observations About The Generalized Almost-spherical Property

In this section, we answer the question posed at the beginning of the chapter for specific values of m and k to gain insight into the topology of L_1 and L_2 balls and into the dependence of r on m and k.

Proposition 1. For $r \leq 1$, *all* subspaces of $\mathbb{R}^m$ of any dimension $k(\leq m)$ are r-spherical.

Proposition 2. For $r > \sqrt{m}$, there does not exist *any* subspace of *any* dimension k that is r-spherical.

Proposition 3. For $r = \sqrt{m}$, only 1-dimensional subspaces are r-spherical.

Proposition 4. For $k = m-1$, the extremal value of $r, r_{lim} = \sqrt{2}$.

We showed that for $r \leq \sqrt{2}$, we can construct a $B^{\perp} = (1, 1, 1, \ldots, 1)$ such that the $(m-1)$-dim. subspace B is r-spherical, and no other $B^{\perp}$ does better. In this case, the number of extremal points on B is given by $2.\binom{m}{2}$.

Proposition 5. For $k = m-2$, $r_{lim} \geq \sqrt{2}$.

We again constructed a $B^{\perp}$ spanned by the vectors $(1, 1, \ldots, 1)$ and $(1, 1, \ldots, -1)$ and showed that the subspace B is r-spherical for $r \leq \sqrt{2}$. Though in this case, we couldn't show that no other $B^{\perp}$ does better. Therefore, we only get a lower bound on r_{lim}. In this case, the number of extremal points on B is given by $2.\binom{m-1}{2}$.

Proposition 6(Conjecture). For m=4 and k=2, $r_{lim} = 1 + 1/\sqrt{2}$.

Proof While we can prove the lower bound, i.e. $r_{lim} \geq 1+1/\sqrt{2}$, we do not have a complete proof for the upper bound; only a proof outline that we believe will give us $r_{lim} \leq 1+1/\sqrt{2}$.

- Lower bound: $r_{lim} \geq 1 + 1/\sqrt{2}$.

  Consider the following 2-dimensional subspace of $\mathbb{R}^4$:

  $$B = \begin{pmatrix} 1 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1 & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

  We first note that the extremal vectors in B (i.e. vectors with L_1 norm r and L_2 norm as close to 1 as possible) will have as many zeroes as possible, which is $k-1$. One such vector is $(1, 1)B = (\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}, 0)$ for $r = 1+1/\sqrt{2}$. For such extremal vectors $\mathbf{x} \in B$, it can be shown that for $r \leq 1 + 1/\sqrt{2}, (||\mathbf{x}||_1 = r) \Rightarrow (|\mathbf{x}||_2 \leq 1)$.

- Upper bound(Conjecture): $r_{lim} \leq 1 + 1/\sqrt{2}$.

  While we do not have a complete proof for the upper bound, we strongly believe the conjecture. We present here a proof outline that we believe can be extended to prove the conjectured upper bound.

  Let a general $B^{\perp}$ be,

  $$B^{\perp} = \begin{pmatrix} c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{pmatrix}$$

This 2×4 subspace can be represented as a point on a *Grassmann manifold* in $\binom{4}{2}$-dimensional space. The co-ordinates of the point are given by the $\binom{4}{2} = 6$ determinants: $D_{12}, D_{13}, D_{14}, D_{23}, D_{24}, D_{34}$, where D_{ij} is,

$$D_{ij} = \begin{vmatrix} c_i & c_j \\ d_i & d_j \end{vmatrix} \neq 0$$

which satisfy the Grassman-Plucker relation,

$$D_{12} \cdot D_{34} + D_{13} \cdot D_{24} + D_{14} \cdot D_{23} = 0 \tag{6.1}$$

Now, any $\mathbf{x} \in B$ will be orthogonal to $B^{\perp}$. Hence,

$$\begin{pmatrix} c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \end{pmatrix} = 0 \tag{6.2}$$

Additionally, $\mathbf{x}$ has a 1-norm equal to r,

$$\sum_i |\mathbf{x}_i| = r \tag{6.3}$$

and should have a 2-norm less than or equal to 1,

$$\sqrt{\sum_i \mathbf{x}_i{}^2} \leq 1 \tag{6.4}$$

Since we know that the extremal vectors in B will have $k-1$ zeros, we set one of the $\mathbf{x}_i$'s in turn to zero and get 4 inequalities from the 2-norm condition

$$(D_{12})^2 + (D_{13})^2 + (D_{23})^2 \leq \frac{1}{r^2}(|D_{12}| + |D_{13}| + |D_{23}|)^2 \tag{6.5}$$

$$(D_{12})^2 + (D_{14})^2 + (D_{24})^2 \leq \frac{1}{r^2}(|D_{12}| + |D_{14}| + |D_{24}|)^2 \tag{6.6}$$

$$(D_{13})^2 + (D_{14})^2 + (D_{34})^2 \leq \frac{1}{r^2}(|D_{13}| + |D_{14}| + |D_{34}|)^2 \tag{6.7}$$

$$(D_{23})^2 + (D_{24})^2 + (D_{34})^2 \leq \frac{1}{r^2}(|D_{23}| + |D_{24}| + |D_{34}|)^2 \tag{6.8}$$

Thus, for the six D_{ij}'s to represent a valid 2-dim. r-spherical subspace of $\mathbb{R}^4$, they have to satisfy the 4 inequalities and the Grassmann-Plucker relationship.

Now, without loss of generality, we assume that $|D_{34}| \geq |D_{12}|$, and so $|D_{34}|$ is at least fourth highest in the ordering of the 6 determinants.

We claim that the maximum value of r that satisfies all 4 inequalities at the same time is attained when the D_{ij}'s take values that are as close to each other as permitted by the GP relation. This maximum value is attained by a particular "extremal" ordering of the $|D_{ij}|$'s.

Assumption: Let us assume the following ordering of D_{ij}'s:

$$|D_{13}| \leq |D_{14}| \leq |D_{23}| \leq |D_{24}| \leq |D_{12}| \leq |D_{34}| \tag{6.9}$$

The objective is to show that the maximum value of r over all valid D_{ij}'s for which all four inequalities simultaneously hold is $1 + 1/\sqrt{2}$.

Also, (6.1) can be written as

$$|D_{12}D_{34}| = |D_{13}D_{24} + D_{14}D_{23}| \tag{6.10}$$

From (6.9) and (6.10) we note that,

$$|D_{34}| \geq \sqrt{2}|D_{13}|$$

From (6.7), we observe that over all permissible values of D_{ij}'s, the expression

$$\sqrt{\frac{(|D_{13}| + |D_{14}| + |D_{34}|)^2}{(D_{13})^2 + (D_{14})^2 + (D_{34})^2}}$$

is maximized when $|D_{34}| = \sqrt{2}|D_{13}|$ and $|D_{14}| = |D_{13}|$ and is equal to $1 + 1/\sqrt{2}$.

Now for this ordering, if the maximum values for r obtained from (6.5), (6.6) and (6.8) are greater than $1 + 1/\sqrt{2}$ then $r \leq 1 + 1/\sqrt{2}$ is the value that satisfies *all four* inequalities. And we know that for the following assignment, *all four* inequalities attain $r \leq 1 + 1/\sqrt{2}$:

$$|D_{13}| = |D_{14}| = |D_{23}| = |D_{24}| = |D_{12}|/\sqrt{2} = |D_{34}|/\sqrt{2}$$

Thus, for this ordering (6.7) acts as the "bottleneck" in preventing the upper bound on r to go above $1 + 1/\sqrt{2}$.

We claim that for each ordering, one can find an inequality analogous to (6.7) that will act as a bottleneck for that ordering. We further claim that none of the other orderings can do better than this, in other words all the total orderings corresponding to the partial order that gives us $|D_{34}| \geq \sqrt{2}|D_{13}|$ are extremal.

Based on our experience working out these cases, we make the following remarks:

1. The exercise here is to pick the inequality/inequalities that give(s) the tighest upper bound on the value of r.
2. At the extremal value of r, the number of inequalities that turn into equalities is equal to the number of variables (i.e., the 6 D_{ij}'s in the m=4, k=2 case).
3. We conjecture that regardless of the value of m and k, we only need to look at a fixed number of orderings to obtain the tightest upper bound on r.
4. For insights into the Forster problem, only a *lower* bound on r_{lim} is needed. But we would have to understand something about the family of constructions/subspaces that establish this lower bound – i.e, be able to say that atleast one of them can be made to cut through any set of orthants that an arbitrary subspace of k dimensions can. Note that with our new approach, a construction is just a setting of values for the D_{ij}.

6.3 Observations on the Figure

After computing the limiting value of r for specific values of m and k and observing its variation with m and k, we make the following observations:

1. $r(m+1,k) > r(m,k)$. The extremal value of r increases with m for a constant k.
2. $r(m,k) \geq r(m,k+1)$. The extremal value of r decreases with increase in k for a constant m. This happens because the number of independent columns in $B^{\perp}$ decreases and hence it becomes easier to find a "bad" $\mathbf{x}$.

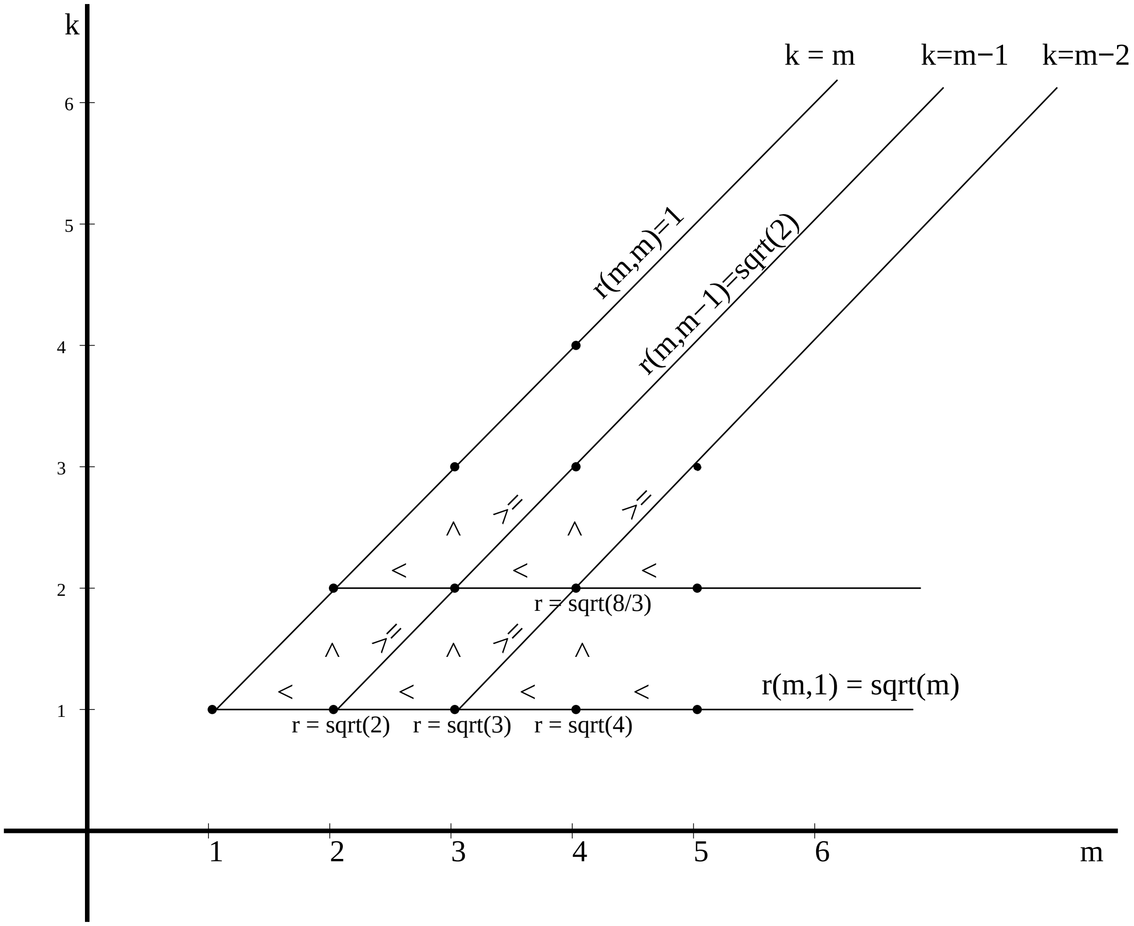

Figure 6–1: Dependence of r on m and k

3. $r(m, m-d) \geq r(m+1, m-d+1)$. For a constant difference between m and k, the extremal value of r is monotonically nonincreasing.

We believe that the closed-form expression for r will look like

$$r(m, k) = \sqrt{m} - a\sqrt{k}$$

for some constant a, though we found it hard to prove that this is the case.

CHAPTER 7
CONCLUSION AND FUTURE WORK

We have presented the study of a geometric lower bound question: what is the minimal dimension of a subspace that intersects a given set of orthants. We start with its 'avatar' as a longstanding open problem in communication complexity and its recent solution where the problem was stated as that of finding a linear arrangement of homogeneous halfspaces in the lowest dimension. We explain why the treatment of the problem and its result seem counter intuitive and justify rephrasing it as finding the minimal dimension of a subspace that intersects a given set of orthants. This rephrasing leads us to some interesting 'nice' properties of these realization subspaces that enable us to show better bounds. We also explain why the new bounds bear out our initial bewilderment over the original result, namely that the minimal dimension of a realization subspace should be a function of the set of orthants its realizes and not just some algebraic quantity such as $||M||$. The almost-spherical property of these realization subspaces leads us to some questions in functional analysis and metric embeddings such as: what is the limiting value r such that the $L_1 = r$ ball is completely contained in the $L_2 = 1$ ball in $\mathbb{R}^m$.

We would also like to identify some directions to proceed further on this problem and also some connections with problems in other areas that could be explored in the future.

1. There are a number of questions about 'niceness' of realization subspaces - can we define other kinds of niceness? Can we show that if there exists a realization, then there exists a realization that is nice in the newly defined sense? Are different kinds of 'niceness'-es equivalent? Do they help us attain tighter bounds? For example,
    (a) Does maintaining maximum distance between the columns of B have any significance as a nice property that helps to attain better results? Or
    (b) Does having columns of B split into k bundles such that each bundle is nearly orthogonal to every other qualify as a 'nice' property of any use?

2. Suppose the inner products of every pair of (a subset of l) rows of M is bounded below by some ϵ. What does it say about the norm of M (in terms of l and ϵ)? Can we use this separation as the property of M we are interested in instead of the norm? Or does it turn out to be equivalent to the norm in some sense? Does it give a better bound than the one we obtained using $ORT(M)$?
3. Though we understood the variation of r with m and k for the generalized almost-spherical property in chapter 6, we were unable to come up with a closed-form expression. Is it possible to obtain a closed-form expression for r in terms of m and k? This leads to a number of questions trying to connect our result with some results from functional analysis and theory of normed spaces. For example,
    (a) It has been shown by V. Milman *et al*[8] that in the right probability space, most k-dimensional ($k > ca^2m$, for some universal constant c) subspaces B of $\mathbb{R}^m$ have the property that the L_1 norm of all unit vectors $\mathbf{w}$ in B is bounded as

    $$(\sqrt{\frac{2}{\pi}} + a)\sqrt{m} \leq ||\mathbf{w}||_1 \leq (\sqrt{\frac{2}{\pi}} + a)\sqrt{m}$$

    Can we show the existence of one k-dimensional subspace that has the above property and also has one of the 'nice' properties we use? Does that improve the bound?

    (b) There is a celebrated theorem by Dvoretzky about embedding of L_p normed spaces into L_2, of which ours is the special case for $p = 1$. Before we state the theorem, we define the *Banach-Mazur distance* between two normed spaces X and Y to be $\leq c$, if there is a *linear* map $f : X \rightarrow Y$ with distortion(f)$\leq c$. Dvoretzky's theorem states that

    Theorem 7.1 (By Dvoretzky) *For every n and $\epsilon > 0$, every n-dimensional normed space contains a $k = \Omega(\epsilon \cdot log\ n)$-dimensional space whose Banach-Mazur distance from L_2 is $\leq 1 + \epsilon$.*

    The proof of this theorem is probabilistic in nature. Is it possible to give a constructive proof of this celebrated theorem?

    (c) Is it possible to characterize the family of all k-dimensional subspaces that are r-spherical?

4. There are a number of questions about cell(orthant)-equivalence of subspaces and linear transformations that preserve cell-equivalence. For example,
    (a) When are 2 k-dim. cell complexes (a cell complex is nothing but a set of orthants) with m hyperplanes *combinatorially* cell equivalent (i.e. have the same set of cells)? In other words, when are 2 k-dim. subspaces of m-space combinatorially cell equivalent?
    (b) When are 2 k-dim. cell complexes with m hyperplanes *real* cell equivalent (i.e. the normals to the hyperplanes for one should be obtainable from that of the other by a transformation that preserves angles)?
    (c) To answer both the above questions, characterize a complete set of ($GL(m)$ – linear transformations in m-space) operations that can be performed to a $k \times m$ matrix that will preserve combinatorial (respectively real) cell equivalence. Are all of these operations contained in $GL(r)$ for some $r \leq m$?
    (d) Given a cell complex C with m hyperplanes in k-space, how to get a k-dim. subspace of m space that is combinatorially cell-equivalent or real cell-equivalent to C?

REFERENCES

[1] Jürgen Forster. "A Linear Lower Bound on the Unbounded Error Probabilistic Communication Complexity" *Journal of Computer and System Sciences*, vol. 65, pp. 612–625, 2002.

[2] R. Paturi and J. Simon. "Probabilistic Communication Complexity," *Journal of Computer and System Sciences*, vol. 33, pp. 106–123, 1986.

[3] N. Alon, P. Frankl and V. Rödl. "Geometrical Realization of Set Systems and Probabilistic Communication Complexity," in *Proceedings of the 26th Annual Symp. on Foundations of Computer Science (FOCS)*, pp. 277–280, Portland OR, 1985.

[4] Jürgen Forster, Matthias Krause, Satyanarayana V. Lokam, Rustam Mubarakzjanov, Niels Schmitt and Hans-Ulrich Simon. "Relations Between Communication Complexity, Linear Arrangements, and Computational Complexity," in *Proceedings of the 21st Conference on Foundations of Software Technology and Theoretical Computer Science*, pp. 171–182, 2001.

[5] Eran Halperin, Jeremy Buhler, Richard Karp, Robert Krauthgamer and Ben Westover. "Detecting Protein Sequences Via Metric Embeddings," in *Proceedings of the 11th International Conference on Intelligent Systems for Molecular Biology (ISMB 2003)* 122-199, Brisbane, Australia, 2003.

[6] Jürgen Forster. "Some Results Concerning Arrangements of Half Spaces and Relative Loss Bounds," *Ph.D Dissertation, Lehrstuhl für Mathematik und Informatik, Bochum, Germany*, 2001.

[7] M. Belcher, S.Hicks, M. Sitharam. "Equiseparations," in *ACM-SE 39: Proceedings of the 39th Annual Southeast Regional Conference*, ACM Press, Athens GA (2001).

[8] V.D. Milman, G. Schechtman. "Asymptotic Theory of Finite-dimensional Normed Spaces," in *Lecture Notes in Mathematics 1200,* Springer-Verlag, Berlin (1986).

[9] Piotr Indyk. "Algorithmic Aspects of Geometric Embeddings," in *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, pp. 10–35, Las Vegas, 2001.

[10] N. Linial. "Finite Metric Spaces - Combinatorics, Geometry and Algorithms," in *Proceedings of the International Congress of Mathematicians III*, pp. 573–586, Beijing, 2002.

BIOGRAPHICAL SKETCH

Pranav Dandekar was born in Indore, India, in 1980. He went to school in Choithram High School, Indore (1983-1995), and in Sri Sathya Sai Vidya Vihar, Indore (1995-1997). He completed his undergraduate studies between 1998 and 2002 from Shri G.S. Institute of Technology and Science, Indore, which is affiliated to Rajiv Gandhi Technical University (Madhya Pradesh), India. He graduated with a B.Eng in computer engineering with distinction in August 2002. He spent the fall of 2002 working as a research assistant at KReSIT, IIT Bombay, on database performance benchmarking and optimization. He has been a graduate student in CISE at the University of Florida since January 2003. He interned with Amazom.com in Seattle, WA, during the summer of 2004 and intends to go back there to work full-time after graduating from the University of Florida.