PROBABILISTIC AND TOPOLOGICAL METHODS IN COMPUTATIONAL GEOMETRY

by

Raghavan S. Dhandapani

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science

New York University

January, 2010

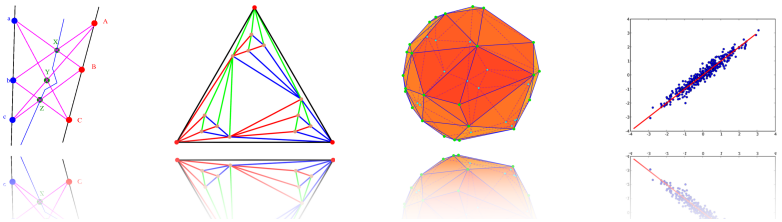
JÁNOS PACH

Probabilistic and Topological methods in Computational Geometry

Raghavan S. Dhandapani
Dissertation Defense

Courant Institute, NYU

14 December, 2009



Compiled on: Wednesday 13th January, 2010 at: 1:08pm

ABSTRACT

We consider four problems connected by the common thread of geometry. The first three arise in applications that *apriori* do not involve geometry but this turns out to be the right language for visualizing and analyzing them. In the fourth, we generalize some well known results in geometry to the *topological* plane. The techniques we use come from *probability* and *topology*.

First, we consider two algorithms that work well in practice but the theoretical mechanism behind whose success is not very well understood.

Greedy routing is a routing mechanism that is commonly used in wireless sensor networks. While routing on the Internet uses standard established protocols, routing in *ad-hoc* networks with little structure (like sensor networks) is more difficult. Practitioners have devised algorithms that work well in practice, however there were no known theoretical guarantees. We provide the first such result in this area by showing that greedy routing can be made to work on *Planar triangulations*.

Linear Programming is a technique for optimizing a linear function subject to linear constraints. *Simplex Algorithms* are a family of algorithms that have proven quite successful in solving Linear Programs in practice. However, examples of Linear Programs on which these algorithms are very inefficient have been obtained by researchers. In or-

der to explain this discrepancy between theory and practice, many authors have shown that Simplex Algorithms are efficient in *expectation* on randomized Linear Programs. We strengthen these results by proving a *partial concentration* bound for the SHADOW VERTEX Simplex Algorithm.

Next, we point out a limitation in an algorithm that is commonly used by practitioners and suggest a way of overcoming this.

Recommendation Systems are algorithms that are used to *recommend* goods (books, movies *etc.*) to users based on the similarities between their past preferences and those of other users. *Low Rank Approximation* is a common method used for this. We point out a common limitation of this method in the presence of *ill-conditioning*: the presence of *multiple local minima*. We also suggest a simple averaging based technique to overcome this limitation and show that this improves the performance of the system.

Finally, we consider some basic results in convexity like Radon's, Helly's and Carathéodory's theorems and generalize them to the *topological* plane, *i.e.,* a plane which has the concept of a linear path that is analogous to a straight line but no notion of a metric.

CONTENTS

Abstract **iv**

List of Figures **ix**

List of Tables **xi**

1 Introduction **1**

1.1 Geometry as the *lingua franca* 1

1.2 Thesis Outline 5

2 Greedy Drawings of Planar Triangulations **7**

2.1 Introduction 7

2.2 Preliminaries and Related Work 10

2.3 Outline 11

2.4 Schnyder Realizers of a Triangulation 12

2.5 Schnyder Drawings and Their Properties 14

2.6 Greedy Paths in Schnyder Drawings 17

2.7 The Main Result 22

2.8 Schnyder Drawings and the Weights of Faces 24

2.9 Conclusions 31
2.10 Recent Developments 32

3 Towards Concentration Bounds for the Simplex Algorithm 33

3.1 Introduction 33
3.2 Our Results 37
3.3 Barycentric Coordinates 38
3.4 Projecting Polyhedra 39
3.5 A Bound on the Perimeter of P 40
3.6 A Bound on Edge Lengths 41
3.7 The Bound on the Shadow Size 43
3.8 Technical Details: Bounds on Edge Lengths and Angles 44
3.9 Technical Details: The Effect of Perturbations 52
3.10 Conclusions 61

4 Low Rank Decompositions for Noisy Data 62

4.1 To Buy Or Not To Buy 62
4.2 Overview of Recommendation Systems 63
4.3 Overview of Matrix Factorization 67
4.4 The NETFLIX PRIZE 68
4.5 Our Results 70
4.6 Matrix Factorization with Regularization 70
4.7 Of Minima, Local and Otherwise 75
4.8 Ill-Conditioning in Real World Data 76
4.9 Stability in the Presence of Ill-Conditioning 78
4.10 Conclusions 81

5 Convexity in the Topological Affine Plane **83**

5.1 Basic Definitions 83

5.2 The Separation Theorem 86

5.3 Some Theorems of Combinatorial Convexity 93

5.4 Conclusions 102

Bibliography **103**

LIST OF FIGURES

2.1 A triangulation and its realizers. 13

2.2 Order and regions defined by outgoing edges. 14

2.3 Regions free of other vertices. 17

2.4 Stacked faces and 60° wedges. 18

2.5 Active and Greedy regions. 19

2.6 Weights of regions. 24

2.7 Weights of new faces. 29

3.1 Shadow size of a 20-constraint polyhedron in 10 dimensions. 37

4.1 Rows of two 500×2 matrices plotted as points in the plane. 66

4.2 Functions with multiple local minima. 75

4.3 The RMSE of the output of two Algorithms. 81

5.1 Opposite directions. 88

5.2 Modifying a pseudoline missing Y to a pseudotriangle. 89

5.3 Monotonicity of tangent direction along a line. 90

5.4 Continuity of the left tangent to a fixed set Y. 91

5.5 The left tangent from point x to set Y. 92

5.6 A strict separator. 93
5.7 $x \in \text{conv}(y_1, y_2, y_3)$. 95
5.8 $\theta_L + \theta_R > \pi$. 96
5.9 $\theta_L + \theta_R < \pi$. 96
5.10 $\theta_L = \pi$. 97
5.11 $\theta_L + \theta_R = \pi, \theta_L > 0, \theta_R > 0$. 97
5.12 Existence of a supporting pseudoline. 98
5.13 Kirchberger's theorem. 100
5.14 Minkowski's theorem on extreme points. 101
5.15 Anti-exchange. 102

LIST OF TABLES

4.1 Data Set Characteristics. 77

4.2 Local Minima for $t = 1$. 78

4.3 Local Minima for $t = 2, 3$ and 5. 79

4.4 Local Minima for $t = 10$. 80

Chapter 1

INTRODUCTION

1.1 Geometry as the *lingua franca*

In this thesis we consider four problems which can be expressed in the common language of Geometry. Three of these, namely, routing in *Sensor Networks*, solving *Linear Programs* using the *Simplex Algorithm* and building *Recommendation Systems* using *Low Rank Approximations*, *prima facie* have no connection to Geometry but as it turns out they can be modeled, visualized and analyzed in Geometrical terms. We describe all four problems briefly below.

Routing in Sensor Networks

Greedy Routing is a routing mechanism that is commonly used in wireless sensor networks. While routing on the Internet uses standard established protocols that rely on the structure in the addresses of nodes, routing in *ad-hoc* networks with little structure (such as sensor networks) is more difficult.

If every sensor has a GPS unit and knows its exact location, then a simple greedy algorithm, where a node forwards a packet to some other node that is "closer" to the

destination, can be used. In other words, a node u forwards a packet with destination v to another node u' which is such that: (a) Node u can directly communicate with u' and (b) the distance between u' and v is less than the distance between u and v. Notice that the distance to the destination decreases at every step. This is called greedy routing.

The first problem with this technique is of course, that it might not always be possible find a sensor "closer" to destination to forward the packet to. Another problem is that GPS units are expensive and the exact location of the sensor does not play a direct role in routing and so it would be helpful to find a way of making the routing work without this. A simple way around both these problems is to take the *Graph* of the network, *i.e.,* a graph G which has a node for every sensor in the network and an edge between two node if corresponding sensors can *directly* communicate with other, and *draw* it in the plane. We can now route in the same greedy fashion as before, using the coordinates of each node as the "location" of the sensor in the network. For this routing to work the drawing has to satisfy the following condition:

> For every pair of nodes $u, v \in G$, there exists a neighbor u' of u which is such that $\|u' - v\| < \|u - v\|$ where $\| \dots \|$ denotes the Euclidean distance.

Such a drawing is called a *Greedy Drawing*.

So the natural question now is if such a drawing exists for any given class of graphs. We show that it does exist for *Planar Triangulations*.

We use a technique called *Schnyder's Realizers* to partition the edges of a planar triangulation into three directed trees. By assigning weights to the faces of the triangulation, a whole family of *planar drawings* can be obtained. We use a fixed point theorem to show that there exists a member in this family that is a greedy drawing of the triangulation.

Simplex Algorithms for Solving Linear Programs

Simplex Algorithms have proven to be quite efficient in solving real world *Linear Programs* [Sha87, Bix02]. However, it has been known for a while that these algorithms can be very inefficient in the worst case [KM72, AZ99, Gol94]. Efforts to explain this apparent contradiction have focused on analyzing the efficiency of Simplex Algorithms on random and perturbed Linear Programs and many results showing that these algorithms are efficient *in expectation* have been obtained [Tod91, ST04, Bor80, Ver06]. It was observed by Shamir [Sha87], that experience with the Simplex Algorithm in practice suggests that it is not only efficient *in expectation* but also that the running time is *concentrated* around the mean. We take the first step towards placing this notion on a formal footing by proving a partial concentration bound for the SHADOW-VERTEX Simplex Algorithm.

Recommendation Systems using Low Rank Approximations

The Internet has given a big boost to modern commerce and has inundated the consumer with choice. In an attempt to help him navigate this multitude of options many online retailers are developing *Recommendation Systems*. These are algorithms that suggest goods to users based on their past preferences and those of other users. One popular algorithm for building these systems is called *Collaborative Filtering*.

We consider one popular approach to Collaborative Filtering called *Low Rank Approximation*. This approach boils down to finding the minimum of an error function. While many approaches exist to find this minimum, we explore the characteristics of the error function itself and argue that looking for the (global) minimum may not be worthwhile.

We show that the error function has multiple *different* local minima that are "as good as" the global minimum. In other words, the function value at each of these local minima

is *almost* the same as its value at the global minimum. This being the case, it is not clear that the global minimum is significant in anyway.

We suggest a simple averaging-based approach for dealing with this multiplicity. This approach has the advantage of being simpler to implement, faster to run and more robust in the presence of noise. Our approach leads to a significant increase in the accuracy of the recommendations made by the system.

Generalizing the idea of Convexity

The idea of *Convexity* is intrinsically tied to that of a *Straight* line. Indeed the standard theorems that deal with Convexity, for *e.g.,* the separation theorem, Radon's theorem, Helly's theorem *etc.* depend on the *geometric* idea of *straightness*. But is this notion really needed?

We show that many of these theorems can very well be extended to the *topological* plane. This is a plane which has no concept of a *metric* at all. A "straight" line in this plane, also called a *pseudoline*, is any simple curve that satisfies certain conditions on its intersections with other "straight" lines.

One can think of this plane as an arbitrarily "*stretched*" version of the more common Euclidean plane. We show that many of the standard theorems of convexity, like Radon's, Helly's and Carathéodory's theorems, hold in this plane as well. Hence it follows that the notion of a *metric* is not needed for these theorems.

The basic result we establish in this chapter, Lemma 67, is a generalization of the following.

Lemma 1. *Given any two disjoint convex sets X and Y in the Euclidean plane, there exists a line separating them.*

Proof: Consider two points $x_0 \in X$ and $y_0 \in Y$ which are such that the Euclidean distance $\|x_0 - y_0\|$ is minimized among all such points. In other words, these points are the solution to $\underset{x \in X, y \in Y}{\operatorname{argmin}} \|x - y\|$.

Let l be the perpendicular bisector of the line segment determined by x_0 and y_0.

Consider the case where $X \cap l \neq \emptyset$ and let $x_1 \in X \cap l$. Then it is easy to see that some point on the segment determined by x_0 and x_1 is closer to y_0 than x_0. Since this segment is contained in X (due to convexity), this is a contradiction. Hence it follows that $X \cap l = \emptyset$.

Similarly, $Y \cap l = \emptyset$. Since l bisects the segment determined by x_0 and y_0, it follows that X and Y lie on opposite sides of l.

Hence l separates X and Y.

□

Notice that the above proof depends crucially on the idea of a metric (in particular the Euclidean metric) while the result itself has only to do with separating sets and does not depend on any metric at all.

The natural question then is if one can do away with a metric in the proof as well. Our result shows that this is indeed possible.

1.2 Thesis Outline

In Chapter 2 we show that all Planar Triangulations have *Greedy Drawings* in the plane. The results in this chapter are based on [Dha08].

In Chapter 3 we prove partial concentration bounds for the SHADOW VERTEX simplex algorithm.

In Chapter 4 we show the limitations of the commonly used approach to *Low Rank Approximations* and describe how this can be overcome.

In Chapter 5 we describe how many standard results of Convexity can be extended to the Topological Affine Plane. The results in this chapter are based on [DGH$^+$06].

Chapter 2

GREEDY DRAWINGS OF PLANAR TRIANGULATIONS

2.1 Introduction

With the increasing use of large wireless communication systems comes an increasing need for reliable and scalable routing algorithms. Internet routing is accomplished using Internet Protocol addresses which are hierarchical and encode topological and geographic information about the nodes in the network. Such a protocol is not possible in an *ad-hoc* network, such as sensor nets, where little information about geographic proximity or network topology can be gleaned from node identifiers.

One important family of routing algorithms used for such networks is *Geographic (or Geometric)* routing. This is a family of algorithms that use the geographic location of the nodes as their addresses. See, for instance [KWZZ03, KK00, BMSU99, GGH$^+$01]. One such algorithm is the *Euclidean Greedy Routing* algorithm which is conceptually quite simple: Each node forwards the packet to the *neighbor*, *i.e.,* a node it can communicate directly with, that has the smallest Euclidean distance to the destination. This algorithm

has the disadvantage of not being able to deal with *lakes* or *voids* in the network, *i.e.,* nodes which have no neighbor closer to the destination. To deal with this, variants of the algorithm (such as *face* routing, which involves routing around *faces*) have been proposed, [KWZZ03, KK00].

Geometric routing has the following two drawbacks: (i) It needs the global position of every node in the network, (ii) it relies entirely on the global position and as such cannot account for local obstructions or the topology of the network. Since GPS units are quite expensive in terms of both money and power requirements, it is quite a restrictive limitation to require every node in the network to have one.

Both these issues were addressed in [RPSS03], where a variant of greedy routing which just uses the local connectivity information of the network without needing the global position of any node, was discussed. The algorithm first computes fictitious or *virtual coordinates* for each node. In other words, it *draws*[1] the graph of the network (where each node in the network is represented by a vertex of the graph and two vertices are adjacent iff the pair of nodes they represent can communicate directly) on the Euclidean plane and routes greedily using these locations. The authors obtain experimental evidence showing that this approach makes greedy routing more reliable. However no theoretical guarantees were obtained.

In a bid to place this approach on a more solid theoretical footing, Papadimitriou and Ratajczak [PR05] investigated classes of graphs on which greedy routing (without having to rely on variants like *face* routing) could be guaranteed to work, *i.e.,* graphs which can be drawn in the plane without *lakes* or *voids*. They came up with the following conjecture:

Definition 2. *A* distance-decreasing *path in a drawing of a graph is a path*

$$s = v_1, v_1, v_2, \ldots, v_k = t \textit{ such that } \|v_i - t\| < \|v_{i-1} - t\|, \;\; 2 \leq i \leq k$$

[1] *i.e.,* maps each node to a point and edges to line segments in the plane.

where $\| \ldots \|$ *denotes the Euclidean distance.*

Conjecture 1 ([PR05])**.** *Any 3-connected planar graph can be drawn*[2] *on the Euclidean plane such that there exists a distance decreasing path between every pair of vertices of the graph.*

Such a drawing is called a **Greedy Drawing** of the graph. It is easy to see that using the greedy drawing of a graph (assuming such a drawing exists) as the virtual coordinates of the vertices guarantees that greedy routing will always work.

Our Results

We settle Conjecture 1 in the affirmative for the case of planar triangulations and thus obtain the first non-trivial class of graphs for which this class of greedy routing algorithms can be guaranteed to work.

We show in fact, that a *planar* greedy drawing of any given triangulation can be obtained, *i.e.,* one in which no pair of edges cross.

The result is obtained by applying the *Knaster-Kuratowski-Mazurkiewicz* Theorem, which is known to be equivalent to the *Brouwer Fixed Point Theorem*. We believe that the technique used in obtaining the result might be of independent interest and might prove helpful in showing the existence of plane drawings with other properties.

Note that greedy drawings can be trivially seen to exist for many simple classes of graphs, like graphs with Hamiltonian circuits, all 4-connected planar graphs (since they have a Hamiltonian circuit by a theorem of Tutte [Tut56]) *etc*. It is not very difficult to show that the *Delaunay triangulation* of any set of points in the plane is also greedy. But thus far no non-trivial class of graphs with this property was known.

[2]Note that the conjecture in [PR05] uses "*embed*" instead of "*draw*". To be consistent with the Graph Drawing literature, we use "*draw*".

2.2 Preliminaries and Related Work

Given a n-vertex graph $G(V, E)$, a *drawing* of G is a mapping of the vertices of G to points and of the edges of G to curve segments (with the images of the corresponding vertices as end points) in the plane. We consider only those drawings in which the edges are mapped to straight-line segments so that the drawing is fully specified by the images of the vertices.

Recall that a *plane graph* is an abstract planar graph whose embedding has been fixed, using, say the Hopcroft-Tarjan algorithm [HT74]. We assume henceforth that G is plane triangulation. We consider only *planar drawings* of graphs, *i.e.,* drawings in which no pair of edges cross. So any reference to a drawing of a graph must be taken to mean a planar straight-line drawing.

Drawing Planar Graphs in the Plane

An overview of graph drawing algorithms can be obtained from [TBET98, NR04]. We describe some well-known algorithms for obtaining planar straight-line drawings of planar graphs:

1. *Rubber Band Embedding* [Tut60]: This algorithm has an elegant physical interpretation: Fix the positions of the vertices of some face of the graph and replace all other edges by springs (or "rubber bands"). It can be shown that if the graph is 3-connected and planar then the equilibrium position of the nodes gives a planar straight-line drawing. Many interesting generalizations of this approach have been obtained, see for instance [LLW88]. The drawback of this method is that the size of the grid required for the drawing may be large (exponential in the number of vertices).

2. *Canonical Ordering* [dFPP88]: This result showed for the first time that a planar straight-line drawing of a planar graph could be obtained on grid of polynomial (in fact $O(n) \times O(n)$) size. This approach was used in [Kan92] to obtain drawings satisfying various bounds on the minimum angle, bends, grid size *etc.*

3. *Schnyder's Realizers* [Sch90]: The author describes an elegant algorithm for partitioning the edges of a triangulation into three trees and obtaining a planar drawing (on a $O(n) \times O(n)$ grid) of the graph based on this. Our result uses the techniques developed here and so this approach is described in detail in Section 2.4. This was generalized to all 3-connected planar graphs in [Fel01]. Also see [Rot05, FPS05, BFM07, Fel04].

On a related note, it was shown recently, [Kle06b], that any graph has a greedy drawing in the *Hyperbolic* plane. But this might require an exponential sized grid, *i.e.,* $\Omega(n)$ bits might be required to store the coordinates of a single vertex, [Kle06a]. This has been further explored in [May06]. In contrast, examples of graphs with no greedy drawing in the *Euclidean* plane were obtained in [PR05].

2.3 Outline

We describe the approach of [Sch90] in Sections 2.4 and 2.5. The details of how the edges of a triangulation can be partitioned into three trees is described in the former section and the latter section describes how a drawing of the triangulation can be obtained from this partitioning along with some interesting geometric properties of these drawings.

In Section 2.6, we investigate greedy paths in drawings and show that any drawing in which every face is *good* (see Definition 11), is greedy. In Section 2.7, we prove our main

result, that there exists a greedy drawing of the triangulation, by showing that there exists a drawing in which every face is *good*.

In Section 2.8 we prove a technical result, on the sum of weights of all *bad* faces of a drawing, which is needed for the main result.

2.4 Schnyder Realizers of a Triangulation

We designate a (triangular) face f_0 of G as the *exterior* face. All vertices (edges) not belonging to f_0 are called the *interior* vertices (edges). Let the vertices of f_0 be P_0, P_1 and P_2. We define the order (P_0, P_1, P_2) to be the "counter-clockwise" (CCW) order.

Theorem 3 ([Sch90])**.** *Given a plane triangulation $G(V, E)$, there exist three directed* edge-disjoint *trees, T_0, T_1 and T_2, called the* realizer *of G, Figure 2.1, such that:*

1. *T_i is rooted at P_i, $i \in \{0, 1, 2\}$ and contains all vertices of G except P_{i+1} and P_{i-1} (the indices are $\mod 3$).*

2. *All edges of T_i are directed towards the root and every edge of G except those belonging to the exterior face are contained in exactly one T_i.*

3. *Each interior vertex, v, has exactly 3 outgoing edges, one for each T_i. The edge belonging to T_0 is followed by the one belonging to T_1 which is followed by the one in T_2 in CCW order around v, Figure 2.2a.*

   Note that there might be any number (including zero) of incoming edges of each T_i at any vertex.

Let $v \in G$ be an interior vertex. Then, it follows from the above that there exist (directed) paths $\mathcal{P}_i(v)$ from v to P_i in $T_i, i = 0, 1, 2$ called the *canonical paths* of v. From the fact the T_i are edge-disjoint and the order of the edges around v, it is clear that $\mathcal{P}_i(v)$

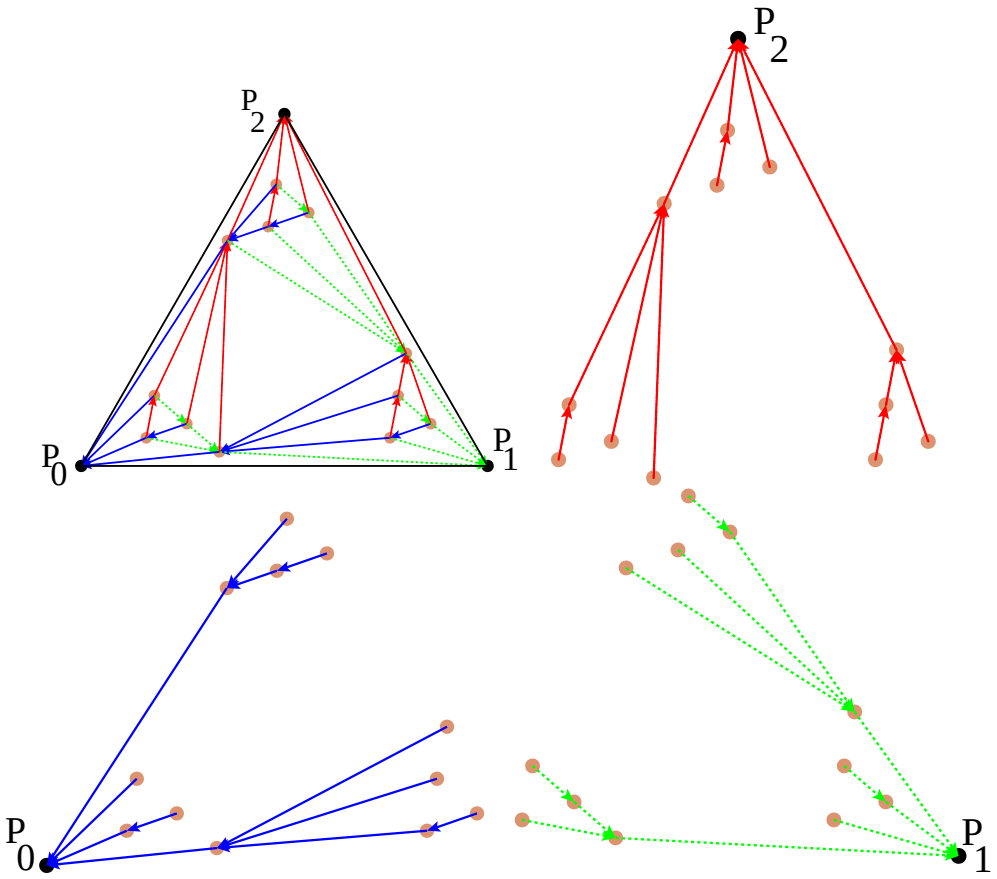


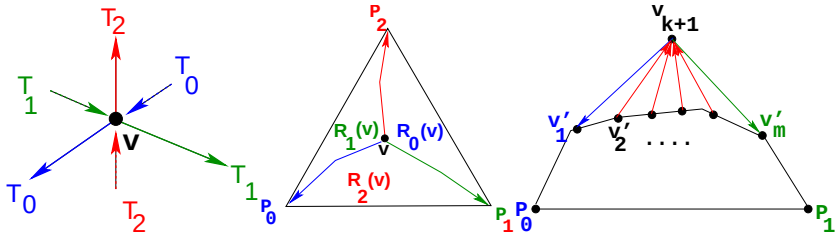
Figure 2.1: A triangulation and its realizers. The top left figure contains all three trees together and the three edges of the exterior face, which do not belong to any tree. The remaining figures show each of the three trees separately.

and $\mathcal{P}_j(v)$ must be vertex disjoint (except for v itself which appears on all three paths) if $i \neq j$. Hence the $\mathcal{P}_i(v), i = 0, 1, 2$ divide the graph G into three "regions", $R_0(v)$, $R_1(v)$ and $R_2(v)$, see Figure 2.2b.

Schnyder Realizers from Canonical Ordering

Let $f_0 = (P_0, P_1, P_2)$ be the external face of G. An ordering of the vertices $v_1 = P_0, v_2 = P_1, \ldots, v_i, \ldots, v_n = P_2$ is called a *Canonical Ordering* [dFPP88], if:

- The graph G_k induced by vertices $v_1, v_2, \ldots, v_k$ is biconnected and the boundary of



(a) (b) (c)

Figure 2.2: (a) The order of the edges belonging to different trees around an internal vertex v. There are exactly three outgoing edges, one belonging to each tree. There can be any number (including 0) of incoming edges. (b) The paths in T_i from v to P_i are vertex disjoint and divide the graph into three regions. (c) Obtaining Realizers from the Canonical Order. Note that $m \geq 2$ since G_{k+1} must be biconnected. If $m = 2$ then the edges shown directed towards v_{k+1} will not exist.

its exterior face is a cycle C_k containing edge P_0P_1.

- Vertex v_{k+1} lies in the exterior face of G_k and its neighbors form a subinterval (of length at least 2) of the path $C_k - P_0P_1$.

A simple way of using the canonical ordering to find the realizers of G was obtained in [dFPP88] and [Bre00]. We describe this below:

We process the vertices in the decreasing order of their rank in the canonical ordering. First, we add all internal edges incident to $v_n(= P_2)$ to tree T_2 and orient them towards v_n. Let the neighbors of v_{k+1} in C_k be $v'_1, v'_2, \ldots, v'_m$. We add the edge $v_{k+1}v'_1$ to tree T_0 and orient it towards v'_1. The edge $v_{k+1}v'_m$ is added to tree T_1 and oriented towards v'_m. All other edges (if any) are added to tree T_2 and oriented towards v_{k+1}, see Figure 2.2c.

2.5 Schnyder Drawings and Their Properties

Let each internal face f_i be assigned a non-negative weight w_i such that $\sum_{i=1}^{2n-5} w_i = 1$. Let $w_{R_i(v)}$ be the sum of weights of all faces in region $R_i(v)$, Figure 2.2b. For the external

vertex P_0 we define $w_{R_0}(P_0) = 1$ and $w_{R_1}(P_0) = w_{R_2}(P_0) = 0$. Weights for the other external vertices are defined in a similar manner.

We can obtain a drawing of G in the following way:

Place vertex v at the point $\big(w_{R_0(v)}, w_{R_1(v)}, w_{R_2(v)}\big)$.

Recall that we only deal with straight-line drawings and so the drawing is specified by the positions of the vertices. Since the total weight of all faces is 1, every vertex of G is placed on the $x+y+z = 1$ plane and we end up with a two-dimensional drawing. Notice that the external vertices P_0, P_1 and P_2 are always placed at the points $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ irrespective of how the weights of the internal faces are assigned and that these points determine an equilateral triangle (in the $x + y + z = 1$ plane). Also notice that all internal vertices are placed inside this equilateral triangle.

The drawing obtained by the above method is defined to be a *Schnyder Drawing* of G.

The set of solutions to the equation $\sum_{i=1}^{2n-5} w_i = 1$ such that the w_i are non-negative can be represented by the unit simplex $\mathfrak{S}$ in $2n - 6$ dimensions, with $2n - 5$ vertices. Hence, for each point $p \in \mathfrak{S}$, a Schnyder drawing of G can be obtained.

The following theorem, while a generalization of the result proved in [Sch90], follows directly from the proofs given there.

Theorem 4 ([Sch90])**.** *In any Schnyder Drawing of a triangulation G, the edges are non-intersecting,* i.e., *the drawing is planar.*

Definition 5. *A **non-degenerate** Schnyder Drawing is defined to be one obtained by assigning strictly positive weights to the faces.*

In what follows, we use the same notation for a vertex v of G and the corresponding point in the plane. The ray $\overrightarrow{P_0P_1}$ is defined to have a slope of 0° and all angles are

measured counter-clockwise from this ray. So, the ray $\overrightarrow{P_0P_2}$ has slope $60°$, $\overrightarrow{P_2P_1}$ has slope $300°$ and so on. Recall that all drawings we consider (and the points P_0, P_1 and P_2) lie on the $x + y + z = 1$ plane.

The following is a key property of Schnyder drawings.

Lemma 6 (The Three Wedges Property [Sch90, Rot05])**.** *In every Schnyder drawing the three outgoing edges at an internal vertex v have slopes that fall in the intervals $[60°, 120°]$ (T_2), $[180°, 240°]$ (T_0) and $[300°, 360°]$ (T_1), with exactly one edge in each interval. See Figure 2.3a.*

Further, if the drawing is non-degenerate, no edge has slope which is a multiple of $60°$ and every edge has positive length.

Lemma 7. *The incoming edges (if present) have slopes in the following ranges T_0 : $[0°, 60°]$, $T_1 : [120°, 180°]$ and $T_2 : [240°, 300°]$, Figure 2.3b.*

Proof: Let $v'v$ be an edge directed from v' towards v. Applying Lemma 6 at v', the result follows. □

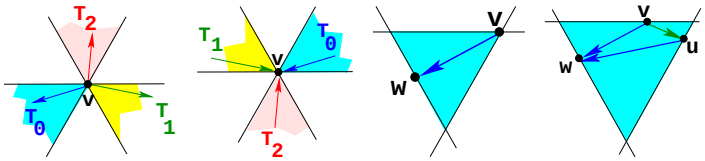
Recall that any number of incoming edges might be present at any vertex.

In the rest of the chapter, we prove many propositions specifically for non-degenerate Schnyder Drawings. Extending them to degenerate Schnyder Drawings would make the proof quite messy as degenerate drawings might have zero length edges. Also, non-degenerate drawings are sufficient for our purpose. So we disregard degenerate drawings.

Let v be a vertex and (v, w) an outgoing (at v) edge. Let the coordinates of v be $(v_{R_0}, v_{R_1}, v_{R_2})$ and the coordinates of w be $(w_{R_0}, w_{R_1}, w_{R_2})$. If $(v, w) \in T_0$, it follows from Lemma 6 that $w_{R_0} > v_{R_0}$, $w_{R_1} < v_{R_1}$ and $w_{R_2} < v_{R_2}$. Similar conclusions follow if $(v, w) \in T_1$ or if $(v, w) \in T_2$.

Let $\max_0(v, w) =$ Max (v_{R_0}, w_{R_0}) with $\max_1$ and $\max_2$ being defined in a similar

manner. The set of all points (x_0, x_1, x_2) in the drawing such that $x_i = c$ is said to be the *line determined by* $x_i = c$, $0 \leq i \leq 2$.



(a) (b) (c) (d)

Figure 2.3: (a) The shaded $60°$ wedges contain exactly one outgoing edge each. The trees containing the edges are marked. (b) All incoming edges (if present) fall in the shaded wedges. (c) The equilateral triangle determined by lines through v and w with slopes $0°, 60°$ and $120°$ is free of other vertices. A similar result holds if edge uw were to belong to T_1 or T_2 (with the equilateral triangle changing appropriately) (d) The enclosing triangle of a face.

Lemma 8 (The Enclosing Triangle Property [Sch90, Rot05])**.** *1. Let (v, w) be an edge of the graph. Consider the* equilateral *triangle determined by the lines $x_0 = \max_0(v, w)$, $x_1 = \max_1(v, w)$ and $x_2 = \max_2(v, w)$, superscribing the edge (v, w), see Figure 2.3c. This triangle is free of other vertices.*

2. *For any face $f = (u, v, w)$ the* equilateral *triangle determined by the lines $x_i = \max_i(u, v, w)$, $0 \leq i \leq 2$ is free of other vertices, see Figure 2.3d. This triangle is called the* enclosing triangle *of f.*

Proof: (1) follows from Lemma 6. (2) follows from (1) and the fact that the drawing is planar. □

2.6 Greedy Paths in Schnyder Drawings

A face of the triangulation is said to be *cyclic* if its edges form a directed cycle and is said to be *acyclic* otherwise. Any cyclic face of a graph can be *stacked* by adding a vertex

adjacent to the three vertices of the face and adding the new edges to each of the trees as shown in Figure 2.4a. This breaks the face into three acyclic faces. After a greedy drawing has been found, the new vertex can be deleted without affecting the greedy paths between the other vertices. Hence, we assume henceforth that every face in the triangulation is acyclic.

Notice that any acyclic face must have a vertex (such as vertex t in face (u, t, v) in Figure 2.4a) with two outgoing face edges which must belong to different trees. The face is said to *belong to tree* T_i if these two outgoing edges belong to trees T_{i-1} and T_{i+1}.

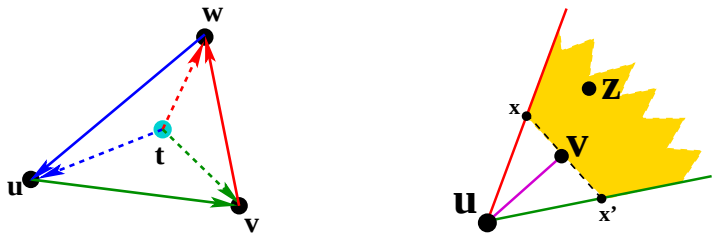


Figure 2.4: (a) The cyclic face (u, v, w) is *stacked* by adding vertex t and its incident edges. The edge tu is added to the same tree as wu, edge tv the same tree as uv and tw the same tree as vw. (b) The triangle (u, x, x') is equilateral. $\|v - z\| < \|u - z\|$ irrespective of where z lies in the shaded region and where v lies on xx'.

The following lemma will prove useful:

Lemma 9. *Let u be some vertex and (u, v) an edge incident to it. Let (u, x, x') be any equilateral triangle superscribing (u, v) with a vertex at u. Let z be any point in the wedge determined by (x, u, x') not on the same side of the line (x, x') as u, see Figure 2.4b.*

Then, $\|v - z\| < \|u - z\|$.

Proof: Let l be the perpendicular bisector of uv. It is easy to see that $z \notin l$ and it lies on the same side of l as v. Hence, it follows that $\|v - z\| < \|u - z\|$. □

To show that a drawing of G is greedy, it clearly suffices to prove:

For every (ordered) pair of distinct vertices $u, v \in V$, there exists some neighbor of u, say u' such that $\|u - v\| > \|u' - v\|$.

We will show that a non-degenerate Schnyder drawing of G exists which satisfies the above property.

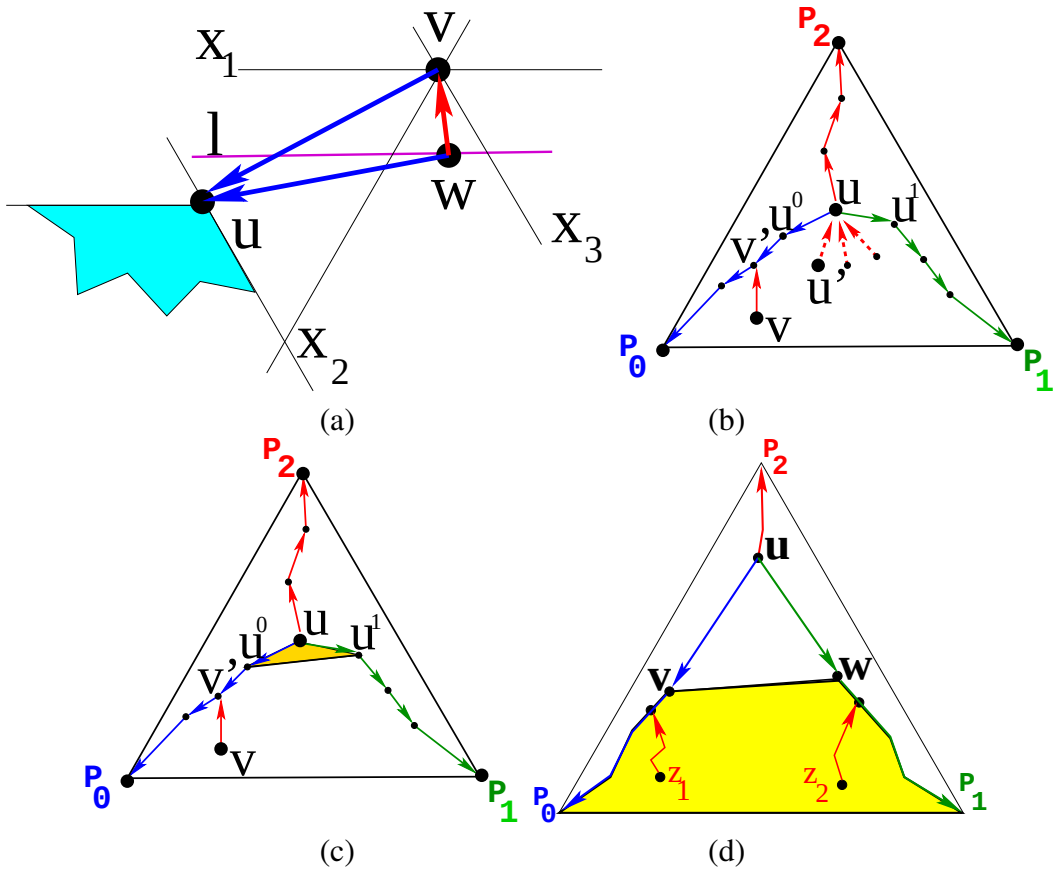


Figure 2.5: (a) An acyclic face with the active region of $\angle uvw$ shaded. The thin lines have slopes that are multiples of $60°$. The active region at u is bounded by rays with slope $180°$ and $300°$. (b) Note that v and v' need not be adjacent. (c) Vertices u^0, u and u^1 form a face. Edge u^0u^1 could be directed either way. (d) The *greedy region* of face $f = (u, v, w)$ is shown shaded.

Let $f = (u, v, w)$ be an acyclic face and let u be the vertex with two incoming edges. Without loss of generality (wlog), we assume that both edges belong[3] to T_0. See

[3]Note that both edges must belong to the same tree, see Theorem 3.

Figure 2.5a. Let the coordinates of u be (u_0, u_1, u_2). It follows that (see Lemma 7) $u_0 = \max_0(u, v, w)$ and $u_2 = \min_2(u, v, w)$.

Let the **active region** of $\angle uvw$, denoted by $\mathcal{A}_{\angle uvw}$, be the set of points (x_0, x_1, x_2) with $x_0 \geq u_0$ and $x_2 \leq u_2$. It is easy to see that this region is the wedge with sides of slopes 180° and 300° at vertex u, Figure 2.5a.

Lemma 10. *Let $f = (u, v, w)$ be an acyclic face of G and in some non-degenerate Schnyder Drawing of G, let z be a vertex in the active region of $\angle uvw$. Then, $\|v - z\| > \min(\|u - z\|, \|w - z\|)$.*

Proof: From Lemma 6, it follows that u lies below the horizontal line (denoted by l in Figure 2.5a) through w.

Since z lies in the active region of vertex u, only two possibilities can arise:

- z lies in the wedge bounded by rays of slope 180° and 240° at vertex v (the wedge x_1vx_2 in Figure 2.5a): From Lemma 9, it follows that $\|v - z\| > \|u - z\|$.

- z lies in the wedge bounded by rays of slope 240° and 300° at vertex v (the wedge x_2vx_3 in Figure 2.5a): It follows that z must lie below the horizontal line through w since u and so the whole active region lies below this line.

  Now applying Lemma 9 again it follows that $\|v - z\| > \|w - z\|$.

Hence, in every case $\|v - z\| > \min(\|u - z\|, \|w - z\|)$. □

Let u and v be a pair of non-adjacent vertices. It follows that v lies in one of three regions $R_0(u)$, $R_1(u)$ or $R_2(u)$ (or their boundaries), Figure 2.2b. Assume, wlog, that v lies in region $R_2(u)$, *i.e.,* the region bounded by the edge P_0P_1 of the external face and the paths $\mathcal{P}_i(u), i = 0, 1$ from u to P_0 and P_1, Figure 2.5b. The path $\mathcal{P}_2(v)$ from v to P_2 must intersect either $\mathcal{P}_0(u)$ or $\mathcal{P}_1(u)$. Assume wlog, that it intersects $\mathcal{P}_0(u)$ and let $v' = \mathcal{P}_2(v) \cap \mathcal{P}_0(u)$. Let u^0 (u^1) be the neighbor of u on $\mathcal{P}_0(u)$ ($\mathcal{P}_1(u)$).

The following possibilities arise:

Case I $v' = u$: Let $\mathcal{P}_2(v) = (v = v_0, v_1, v_2, \ldots, v_{k-1}, v_k = v' = u, v_{k+1}, \ldots, P_2)$. It follows from Lemmas 9 and 7 that $\|u - v\| > \|v_{k-1} - v\|$ in every non-degenerate Schnyder Drawing of G.

Case II *u has one or more edges directed inwards lying between the edges uu^0 and uu^1 in the embedding*: Let the edge following uu^0 (in CCW direction) be uu', Figure 2.5b. It follows from Lemma 6 that v lies in the active region of $\angle u^0uu'$.

Hence from Lemma 10 it follows that either $\|u^0 - v\| < \|u - v\|$ or $\|u' - v\| < \|u - v\|$ in every non-degenerate Schnyder drawing.

Case III *The vertices u, u^0 and u^1 form an acyclic face of G, see Figure 2.5c*: In this case there might exist some Schnyder drawings in which for every neighbor u^i of u, $\|u^i - v\| > \|u - v\|$. But we will show below that there must exist *some* non-degenerate Schnyder drawing in which $\|u^0 - v\| < \|u - v\|$.

The **greedy region** of a face $f = (u, v, w)$ is the region bounded by the edge vw and the paths $\mathcal{P}_0(v)$ and $\mathcal{P}_1(w)$ as shown in Figure 2.5d. Note that even though the greedy region depends on the drawing, the set of vertices falling in this region is fixed by the realizer of G.

Definition 11. *Let $f = (u, v, w)$ be a triangular face with edges uv and uw directed away from u, see Figure 2.5d, and let $\epsilon > 0$ be some constant depending only on the number of vertices of G, whose value will be fixed later. Then, in a Schnyder Drawing of G, f is said to be **good** if*

I *The length of every edge of f is at least $\sqrt{\epsilon}$.*

II *For every vertex z in the greedy region*

$$\|u-z\|^2 - \|v-z\|^2 \geq \epsilon \quad \textit{if} \quad \mathcal{P}_2(z) \cap \mathcal{P}_0(u) \neq \emptyset, \textit{ and}$$
$$\|u-z\|^2 - \|w-z\|^2 \geq \epsilon \quad \textit{if} \quad \mathcal{P}_2(z) \cap \mathcal{P}_1(u) \neq \emptyset,$$

*and is said to be **bad** otherwise.*

Note that for every vertex z in the greedy region exactly one of $\mathcal{P}_2(z) \cap \mathcal{P}_0(u)$ and $\mathcal{P}_2(z) \cap \mathcal{P}_1(u)$ is non-empty. *Clearly, a non-degenerate drawing in which every face is good, is greedy.*

The following Lemma is not used directly in our proofs but is helpful because it provides some intuition as to why the Schnyder drawing framework can lead to greedy drawings of graphs.

Lemma 12. *Given any two vertices $u, v \in G$, then in* any *non-degenerate Schnyder drawing of G, there exists a neighbor of u, say u' and a neighbor of v, say v' such that* $\|u-v\| > \min(\|u'-v\|, \|u-v'\|)$.

Proof: Follows[4] from Lemmas 6 and 9. □

2.7 The Main Result

The following theorem will prove useful:

Theorem 13 (Knaster-Kuratowski-Mazurkiewicz [KKM29])**.** *Let a d-simplex with vertices $\{v_0, \ldots, v_d\}$, be covered by* closed *sets $C_i, i \in \{0, \ldots, d\}$ such that the following* covering condition *holds:*

[4]This lemma holds more generally for all 3-connected planar graphs and not just triangulations. We will not prove this generalization here as we deal only with triangulations.

For any $Q \subseteq \{0, \ldots, d\}$ the face spanned by the vertices $\{v_i \mid i \in Q\}$ is covered by $\bigcup_{i \in Q} C_i$.

Then, $\bigcap_{i \in \{0,\ldots,d\}} C_i \neq \emptyset$.

This theorem is known to be equivalent to the Brouwer Fixed Point Theorem.

The main result is the following:

Theorem 14. *Given an n-vertex plane triangulation G, there exists a non-degenerate Schnyder drawing of G which is greedy.*

Proof: Recall that for each point $p \in \mathfrak{S}$, the unit simplex with $2n-5$ vertices (in $2n-6$ dimensions), a Schnyder Drawing of G can be obtained.

We define **good sets** $G_{f_1}, \ldots, G_{f_{2n-5}} \subseteq \mathfrak{S}$, in the following way:

Let $w = (w_1, w_2, \ldots, w_{2n-5}) \in \mathfrak{S}$. Then $w \in G_{f_i}$ iff in the Schnyder drawing of G corresponding to w, the face f_i is good. Note that the definition of these good sets depends on the value of ϵ (Definition 11).

In Section 2.8, it is shown that in any Schnyder drawing of G the sum of the weights of all the bad faces is always strictly less than 1, if ϵ is small enough (Theorem 17). Let $p = (p_0, \ldots, p_{2n-5}) \in \mathfrak{S}$ lie in the interior of some k-face of $\mathfrak{S}$. Wlog, we can assume that $p_0, p_1, \ldots, p_k > 0$ and $p_{k+1} = \ldots = p_{2n-5} = 0$. Since the sum of weights of bad faces is always less than 1, it follows that some face f_i where $i \in [0, k]$ must be good in the drawing corresponding to point p. Hence $p \in G_{f_i}$ and so the KKM covering condition is satisfied.

It is easy to see that the sets G_{f_i} are closed. The condition that the length of the edges of f_i are at least $\sqrt{\epsilon}$ can be expressed in the form $P \geq \epsilon$ where P is a quadratic polynomial, equation 2.1. It is not very difficult to see that Condition II in Definition 11 can also be expressed as a polynomial (in fact quadratic) inequality. Hence, the set G_{f_i}

can be expressed as the set of all points satisfying some weak polynomial inequalities. Hence G_{f_i} is closed.

From this it follows that the G_{f_i} satisfy the conditions of Theorem 13.

Hence, $\bigcap_{i\in\{1,\ldots,2n-5\}} G_{f_i} \neq \emptyset$. Let $g \in \bigcap_{i\in\{1,\ldots,2n-5\}} G_{f_i}$. It follows that every face is good in the Schnyder drawing corresponding to g, which implies that this drawing is greedy.

It is possible that the drawing corresponding to g is degenerate. But since $\epsilon > 0$ and the drawing varies continuously with the set of face weights, we can always pick another point g' close enough to g such that the drawing corresponding to g' is non-degenerate and greedy.

□

2.8 Schnyder Drawings and the Weights of Faces

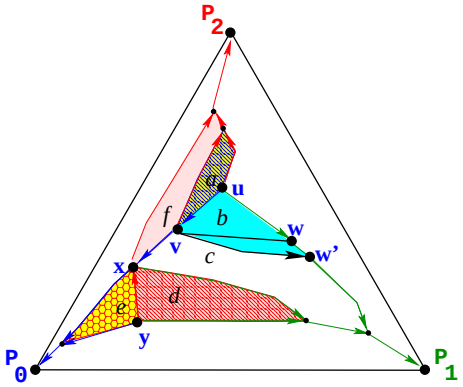


Figure 2.6: The sum of weights of faces in different regions are denoted by a, b, c, d, e and f. Note that w and w' could possibly be the same vertex, depending on how the edge wv is directed. The analysis below remains the same in either case.

In this section we show that sum of weights of the bad faces in a drawing of the triangulation is always *strictly* less than 1 for ϵ small enough where ϵ depends just on n

(the total number of vertices of G).

Consider the face $F = (u, v, w)$ in Figure 2.6. The sum of weights of faces in various regions are marked. All paths shown in the figure are canonical paths ($\mathcal{P}_i(\cdot)$) starting from some vertex. Note that b is the weight of the region demarcated by $uvw'w$ where w and w' could possibly be the same vertex.

The coordinates of the points the vertices are mapped to are given below. Recall that the graph is being drawn on the $x + y + z = 1$ plane, so the points lie on this plane. The vectors corresponding to various edges are also given below. Note that u_0 represents the first coordinate of vertex u, x_1 represents the second coordinate of vertex x and y_2 the third coordinate of y.

$$\begin{aligned}
u &= (u_0, & x_1 + a + f, & \quad y_2 + b + c + d + e),\\
v &= (u_0 + a + b, & x_1 + f, & \quad y_2 + c + d + e),\\
y &= (u_0 + a + b + c + d + f, & x_1 + e, & \quad y_2),\\
\overrightarrow{u - y} &= (-a - b - c - d - f, & a + f - e, & \quad b + c + d + e),\\
\overrightarrow{v - y} &= (-c - d - f, & f - e, & \quad c + d + e),\\
\overrightarrow{u - v} &= (-a - b, & a, & \quad b).
\end{aligned}$$

It follows that the length of the edge uv is given by

$$\|u - v\|^2 = 2(a^2 + b^2 + ab). \tag{2.1}$$

Lemma 15. *Let $\mathcal{W}_{uvw}$ be the weight of face (u, v, w). If $\mathcal{W}_{uvw} \geq \sqrt{\frac{\epsilon}{2}}$, every edge of face F has length at least $\sqrt{\epsilon}$.*

Proof:

$$\|u - v\|^2 = 2(a^2 + b^2 + ab) \geq 2\mathcal{W}_{uvw}^2 \geq \epsilon. \tag{2.2}$$

An identical argument applies to edge uw. For edge vw, notice that from Lemma 6, $\angle vuw \geq 60°$. Hence the edge vw is longer than at least one of the other two edges.

□

Note that $b \geq \mathcal{W}_{uvw}$ since b is the weight of all faces in region $uvw'w$.

Theorem 16. *Assuming that $b \geq \sqrt{\frac{\epsilon}{2}}$, the following conditions are necessary (but not sufficient) for the inequality $\|u-y\|^2 - \|v-y\|^2 < \epsilon$ to hold.*

$$a > b, \tag{2.3}$$

$$e > b, \tag{2.4}$$

$$a < \sqrt{a-b} \text{ and } b < \sqrt{a-b}. \tag{2.5}$$

Proof:

$$\begin{aligned} & \|u-y\|^2 - \|v-y\|^2 < \epsilon, \\ \implies & a(a+b+c+d+2f-e) + b(b+2c+2d+f+e) < \frac{\epsilon}{2}. \end{aligned} \tag{2.6}$$

Since $b \geq \sqrt{\frac{\epsilon}{2}}$ and all variables are non-negative, we must have:

$$\begin{aligned} & a+b+c+d+2f-e < 0, \\ \implies \quad & b < e. \end{aligned}$$

Rearranging the terms of equation 2.6, we obtain:

$$\begin{aligned} & a(a+b+c+d+2f) + b(b+2c+2d+f) + e(b-a) < \frac{\epsilon}{2}, \\ \implies \quad & b-a < 0 \implies a > b. \end{aligned}$$

for the same reason as before.

Rearranging the terms of equation 2.6 again we obtain:

$$
\begin{aligned}
& a(a+b+c+d+2f)+b(b+2c+2d+f)+e(b-a)<\frac{\epsilon}{2}, \\
\implies \quad & a^2+e(b-a)<0, \\
\implies \quad & a^2<e(a-b), \\
\implies \quad & a<\sqrt{a-b} \quad \text{(since } e<1\text{)}.
\end{aligned}
$$

□

The Maximum Weight of Bad Faces

Let point $w=(w_1,\ldots,w_{2n-5})\in\mathfrak{S}$ be such that in the Schnyder drawing, every face f_i with weight $w_i>0$ is bad. Then, the faces can be divided into three types:

Type A : The face has weight 0 and can be either good or bad.

Type B : The face has weight strictly less than $\sqrt{\frac{\epsilon}{2}}$ and is bad because either one of its edges is shorter than ϵ (and so violating Condition I in Definition 11) or because it violates Condition II in Definition 11.

Type C : The face has weight at least $\sqrt{\frac{\epsilon}{2}}$ and is bad because it violates Condition II in Definition 11.

If ϵ is small enough, then "most" of the weight must be present in faces of Type C.

Theorem 17. *There exists a positive function $\epsilon'(n)$ such that in any Schnyder drawing of G, the sum of weights of all faces of type B and C is at most $1-\epsilon'(n)$ where n is the number of vertices of G.*

We first give a brief description of the main idea behind the proof.

We try to find a point in $\mathfrak{S}$ such that, in the Schnyder drawing corresponding to it, every face with positive weight is bad. But we run into a contradiction, thus showing that such a point cannot exist.

Let $v_1, v_2, v_3, \ldots v_n$ be the canonical order of G where v_1 and v_2 are the vertices of the bottom edge of the external face and v_n is the topmost node. We start with the edge v_1v_2 and construct the triangulation by adding vertices one by one according to the canonical order. This also gives us an ordering on the faces. As faces are added, we try to assign weights to them in such a way that no face with positive weight is good. This condition places an upper bound on the weight each face can be assigned. Once we are done with all faces, we show that the sum of weight of all faces (good or bad) is forced to be less than 1, which is a contradiction.

Proof: Let $v_1, v_2, v_3, \ldots v_n$ be the canonical order of G where v_1 and v_2 are the edges of the bottom edge of the external face and v_n is the topmost node.

We start with the edge v_1v_2 and build the graph by adding vertices one by one according to the canonical order. The vertex v_3 and the face, f_1, it forms with v_1 and v_2 are shown in Figure 2.7a. Since the greedy region of v_3 contains no vertices, it is clear that f_1 cannot be a type C face. Hence $w_{f_1} < \sqrt{\frac{\epsilon}{2}}$.

Let G_k be the graph induced by the vertices $v_1, v_2, \ldots, v_k$ and let the sum of weight of all faces of G_k be W_k. We will show that to ensure that no face with positive weight is good, we must have $W_k \to 0$ as $\epsilon \to 0$ where $3 \leq k \leq n$. This is clearly satisfied by $W_3 = w_{f_1} < \sqrt{\frac{\epsilon}{2}}$. We show next that W_{k+1} also satisfies this property if W_k does. Let $W = \max\left(W_k, \sqrt{\frac{\epsilon}{2}}\right)$.

We now add v_{k+1} to G_k and try to assign weights to the newly formed faces. Let W_{new} be the maximum weight that can be assigned to the new faces while ensuring that every face with positive weight is bad. We have the two following possibilities:

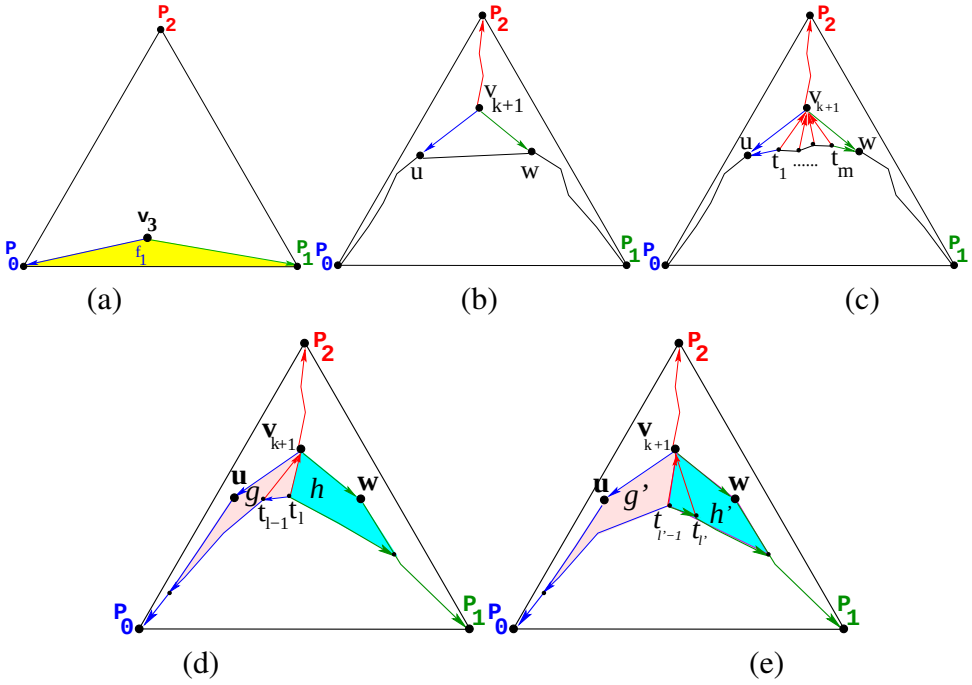


Figure 2.7: (a) The first face added. Note that the vertex P_2 and the edges P_2P_0 and P_2P_1 have not yet been added to the graph and are shown only for clarity. (b) and (c) Faces obtained when the vertex v_{k+1} is added to G_k. Note that the path from v_{k+1} to P_2 is not present in G_k and is shown only for clarity. (d) Note that only face $f_l = (v_{k+1}, t_l, t_{l-1})$ is shown to avoid clutter. (e) Only face $f'_l = (v_{k+1}, t_{l'}, t_{l'-1})$ is shown for the same reason. Note that $f_{l'}$ lies to the **right** of f_l.

Case I : v_{k+1} has only two neighbors in G_k. Let them be vertices u and w as shown in Figure 2.7b. In this case, the only new face is (u, v_{k+1}, w) which belongs to tree 2. Then it follows from equation 2.4 that $W_{new} < W$, as otherwise the face is good.

Case II : v_{k+1} has more than two neighbors, say $u, t_1, \ldots, t_m$ and w as shown in Figure 2.7c. The new faces are $f_1 = (u, t_1, v_{k+1})$ which belongs to tree 1, $f_m = (v_{k+1}, t_m, w)$ which belongs to tree 0 and $m-1$ faces of the form $f_i = (t_i, v_{k+1}, t_{i+1})$ each of which may belong to either tree 0 or 1. Let w_{f_i} be the weight of face f_i.

Case i: *None of the new faces have weight more than* W. Hence, $W_{new} < nW$ as $m + 1 < n$.

Case ii: *At least one face, say belonging to tree* 1*, has weight more than* W*, Figure 2.7d.*

Let $l \in [1..m]$ be the maximum value such that: (a) $w_{f_l} > W$ and (b) f_l belongs to tree 1. Let $l' \in [l+1..m]$ be the minimum value such that: (a) $w_{f_{l'}} > W$ and (b) $f_{l'}$ belongs to tree 0.

Of course such an l' need not exist. If it does not, then, in Figure 2.7d every face in the region h has weight at most W. By equation 2.3 applied[5] to face f_l, $g < h$ if f_l is to be bad. Every new face must fall in one of the regions g or h. Since $h < nW$, we have $W_{new} < g + h < 2nW$ (since $m + 1 < n$).

If l' does exist, then by equation 2.3, applied to face $f_{l'}$, $g' > h'$ in Figure 2.7e. Note that is possible to have $l = 1$ and/or $l' = m$.

Let $S(g), S(g'), S(h)$ and $S(h')$ denote the set of faces in the regions so marked in Figure 2.7d and 2.7e. Since $f_{l'}$ lies to the right of f_l, it is clear that $S(g) \subset S(g')$ and $S(h') \subset S(h)$ and by equation 2.3 applied to faces f_l and $f_{l'}$, $h > g$ and $g' > h'$.

Let $D_{gg'} = S(g')\backslash S(g)$ and $D_{hh'} = S(h)\backslash S(h')$ and $W_{gg'}$ ($W_{hh'}$) be the weight of the faces in $D_{gg'}$ ($D_{hh'}$). We have:

$$\begin{aligned} g + W_{gg'} \quad &= \quad g' \text{ and } h' + W_{hh'} = h \\ &\Longrightarrow \quad W_{gg'} + W_{hh'} > g' - h' \text{ since } h > g \\ and \quad &W_{gg'} + W_{hh'} > h - g \text{ since } g' > h' \end{aligned}$$

The only new faces in the sets $D_{gg'}$ and $D_{hh'}$ are $f_i, i \in [l+1, l'-1]$. Each of these faces have weight at most W (by definition of l and l'). Since the sum

[5] Note that the face shown in Figure 2.6 in the derivation of equation 2.3 belongs to tree 2 while face f_l and f'_l belong to trees 0 and 1. Of course this does not really change anything as the same argument applies. To see how equation 2.3 (or equation 2.5) applies to face f_l, compare Figures 2.6 and 2.7d where vertices v_{k+1}, t_l, t_{l-1} map to v, u, w in that order.

of all the old faces is at most W, we have: $W_{gg'} + W_{hh'} < 2(W + nW) = 2(n+1)W$.

From equation 2.5 applied to faces f_l and $f_{l'}$, it follows that $g, g', h, h' < \sqrt{2(n+1)W}$.

Hence $W_{new} < \max(2nW, c\sqrt{nW})$ where c is some small (absolute) constant. Since $W \to 0$ as $\epsilon \to 0$, we can assume that ϵ is small enough that $2nW < c\sqrt{nW}$ and so $W_{new} < c\sqrt{nW}$. Recall that $W = \max\left(W_k, \sqrt{\frac{\epsilon}{2}}\right)$. It follows that $W_{k+1} = W_k + W_{new} < W + c\sqrt{nW} < c'\sqrt{nW}$. Hence $W_{k+1} < c'\sqrt{n \max(W_k, \sqrt{\frac{\epsilon}{2}})}$. Notice that $W_{k+1} \to 0$ as $\epsilon \to 0$.

Hence it easy to see that by picking ϵ small enough, we can make $W_n < 1$ (in fact, we can make $W_n \to 0$). But the total weight of all faces must be exactly 1 and so this gives us a contradiction and the result follows.

□

2.9 Conclusions

We have been able to show that every triangulation has a planar greedy drawing in the Euclidean plane. As for algorithmic questions, the following iterative approach works quite well in practice:

- Let $\mathcal{W}^i = (w_0, w_1, \ldots, w_{2n-5}) \in \mathfrak{S}$ be the weights of the faces in iteration i.
- Let $\mathcal{W}^{i+1} = \frac{1}{W}\left(w'_0, w'_1, \ldots, w'_{2n-5}\right)$ where $w'_j = w_j$ if f_j is good in the drawing corresponding to $\mathcal{W}^i$ and $w'_j = 2w_j$ otherwise and W is the normalizing factor such that $\mathcal{W}^{i+1} \in \mathfrak{S}$.

- For $i = 0$, let $w_0 = w_1 = \ldots = w_{2n-5} = \frac{1}{2n-5}$.

This algorithm converges quite fast, but so far no theoretical bounds are known.

2.10 Recent Developments

Several results about Greedy Drawings have appeared in the literature after [Dha08]. We mention the most important ones below.

Conjecture 1 was settled in the affirmative by [ML08]. Their approach is combinatorial and they show that all 3-connected planar graphs have a subgraph with a specific structure. They then show that this subgraph has a greedy drawing in the plane. Independently, a proof based on a similar approach (but just for triangulations) was obtained in [AFG09]. However, it is not known if a *succinct* greedy drawing, *i.e.,* a greedy drawing on a polynomial sized grid, exists in the Euclidean plane.

On a slightly different track, some results about Greedy Drawing in the *Hyperbolic* plane were obtained in [EG09]. The authors describe an algorithm to obtain a succinct greedy drawing in which the location of each node can be described in $O(\log n)$ bits where n is total number of nodes.

Chapter 3

TOWARDS CONCENTRATION BOUNDS FOR THE SIMPLEX ALGORITHM

3.1 Introduction

A Linear Programming (LP) problem is a problem of the form:

$$\begin{aligned} \text{minimize:} \quad & \mathbf{c}^T\mathbf{x} \\ \text{Subject to:} \quad & \overline{A}\mathbf{x} \leq \overline{\mathbf{b}} \end{aligned}$$

where $\mathbf{x}, \mathbf{c} \in \mathbb{R}^d$, $\overline{A}$ is a $n \times d$ matrix and $\mathbf{b} \in \mathbb{R}^n$. The polyhedron $\overline{\mathcal{P}}$ defined by $\overline{A}\mathbf{x} \leq \overline{\mathbf{b}}$ is called the *constraint* polyhedron. It is easy to see that the solution to any LP, if bounded, is a vertex of this polyhedron.

LP problems arise frequently in various contexts and numerous algorithms have been developed to solve them [MG06]. Of these, the Simplex family of Algorithms have proven particularly successful in practice [Sha87, Bix02]. This family of algorithms start from a vertex of $\overline{\mathcal{P}}$ and "walk" along the edges towards the vertex minimizing $\mathbf{c}^T\mathbf{x}$. At each step a *pivot rule* determines which adjacent vertex the algorithm moves to and variants of Simplex Algorithm are characterized by the *pivot* rule they use.

Of all the variants of Simplex, the one that has proven most amenable to analysis is the SHADOW-VERTEX variant. The path followed by this algorithm can be described as the pre-image of the convex hull of the projection of $\overline{\mathcal{P}}$ onto a 2-plane as described below.

Suppose we know the vertex z' of $\overline{\mathcal{P}}$ that optimizes the objective function determined by $\mathbf{c}'$. Let z be the vertex optimizing $\mathbf{c}$. Let $\overline{P}$ be the polygon obtained by projecting $\overline{\mathcal{P}}$ onto the 2-plane determined by $\mathbf{c}'$ and $\mathbf{c}$. It is easy to see [ST04], that every vertex and edge of $\overline{P}$ is the image of some vertex and edge of $\overline{\mathcal{P}}$. In particular the images of both z' and z lie on $\overline{P}$ and so the edges of $\overline{P}$ form a path from z' to z. The SHADOW-VERTEX algorithm follows this path to find z starting from z'. In practice, this algorithm has two phases. In the first the optimal vertex z' for a specially designed objective function $\mathbf{c}'$ is found. In the second, the optimal vertex for $\mathbf{c}$, the objective function of interest, is found as described above.

Efficiency of the Simplex Algorithm

While the Simplex Algorithms have been found to work well in practice [Sha87, Bix02], most variants are known to be inefficient in the worst case [KM72, Gol94, AZ99]. In particular, the SHADOW-VERTEX algorithm is known to require 2^d steps on the d-dimensional *Goldfarb* cube.

In an attempt to explain this apparent contradiction, the effect of noise on the equations determining the constraint polyhedron was investigated in [ST04]. A *smoothed Linear Program* is an LP of the form:

$$\begin{aligned} \text{minimize:} \quad & \mathbf{c}^T\mathbf{x} \\ \text{Subject to:} \quad & A\mathbf{x} \leq \mathbf{b} \end{aligned} \tag{3.1}$$

where $A = \overline{A} + \sigma G$, $\mathbf{b} = \overline{\mathbf{b}} + \sigma \mathbf{g}$, and G ($\mathbf{g}$) is a Gaussian matrix (vector) of

appropriate dimensions and σ is the standard deviation scaled appropriately by the size of the elements of $\overline{A}$ and $\overline{\mathbf{b}}$.

Theorem 18 ([ST04, Ver06])**.** *Given any LP with $d > 3$ variables and $n > d$ constraints, the **expected** number of pivot steps in a two-phase* SHADOW-VERTEX *simplex method for the smoothed program is at most a polynomial, $Poly(n, d, \frac{1}{\sigma})$.*

In [ST04], the bound obtained was $Poly(n, d, \frac{1}{\sigma}) = O^*(n^{86}d^{55}\sigma^{-30})$. This was improved in [Ver06] to $O(\max(d^5 \log^2 n, d^9 \log^4 d, d^3\sigma^{-4}))$.

These results provide some intuition why the Simplex Algorithm is efficient in practice. However, they only show that Simplex is efficient in **expectation**. But experience suggests that Simplex Algorithms are not just efficient in expectation but in fact are efficient with *very high probability* and satisfy strong *concentration bounds* [Sha87], see Section 3.1.

Bounds on the *Shadow* Size

Let $\mathcal{P}$ be the polyhedron defined by equation (3.1). The polygon $\mathcal{P}^{\Pi}$ obtained by projecting $\mathcal{P}$ onto a (fixed) 2-plane Π is called the *Shadow* of $\mathcal{P}$. The number of vertices of $\mathcal{P}^{\Pi}$, denoted by $|\mathcal{P}^{\Pi}|$, plays an important role in determining the complexity of the SHADOW-VERTEX algorithm. In particular, this gives an upper bound on the number of steps required in the second phase [Ver06].

In the rest of the chapter, we assume that the constraint polyhedron $\mathcal{P}$ is given by the equations $A\mathbf{x} \leq \mathbf{1}$ where A is as defined in equation (3.1). It was shown in [Ver06] that any constraint polyhedron can be reduced to this form and so this assumption does not lead to loss of generality. Also, it will be be convenient to deal with the polar, $\mathcal{C} = \mathrm{Conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n)$, of $\mathcal{P}$ where the $\mathbf{a}_i$ are the rows of A. Note that since we are mainly

interested in $\mathcal{C} \cap \Pi$ (the polar of $\mathcal{P}^{\Pi}$) we can restrict our analysis to $\mathcal{C}' = \mathrm{Conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ since $|\mathcal{C} \cap \Pi| \leq |\mathcal{C}' \cap \Pi| + 1$.

Theorem 19 ([ST04, Ver06]). *Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be independent Gaussian vectors in $\mathbb{R}^d$ with centers of norm at most 1, and whose components have standard deviation $\sigma \leq \frac{1}{6\sqrt{d \log n}}$. Let Π be a fixed 2-plane in $\mathbb{R}^d$. Then the random polytope $\mathcal{C}$ satisfies*

$$\mathop{\mathbf{E}}_{\mathbf{a}_1,\ldots,\mathbf{a}_n} [|\mathcal{C} \cap \Pi|] \leq Poly(n, d, \frac{1}{\sigma}).$$

The bound obtained in [ST04] was $Poly(n, d, \frac{1}{\sigma}) = Cnd^3\sigma^{-6}$. This was improved in [Ver06] to $Cd^3\sigma^{-4}$.

Improving Bounds on the Shadow Size

Our goal is to improve the result in Theorem 19 by proving concentration bounds on the shadow size. This, together with results derived in [Ver06], will lead to a similar concentration bound on the complexity of the SHADOW-VERTEX Simplex Algorithm.

A lower bound for the concentration is provided by the d-dimensional *Goldfarb* cube [AZ99] which has a shadow of size 2^d. It was observed in [ST04] that this cube "survives" an exponentially small perturbation, *i.e.,* the facets of the cube can be perturbed by *roughly* 2^{-d} without effecting the size of the shadow. This leads to a lower bound of the form[1] $\mathop{\mathbf{Pr}}_{\mathbf{a}_1,\ldots,\mathbf{a}_n} [|\mathcal{C} \cap \Pi| \geq Poly(n, d)] \geq 2^{-d^3}$ where $Poly(n, d)$ is any polynomial in n and d. So one cannot hope for better bound.

Experiments indicate that the shadow size is highly concentrated, see Figure 3.1. We believe that a result of the form: $\mathop{\mathbf{Pr}}_{\mathbf{a}_1,\ldots,\mathbf{a}_n} [|\mathcal{C} \cap \Pi| \geq Poly(n, d)] \leq 2^{-d \log n}$ should be true.

[1] Note that this is only a rough estimate. Also note that this is a "bad" example only if $n = o(2^d)$ as otherwise the size of the shadow would be polynomial in n.

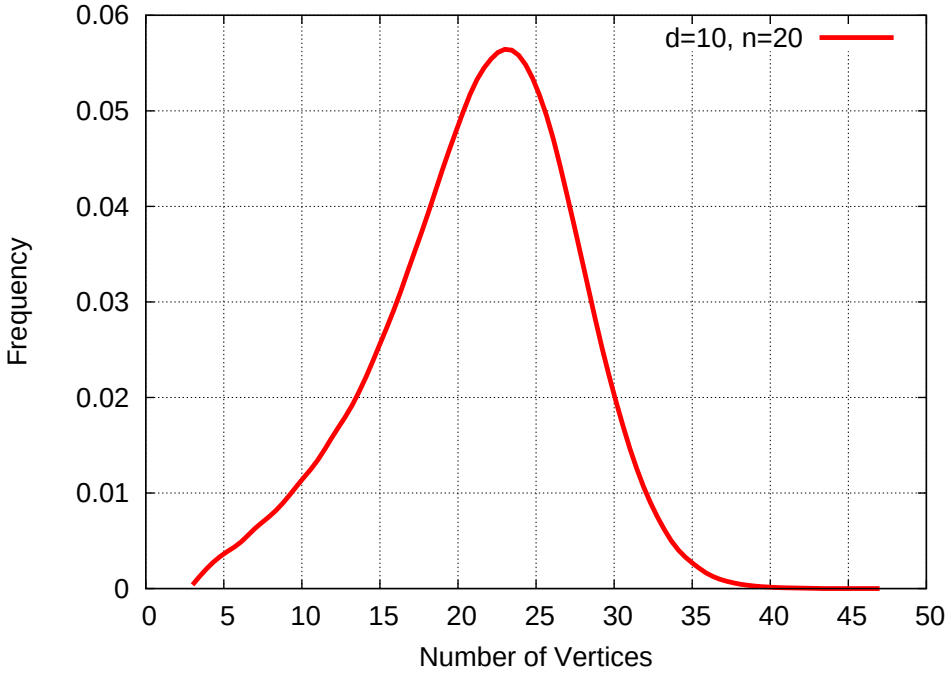


Figure 3.1: The number of vertices of the convex polygon obtained by intersecting a fixed 2-plane Π with a 20-vertex random polytope in 10 dimensions. The vertices were generated independently from the origin centered Gaussian. A total of 1.28×10^6 polytopes were generated of which $230,000$ did not intersect Π at all. These were discarded and the size of the intersection was computed for the rest 1.05×10^6 samples and the frequencies were normalized. The mean is 21.3 and the maximum is 47.

A word on the notation

In the rest of the chapter, we will use the same notation for both a point and the vector it represents. The meaning should be clear from context. The coordinates of a point $\mathbf{a}$ will be represented by superscripts, *i.e.,* $(a^{(1)}, a^{(2)}, \ldots)$. The *size* of a convex polygon, P, is defined to be the number of its vertices and is denoted by $|P|$.

3.2 Our Results

We will assume, without loss of generality *(wlog)*, that the fixed 2-plane Π is the one spanned by the last two coordinate axes. Let $\mathbf{a}_i = (a_i^{(1)}, \ldots, a_i^{(d)})$, $1 \leq i \leq n$. We

define $\mathbf{a}'_i = (a_i^{(1)}, \ldots, a_i^{(d-2)})$ and $l_i = (a_i^{(d-1)}, a_i^{(d)})\ \ \forall i$. Using this notation, the result in Theorem 19 can be written in the form:

$$\mathop{\mathbf{E}}_{l_1,\ldots,l_n,\mathbf{a}'_1,\ldots,\mathbf{a}'_n}[|\mathcal{C} \cap \Pi|] \leq Cd^3\sigma^{-4}.$$

We improve this to:

$$\mathop{\mathbf{Pr}}_{l_1,\ldots,l_n}\left[\mathop{\mathbf{E}}_{\mathbf{a}'_1,\ldots,\mathbf{a}'_n}[|\mathcal{C} \cap \Pi|] \geq \frac{Cd^{29}n^{66}\sqrt{\log n}}{\sigma^{16}}\right] \leq 2n^{-d+2} \tag{3.2}$$

where each $\mathbf{a}_i$ is a Gaussian with mean $\overline{\mathbf{a}}_i$ and standard deviation σ.

3.3 Barycentric Coordinates

Let $A = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\},\ A' = \{\mathbf{a}'_1, \ldots, \mathbf{a}'_n\}$ and $\mathrm{Affine}(\mathbf{a}_{i_1}, \ldots \mathbf{a}_{i_k})$ ($\mathrm{Conv}(\mathbf{a}_{i_1}, \ldots \mathbf{a}_{i_k})$) denote the *Affine hull* (*Convex Hull*) of the points $\mathbf{a}_{i_1}, \ldots \mathbf{a}_{i_k}$. We assume that all points are in general position, *i.e.,* the affine hull of any set of at most d points does not contain any of the other points in A or A'.

Any subset of $d-1$ elements of A is called a *ridge*. Consider a ridge $R = \{\mathbf{a}_1, \ldots, \mathbf{a}_{d-1}\}$ of A. Let $p = \mathrm{Affine}(R) \cap \Pi$. Since the points are in general position it follows that $p \neq \emptyset$. Let $\alpha = (\alpha^{(1)}, \ldots, \alpha^{(d-1)})$ be such that $\alpha^{(1)} + \ldots + \alpha^{(d-1)} = 1$ and $\alpha^{(1)}\mathbf{a}_1 + \ldots \alpha^{(d-1)}\mathbf{a}_{d-1} = p$. It is easy to see that α exists and is unique. In addition, $\mathrm{Conv}(R) \cap \Pi \neq \emptyset$ if and only if $\alpha^{(i)} \geq 0,\ \ \forall i$. Of course, in this case, $\mathrm{Conv}(R) \cap \Pi = p$. The vector α is defined to be the *Barycentric Coordinate* of the point p and is denoted by α_R. Let $\mathcal{R}$ be the set of ridges of A, $\mathcal{R}_{\geq 0} = \{R \mid R \in \mathcal{R} \text{ and } \mathrm{Conv}(R) \cap \Pi \neq \emptyset\}$ and $\alpha_{\mathcal{R}_{\geq 0}} = \{\alpha_R \mid R \in \mathcal{R}_{\geq 0}\}$.

Let $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_d\}$ be a set of points in n dimensions such that[2] $b_i = (a_1^{(i)}, \ldots, a_n^{(i)})$ where $1 \leq i \leq d$ and let $B' = \{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}\}$. Let $\mathcal{B}' = \mathrm{Affine}(\mathbf{0}, \mathbf{b}_1, \ldots, \mathbf{b}_{d-2})$ and

[2]In terms of matrices, B is the transpose of A.

$\mathcal{B}'_\perp$ be the $n-d+2$ space perpendicular to $\mathcal{B}'$ and containing $\mathbf{0}$. Let $\mathcal{S}$ be the $(n-1)$-dimensional unit simplex defined by $\{\mathbf{x} \mid x^{(1)} + \ldots + x^{(n)} = 1 \text{ and } x^{(i)} \geq 0 \ \forall i \in [1,n]\}$ where $\mathbf{x} = (x^{(1)}, \ldots, x^{(n)})$. Let $\mathcal{P}_{B'} = \mathcal{B}'_\perp \cap \mathcal{S}$. The following properties are easy to establish:

- Every $\alpha \in \alpha_{\mathcal{R}_{\geq 0}}$ is a vertex of $\mathcal{P}_{B'}$ and *vice-versa*.
- The polygon $\mathrm{Conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \cap \Pi$ is the same as the one obtained by projecting $\mathcal{P}'_B$ onto the 2-plane spanned by $\mathbf{b}_{d-1}$ and $\mathbf{b}_d$. In particular, a simple one-to-one correspondence exists between the vertices of these two polygons.

In the rest of the chapter, we denote the 2-plane spanned by $\mathbf{b}_{d-1}$ and $\mathbf{b}_d$ by Γ.

3.4 Projecting Polyhedra

For a $(d-2)$-dimensional face, F_{d-2}, of the unit simplex $\mathcal{S}$, let $u = \mathcal{B}'_\perp \cap F_{d-2}$. If $u \neq \emptyset$, then it follows that u is a vertex of $\mathcal{P}_{B'}$. Similarly, if the intersection of $\mathcal{B}'_\perp$ with any $(d-1)$ dimensional face is non-empty, then it is an edge of $\mathcal{P}_{B'}$.

Let $\mathcal{F}_{d-1}$ denote the set of $(d-1)$-dimensional faces of $\mathcal{S}$. For any $f \in \mathcal{F}_{d-1}$ we let $e_f = \mathcal{B}'_\perp \cap f$. We denote by $E(e_f)$ the event that $e_f \neq \emptyset$. Let the image of e_f after projection to Γ be denoted by e_f^Γ. We denote by $S(e_f)$ the event $(e_f^\Gamma \in P) | E(e_f)$ where $P = Shadow^\Gamma(\mathcal{P}_{B'})$. In other words $E(e_f) \wedge S(e_f)$ is the event that edge e_f exists and and its image on Γ lies on the boundary of the shadow of $\mathcal{P}_{B'}$.

Let Γ be such that the following two conditions hold:

1. The perimeter of P is at most C_1,
2. For any $f \in \mathcal{F}_{d-1}$, $\mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}} \left[\|e_f^\Gamma\| \mid E(e_f) \wedge S(e_f) \right] \geq C_2$,

where C_1 and C_2 are functions of n, d and σ and $\| \ldots \|$ denotes length, then we show that $|P|$ cannot be too high.

Theorem 20 ([KS06]). *Let C_1 and C_2 be as defined above. Then,* $\mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[|P|] \leq \frac{C_1}{C_2}$.

Proof: Let random variables $X_1, \ldots, X_{|\mathcal{F}_{d-1}|}$ be defined as follows:

$$\begin{aligned} X_i &= \|e^{\Gamma}_{f_i}\| \text{ if } E(e_{f_i}) \wedge S(e_{f_i}), 1 \leq i \leq |\mathcal{F}_{d-1}|, \\ &= 0 \text{ otherwise.} \end{aligned}$$

Let $Y_1, \ldots, Y_{|\mathcal{F}_{d-1}|}$ be such that $Y_i = \lceil X_i \rceil$.

It follows that $Perimeter(P) = \sum_{i=1}^{|\mathcal{F}_{d-1}|} X_i$ and $|P| = \sum_{i=1}^{|\mathcal{F}_{d-1}|} Y_i$.

$$\begin{aligned} \mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[Y_i] &= \mathop{\mathbf{Pr}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[X_i > 0] \leq \frac{1}{C_2} \mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[X_i], \\ \mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[Perimeter(P)] &= \sum_{i=1}^{|\mathcal{F}_{d-1}|} \mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[X_i], \\ \implies \frac{1}{C_2} \mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[Perimeter(P)] &\geq \sum_{i=1}^{|\mathcal{F}_{d-1}|} \mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[Y_i], \\ \implies \frac{C_1}{C_2} &\geq \mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[|P|]. \end{aligned}$$

□

Hence, it suffices to show that with high probability, C_1 is not too high and C_2 is not too low. We show the first in Section 3.5 and the second in Section 3.6.

3.5 A Bound on the Perimeter of P

As before, let the l_i denote the last two coordinates of $\mathbf{a}_i$. Since $\|\bar{\mathbf{a}}_i\| \leq 1$, it follows that $\|\bar{l}_i\| \leq 1$.

Theorem 21.

$$\Pr_{l_1,\ldots,l_n}\left[Perimeter(P) > 2\pi c\sqrt{d\log n}\sigma + 2\pi\right] \leq \frac{n^{-\frac{c^2}{2}d+1}}{2}$$

where c is a positive constant.

Proof: From Lemma 50 it follows that

$$\Pr_{l_1,\ldots,l_n}\left[\|l_1\| \leq 1 + c\sqrt{d\log n}\sigma \wedge \ldots \wedge \|l_n\| \leq 1 + c\sqrt{d\log n}\sigma\right] \geq 1 - \frac{n^{-\frac{c^2}{2}d+1}}{2}$$

where c is a constant. Since $P \subseteq \mathrm{Conv}(l_1, \ldots, l_n)$, the result follows. □

3.6 A Bound on Edge Lengths

Let $\mathbf{e}_{d-1}$ and $\mathbf{e}_d$ be the vector formed by the first $d-1$ components of $\mathbf{b}_{d-1}$ and $\mathbf{b}_d$ respectively and let the $(d-1)$-dimensional space formed by the first $(d-1)$ coordinate axes be denoted by Q. Recall that $\mathbf{b}_i$ are vectors in n-dimensions. Hence $\mathbf{e}_i,\ i = d-1, d$ is the projection of $\mathbf{b}_i,\ i = d-1, d$ onto Q and so $\mathbf{e}_i,\ i = d-1, d$ are Gaussian vectors with standard deviation σ centered at $\overline{\mathbf{e}}_i,\ i = d-1, d$.

Since $\|\overline{\mathbf{a}}_i\| \leq 1$, it follows that $\|\overline{\mathbf{e}}_i\| \leq d-1,\ i = d-1, d$. Let L be the $(d-2)$-dimensional plane through the origin perpendicular to $\mathbf{e}_{d-1}$ and let $\mathbf{e}_d^L$ be the projection of $\mathbf{e}_d$ onto L.

Since $\|\overline{\mathbf{a}}_i\| \leq 1$, it follows that $\|\overline{\mathbf{e}}_i\| \leq d-1, i = d-1, d$. From Lemma 52 it follows that $\Pr_{\mathbf{e}_i}\left[\|\mathbf{e}_i\| \geq d + c\sqrt{d\log n}\sigma\right] \leq n^{-c'd}$ where c and c' are constants.

Lemma 22.

$$\begin{aligned}\Pr_{\mathbf{e}_{d-1}}\left[\|\mathbf{e}_{d-1}\| \leq \epsilon\right] &\leq \left(\frac{\epsilon}{\sigma}\right)^{d-1}, \text{ and} \\ \Pr_{\mathbf{e}_{d-1},\mathbf{e}_d}\left[\|\mathbf{e}_d^L\| \leq \epsilon\right] &\leq \left(\frac{\epsilon}{\sigma}\right)^{d-2}\end{aligned}$$

Proof: The result follows from Lemma 49 since the vectors are all Gaussian. □

Lemma 23. *For every $(d-2)$-dimensional plane $N \in Q$, there exists a vector $\mathbf{q} = \mathbf{e}_{d-1}\cos(\theta) + \mathbf{e}_d \sin(\theta)$ such that projection of $\mathbf{q}$ onto N has magnitude at least ϵ with a probability at least $1 - \left(\frac{\epsilon}{\sigma}\right)^{d-1} - \left(\frac{\epsilon}{\sigma}\right)^{d-2}$.*

In particular, the magnitude is at least $\frac{\sigma}{n^c}$ with probability at least $1 - 2n^{-c(d-2)}$ for any constant c.

Recall that an edge e_{f_1} is the intersection of f_1, a $(d-1)$-dimensional face of the unit simplex $\mathcal{S}$ and $\mathcal{B}'^{\perp}$. Let the hyperplane perpendicular to e_{f_1} and passing through the origin be $H_{e_{f_1}}$. Recall that Γ is the plane spanned by $\mathbf{b}_{d-1}$ and $\mathbf{b}_d$. Any vector in Γ can be written in the form $\mathbf{l}(\theta) = \mathbf{b}_{d-1}\cos(\theta) + \mathbf{b}_d \sin(\theta)$. Let $\mathbf{l}(\theta_1)$ be the vector along the intersection of H_{f_1} and Γ. Let $\mathbf{l}(\theta_1 + \frac{\pi}{2})$ be the vector perpendicular to it. Let $\widehat{e_{f_1}}$ be the unit vector along e_{f_1} and $\widehat{\mathbf{l}}(\theta_1 + \frac{\pi}{2})$ the unit vector along $\mathbf{l}(\theta_1 + \frac{\pi}{2})$.

The following is easy to see.

Observation 24. *If e_{f_1} appears on the boundary of P, then its magnitude is given by :*

$$\|e^{\Gamma}_{f_1}\| = \|e_{f_1}\| \ |\langle \widehat{e_{f_1}} | \mathbf{l}(\theta_1 + \frac{\pi}{2})\rangle| = \|e_{f_1}\| \ |\langle \widehat{e_{f_1}} | \widehat{\mathbf{l}}(\theta_1 + \frac{\pi}{2})\rangle| \ \|\mathbf{l}(\theta_1 + \frac{\pi}{2})\|.$$

This just puts a dot product $\langle a|b\rangle$ in the form $\|a\|\cos(\theta)\|b\|$.

Now let $A = \mathrm{Affine}(f_1)$. We know that $e_{f_1} \in A$. Hence, the above observation can be put in a more convenient form:

Observation 25. *If e_{f_1} appears on the boundary of P, then its magnitude is given by :*

$$\|e^{\Gamma}_{f_1}\| = \|e_{f_1}\| \ |\langle \widehat{e_{f_1}} | \widehat{\mathbf{l}^A}(\theta_1 + \frac{\pi}{2})\rangle| \ \|\mathbf{l}^A(\theta_1 + \frac{\pi}{2})\|.$$

where $\mathbf{l}^A(\theta_1 + \frac{\pi}{2})$ is the projection of $\mathbf{l}(\theta_1 + \frac{\pi}{2})$ onto $H_{f_1} \cap A$.

Theorem 26. *For any given $(d-1)$-dimensional face $f_1 \in \mathcal{S}$:*

$$\mathop{\mathbf{Pr}}_{\mathbf{b}_{d-1},\mathbf{b}_d}\left[\|\mathbf{l}^A(\theta_1+\frac{\pi}{2})\| \leq \frac{\sigma}{n^c}\right] \leq 2n^{-c(d-2)}.$$

The probability that this condition fails for any $(d-1)$-dimensional face $f_i \in \mathcal{S}$ is at most $2n^{-(c-1)(d-1)+1}$.

Proof: The first part follows from Lemma 23 and the second one by union bound. □

Recall that $E(e_f)$ is the event that edge e_f exists and $S(e_f)$ is the event that the image of $e_f \in P$.

Theorem 27.

$$\mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[\|e_f\| \mid E(e_f) \wedge S(e_f)] \geq \frac{C\sigma^{16}}{n^{64}d^{28}}.$$

Proof: This follows directly from Theorem 47. □

Theorem 28.

$$\mathop{\mathbf{Pr}}_{\mathbf{b}_{d-1},\mathbf{b}_d}\left[\mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[\|e_f^\Gamma\| \mid \ E(e_f) \wedge S(e_f)\right] \leq \frac{C\sigma^{17}}{n^{64+c}d^{28}}\right] \leq 2n^{-(c-1)(d-1)+1},$$

where e_f^Γ is the projection of edge e_f onto plane Γ.

Proof: Follows from Theorems 26 and 27 and observation 25. □

3.7 The Bound on the Shadow Size

Theorem 29.

$$\mathop{\mathbf{Pr}}_{\mathbf{b}_{d-1},\mathbf{b}_d}\left[\mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[|P|] \geq \frac{Cd^{29}n^{64+c}\sqrt{\log n}}{\sigma^{16}}\right] \leq 2n^{-(c-1)(d-1)+1},$$

where c is some constant greater than 1.

Proof: The result follows from Theorems 20, 21 and 28. □

Recall that $P = \mathcal{C} \cap \Pi$. Hence, it follows that:

Theorem 30.

$$\Pr_{l_1,l_2,\ldots,l_n}\left[\mathop{\mathbf{E}}_{\mathbf{a}'_1,\ldots,\mathbf{a}'_n}[|\mathcal{C}\cap\Pi|] \geq \frac{Cd^{29}n^{66}\sqrt{\log n}}{\sigma^{16}}\right] \leq 2n^{-d+2},$$

which completes the result.

3.8 Technical Details: Bounds on Edge Lengths and Angles

As before, we let $f \in \mathcal{F}_{d-1}$ and $e_f = \mathcal{B}'_\perp \cap f$. In this section we obtain lower bounds on $\mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[\|e_f^\Gamma\| \mid E(e_f) \wedge S(e_f)\right]$ for the case where Γ is not "too bad". For now, we will consider Γ to be fixed. It is intuitively clear that if $\mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[\|e_f^\Gamma\| \mid E(e_f) \wedge S(e_f)\right]$ is large then the expected shadow size, $\mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[|P|]$, cannot be too high.

Consider the edge $e_f = \mathcal{B}'_\perp \cap f$ where f is a $(d-1)$-dimensional face of $\mathcal{S}$. Let g be some $(d-2)$-dimensional face of f. Let $u_g = \mathcal{B}'_\perp \cap g$ and let $E(u_g)$ be the event that $u_g \neq \emptyset$. If $E(u_g)$ holds then it follows that $e_f \neq \emptyset$ and that u_g is a vertex of e_f.

Let the $(d-1)$-dimensional faces incident on g be $\{f_1 = f, \ldots, f_{n-d+1}\}$ and let $e_{f_i} = \mathcal{B}'_\perp \cap f_i$. Let $\widehat{e}_{f_i}$ be the *unit* vector along e_{f_i} (directed away from u_g). Now let $u_g = (u_g^{(1)}, \ldots, u_g^{(d-1)}, 0, \ldots, 0)$ where the last $n-d+1$ coordinates are zero. It follows that $E(u_g) \iff u_g^{(i)} \geq 0, \ \forall i$.

Let the vertices of the unit $(n-1)$-dimensional simplex $\mathcal{S}$ be $h_1, \ldots, h_n$ where each h_i lies at a unit distance along the i^{th} coordinate axis. Without loss of generality (*wlog*) we let the vertices of g be $\{h_1, \ldots, h_{d-1}\}$ and f_i be $g \cup \{h_{d-1+i}\}$, $1 \leq i \leq n-d+1$. In the rest of the chapter we use $h_i, \ 1 \leq i \leq n$, to denote both a vertex of $\mathcal{S}$ and the unit vector

along the i^{th} coordinate axis. The context will provide the necessary disambiguation. We let $h_0 = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$ be the vector perpendicular to the hyperplane containing $\mathcal{S}$.

The set of points $B' = \{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}\}$ is called a $(d-2)$-*figure* and it is easy to see that it can be represented as a point in $\mathbb{R}^{(d-2)n}$. In the rest of this section, we let $\mathbf{p} \in \mathbb{R}^{(d-2)n}$ be the vector obtained by concatenating the $\mathbf{b}_i$.

We now investigate how $\mathbf{p} \in \mathbb{R}^{(d-2)n}$ can be varied without effecting the values of $\widehat{e}_{f_i}$. But first, we make the following observation:

Observation 31. *Let $\mathbb{L} \subseteq \mathbb{R}^{(d-2)n}$ be such that for every $\mathbf{p} \in \mathbb{L}$:*

1. *$E(u_g)$ holds (*i.e., *the vertex u_g exists) and,*
2. *The directions of the unit vectors $\widehat{e}_{f_i}$ do not change as $\mathbf{p}$ varies within $\mathbb{L}$.*

Then, the value of $S(e_{f_i})$ is fixed for all $\mathbf{p} \in \mathbb{L}$. In other words, if $S(e_{f_i})$ holds for some $\mathbf{p}_1 \in \mathbb{L}$, then it holds for all $\mathbf{p} \in \mathbb{L}$.

Recall that $e_{f_1} = \mathcal{B}'_\perp \cap f_1$ where f_1 has vertices $\{h_1, \ldots, h_{d-1}, h_d\}$. It follows that the vectors $\{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}, h_0, h_{d+1}, h_{d+2}, \ldots, h_n\}$ are all perpendicular to e_{f_1}.

Let $H_1 = Span(\{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}, h_0, h_{d+1}, \ldots, h_n\})$. Note that H_1 is a $(n-1)$-plane. It follows that $\widehat{e}_{f_1}$ is fixed if and only if H_1 is fixed. Let $H_i,\ 1 < i \leq n-d+1$ be similarly defined. Let $Q = \bigcap_{1 \leq i \leq n-d+1} H_i$. It follows that $Q = Span(\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}, h_0)$ and that if the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}$ are varied such that Q does not change, then the directions of $\widehat{e}_{f_i}$ do not change[3].

Observation 32. *Let $\mathbb{L} \subseteq \mathbb{R}^{(d-2)n}$ be such that for every $\mathbf{p} \in \mathbb{L}$, $Span(\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}, h_0)$ is fixed. Then, whenever $e_{f_i} \neq \emptyset$, $\widehat{e}_{f_i},\ 1 \leq i \leq n-d+1$ lies along some fixed unit vector.*

[3]Notice that in this case the direction of every edge of $\mathcal{P}'_B$ is fixed. However the edge itself may cease to exist.

Recall that $u_g \in g$. The vectors $\{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}, h_d, \ldots, h_n\}$ are all perpendicular to the vector u_g. Reasoning as before, it follows that:

Observation 33. *Let $\mathbb{L} \subseteq \mathbb{R}^{(d-2)n}$ be such that for every $\mathbf{p} \in \mathbb{L}$ $Span(\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}, h_d)$ is fixed, then, the vertex u_g and the edges $e_{f_2}, e_{f_3}, \ldots, e_{f_{n-d+1}}$ are fixed. In other words for different values of $\mathbf{p} \in \mathbb{L}$, where $\mathbb{L}$ is as above, the only quantity that varies is the direction and length of the edge e_{f_1}. The edges $e_{f_2}, \ldots, e_{f_{n-d+1}}$ and the vertex u_g are all fixed.*

An upper bound on $\Pr_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[\|e_f\| \leq \epsilon | E(u_g) \wedge S(e_f)]$

Recall that $g = \mathrm{Conv}(h_1, \ldots, h_{d-1})$ and let $G = \mathrm{Affine}(h_1, \ldots, h_{d-1}, \mathbf{0})$. Since G has $(d-1)$ dimensions and contains the origin, it follows that $G \cap \mathcal{B}'_\perp$ is a line though the origin. Also $u_g = (G \cap \mathcal{B}'_\perp) \cap g$.

Let Q be some arbitrary (and fixed) $(d-1)$-plane containing the origin and the vector h_0. We now condition on the event $\mathbf{b}_i \in Q, 1 \leq i \leq d-2$ and derive our bound under this restriction. Let $\mathbf{l} = G \cap \mathcal{B}'_\perp$ and $\widehat{\mathbf{l}}$ be the unit vector along $\mathbf{l}$.

From observation 32 it follows that the edge directions, $\widehat{e}_{f_1}, \ldots, \widehat{e}_{f_{n-d+1}}$ do not change as $\mathbf{b}_i$ vary within Q. The edges may cease to exist however and their magnitude may change. Let $h'_0 = (\frac{1}{\sqrt{d-1}}, \ldots, \frac{1}{\sqrt{d-1}}, 0, 0, \ldots, 0)$ where the first $d-1$ coordinates are non-zero and the rest $n-d+1$ are zero. Let $\Xi = Span(h'_0, \widehat{\mathbf{l}})$. We now further restrict $\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}$ to vary in a such in a manner that Ξ is fixed (and arbitrary).

Since $h'_0 \in \Xi$ it follows that Ξ intersects the face g of the simplex $\mathcal{S}$. Let h''_0 be a unit vector perpendicular to h'_0 within Ξ. Since $\widehat{\mathbf{l}} \in \Xi$, it can be parametrized in terms of h'_0 and h''_0. Let $\widehat{\mathbf{l}}(\theta) = h'_0 \cos(\theta) + h''_0 \sin(\theta)$. An angle $\theta \in [0, 2\pi]$ is said to be *active* if the line containing $\widehat{\mathbf{l}}(\theta)$ intersects the face g (*i.e.,* if $E(u_g)$ holds).

Observation 34. *The set of active angles is an interval of the form* $[0, \theta_0] \cup [\theta_1, 2\pi]$. *This follows from the fact that "moving"* $\widehat{\mathbf{l}}$ *towards* h'_0 *cannot destroy* u_g. *Also it is easy to see that:* $\theta_0, 2\pi - \theta_1 \geq \cos^{-1}(\sqrt{1 - \frac{1}{d-1}})$.

Since $\|e_f\| \geq \min_{i \in [0,d-1]}(u_g^{(i)})$, a lower bound on $\min_{i \in [0,d-1]}(u_g^{(i)})$ implies a lower bound on $\|e_f\|$. Notice that if the vertex u_g lies in the "interior" of the face g, *i.e.,* is not close to any facet of g, then it follows that $\min_{i \in [0,d-1]}(u_g^{(i)})$ cannot be too small. Recall that the vertex u_g is the point intersection between the line containing the vector $\widehat{\mathbf{l}}$ and the face g. It is easy to see that smaller the angle between $\widehat{\mathbf{l}}$ and h'_0, the larger the value of $\min_{i \in [0,d-1]}(u_g^{(i)})$. In fact, if the vectors coincide, $\min_{i \in [0,d-1]}(u_g^{(i)}) = \frac{1}{d-1}$ and this is the largest value of $\min_{i \in [0,d-1]}(u_g^{(i)})$ possible. Hence, it clear that the "bad" points which must be avoided are the boundary points, θ_0 and θ_1, of the active interval, see observation 34. We formalize this notion in the next Lemma.

Lemma 35. *Let* u_g, $\widehat{\mathbf{l}}$, h'_0, h''_0 *and* Ξ *be as above. Let* $\widehat{\mathbf{l}}(t) = h'_0 t + h''_0\sqrt{1 - t^2}$ *where* $t \in [-1, 1]$ *and the active interval be* $[0, \theta_0] \cup [\theta_1, 2\pi]$. *Let* $t_0 = \cos(\theta_0)$ *and* $t_1 = \cos(\theta_1)$. *Then: (i)* $\min_{i \in [0,d-1]}(u_g^{(i)}(t_0)) = \min_{i \in [0,d-1]}(u_g^{(i)}(t_1)) = 0$, *(ii)* $\min_{i \in [0,d-1]}(u_g^{(i)}(t_0 + \epsilon)) \geq \frac{\epsilon}{(d-1)^{\frac{3}{2}}}$.

Proof: Part(i) follows simply from the fact that $\widehat{\mathbf{l}}(t_0)$ and $\widehat{\mathbf{l}}(t_1)$ intersect g at one of its facets.

Part(ii): Let the intersection point of the line containing the vector h'_0 and g be $\mathbf{c}$. Notice that $\mathbf{c} = (\frac{1}{d-1}, \ldots, \frac{1}{d-1}, 0, \ldots, 0)$ where the first $d-1$ coordinates are non-zero and the rest $n - d + 1$ coordinates are zero.

Let $l = \|\mathbf{c} - u_g(t_0)\|$. Notice that $l \leq 1$ since the face g has a circumsphere of radius less than 1. Therefore,

$$u_g(t_0 + \epsilon) = u_g(t_0)\left(\frac{l - \|u_g(t_0 + \epsilon) - u_g(t_0)\|}{l}\right) + \mathbf{c}\left(\frac{\|u_g(t_0 + \epsilon) - u_g(t_0)\|}{l}\right),$$

since by design, $u_g(\theta)$ moves towards $\mathbf{c}$ along the line containing $u_g(\theta_0)$ and $\mathbf{c}$ as θ varies from θ_0 to 0.

Every point in g is at least $\frac{1}{\sqrt{d-1}}$ from the origin. Hence, we have $\|u_g(t_0+\epsilon) - u_g(t_0)\| \geq \frac{1}{\sqrt{d-1}}\epsilon$.

It follows that $u_g^{(i)}(t_0+\epsilon) \geq \frac{\epsilon}{(d-1)^{\frac{3}{2}}},\ 1 \leq i \leq d-1$ since $u_g^{(i)}(t_0) \geq 0\ \forall i$. □

Theorem 36. *Let $\mathbf{b}_i$, $h_0, h_0', h_0'', t_0, t_1, \Xi$ and $\widehat{\mathbf{l}}$ be as before. Let Q be some $(d-1)$-space containing the origin and h_0. Let $\widehat{h'^Q}_0$ and $\widehat{h''^Q}_0$ be the unit vectors along the projection of h_0' and h_0'' onto Q and let $\delta = \langle \widehat{h'^Q}_0 | \widehat{h''^Q}_0 \rangle$.*

$$\mathop{\mathbf{Pr}}_{\mathbf{b}_i \in Q,\ 1 \leq i \leq d-2, \widehat{\mathbf{l}} \in \Xi} \left[\min_{i \in [0,d-1]} (u_g) \leq \epsilon \right] \leq \frac{4C(d-1)^{\frac{3}{2}}}{(1-|\delta|)^2} \frac{\epsilon}{\epsilon_0} \leq \frac{C' n^2 d^{3.5}}{(1-|\delta|)^2 \sigma^2} \epsilon$$

where $n > d > 3$, $\sigma < 1.0$, $0 < \epsilon < \frac{\sigma^2}{6n^2d^2} = \epsilon_0$ and C, C' are constants.

Proof: It follows directly from Theorem 58 that

$$\mathop{\mathbf{Pr}}_{\mathbf{b}_i \in Q,\ 1 \leq i \leq d-2, \widehat{\mathbf{l}} \in \Xi} \left[\langle \widehat{\mathbf{l}} | h_0' \rangle \in [t_0, t_0+\epsilon] \right] \leq \frac{4C}{(1-|\delta|)^2} \frac{\epsilon}{\epsilon_0}.$$

Lemma 35 does the rest. We only need to bound the values of $1-|t_0|$ and $1-|t_1|$. It follows from observation 34 that $|t_0|, |t_1| \leq \sqrt{1-\frac{1}{d-1}}$. Since $\sqrt{1-x} \leq 1 - \frac{x}{2}, \forall x \in [0,1]$ it follows that $1-|t_0|, 1-|t_1| \leq \frac{1}{2(d-1)}$ and the result follows. □

Now we only need to bound δ. From Lemma 61 it follows that for a given edge e_f,

$$\mathop{\mathbf{Pr}}_{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}} \left[1 - |\delta_{e_f}| \leq \epsilon \right] \leq 4 \left(\frac{2\sqrt{2} n \epsilon}{\sigma} \right)^{d-3} + n^{-2.9d+3.9}.$$

Hence it follows that for any edge e_f:

Theorem 37.

$$\begin{aligned} \mathop{\mathbf{Pr}}_{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}} \left[\|e_f\| \leq \epsilon \mid E(u_g) \wedge S(e_f) \right] &\leq \mathop{\mathbf{Pr}}_{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}} \left[\min_{i \in [0,d-1]} (u_g^{(i)}) \leq \epsilon \mid E(u_g) \wedge S(e_f) \right] \\ &\leq \frac{C n^8 d^{3.5}}{\sigma^4} \epsilon + 2n^{-2.9d+3.9}, \end{aligned}$$

where C is some constant and $\sigma < 1.0$ and $n > d > 3$ and $0 < \epsilon < \frac{\sigma^2}{6n^2d^2}$.

A Bound on the Angles

In the last section, we obtained an upper bound on $\mathop{\mathbf{Pr}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}[\|e_f\| \leq \epsilon \mid E(u_g) \wedge S(e_f)]$ via an upper bound on $\mathop{\mathbf{Pr}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[\min_{i\in[0,d-1]}(u_g^{(i)}) \leq \epsilon | E(u_g) \wedge S(e_f)\right]$. Since our goal is to find an upper bound on $\mathop{\mathbf{Pr}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[\|e_f^{\Gamma}\| \leq \epsilon \mid E(u_g) \wedge S(e_f)\right]$ we now analyze the angle between e_f and the 2-plane Γ.

Let $e_{f_1}^{\perp}$ be the hyperplane perpendicular to the edge e_{f_1} and passing through the origin. Let $\mathbf{j}_1$ be unit vector along $e_{f_1}^{\perp} \cap \Gamma$. Let $ExtAngle(e_{f_1})$ denote the external angle (formed by the normals to the facets) at the edge. The following is easy to see:

Observation 38. *The edge e_{f_1} appears on the boundary of the polygon P (*i.e., *$S(e_{f_1})$ holds) if and only if $\mathbf{j}_1 \in ExtAngle(e_{f_1})$.*

Now let $\mathbf{j}_2 \in \Gamma$ be a unit vector perpendicular to $\mathbf{j}_1$. It is easy to see that if the angle between e_{f_1} and $\mathbf{j}_2$ is close to $90°$ then $\|e_{f_1}^{\Gamma}\|$ is close to 0. Thus we need to bound this probability.

Let $\mathbf{j}_2^{f_1}$ be the projection of $\mathbf{j}_2$ onto $\mathrm{Affine}(f_1, \mathbf{0})$ and let $\widehat{\mathbf{j}^{f_1}}_2$ be the unit vector in this direction. As before, let $\widehat{e}_{f_1}$ be the unit vector along e_{f_1} directed away from u_g. We seek an upper bound on: $\mathop{\mathbf{Pr}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[|\langle \widehat{e}_{f_1} | \widehat{\mathbf{j}^{f_1}}_2 \rangle| \leq \epsilon | E(u_g) \wedge S(e_f)\right]$.

Observation 39. *Given u_g and $e_{f_1}, \ldots, e_{n-d+1}$ such that $E(u_g)$ and $S(e_{f_1})$ hold, varying e_{f_1} in the 2-plane spanned by $\widehat{e}_{f_1}$ and $\widehat{\mathbf{j}^{f_1}}_2$ preserves $S(e_{f_1})$ (and $E(u_g)$).*

Now let Q be some (fixed) $(d-1)$-dimensional plane through the origin containing h_d. From observation 33 it follows that if the $\mathbf{b}_i$ are restricted to vary such that $\mathcal{B}' \in Q$ then the only e_{f_1} varies. Fix some 2-plane Ξ' perpendicular to $\mathbf{j}_1$ and containing $\widehat{\mathbf{j}^{f_1}}_2$. Now, if B' is permitted to vary such that (i) $\mathcal{B}' \in Q$ and (ii) $\widehat{e}_{f_1} \in \Xi'$, then it follows from observation 39 that $S(e_{f_1})$ is preserved.

Let $\mathbf{j}_3 \in \text{Affine}(f_1, \mathbf{0})$ be a unit vector perpendicular to $\mathbf{j}_1$ and $\widehat{\mathbf{j}^{f_1}}_2$ and let $\Xi' = Span(\mathbf{j}_3, \widehat{\mathbf{j}^{f_1}}_2)$. If $E(u_g)$ holds then it follows that the line containing $\mathbf{j}_3$ and passing through u_g intersects some facet of f_1 (other than g). Without loss of generality we assume that this face is $\text{Conv}(h_2, \ldots, h_d)$. Let this intersection point be w_3. Let the line containing $\widehat{\mathbf{j}^{f_1}}_2$ and passing through u_g intersect $\text{Affine}(h_2, \ldots, h_d)$ at point w_2. Let the line containing w_3 and w_2 be denoted by l. It is easy to see that as $\widehat{e}_{f_1}$ varies in Ξ' the intersection point of the line containing this vector with $\text{Affine}(h_2, \ldots, h_d)$, w_1, varies along l. Also, as long as w_1 is not close to w_3, $|\langle \widehat{e}_{f_1} | \widehat{\mathbf{j}^{f_1}}_2 \rangle|$ is large.

Let $\min_{i \in [0,d-1]}(u_g^{(i)}) = \gamma_0$. Notice that $\gamma_0 \leq \|w_3 - u_g\| \leq 2$, $\gamma_0 \leq \|w_2 - u_g\|$ and $\|w_3 - w_2\| \geq \gamma_0$. Since $w_3 \in f_1$, $\frac{1}{\sqrt{d-1}} \leq \|w_3\| \leq 1$ and $\frac{1}{\sqrt{d-1}} \leq \|w_2\|$.

Lemma 40. *Let w_i, $i \in \{1, 2, 3\}$, l and u_g be as above. Let $w \in l$ be such that $\|w - w_3\| \geq \epsilon$ where $\epsilon \leq 1.0$. Let $\mathbf{j}$ be the unit vector determined by w and u_g. Then: $|\langle \widehat{\mathbf{j}^{f_1}}_2 | \mathbf{j} \rangle| \geq \frac{\epsilon \gamma_0}{9}$.*

Proof: We have: $\|w - u_g\| \leq 3$. Let the angle determined by u_g and w_2 at w_3 be θ. It is easy to check that $\sin(\theta) \geq \frac{\gamma_0}{3}$. Let the angle determined by w and w_3 at u_g be α. Then, $\sin(\alpha) \geq \frac{\epsilon \gamma_0}{9}$ and the result follows.

□

Let the unit vector determined by w_3 (w_2) be $\widehat{w_3}$ ($\widehat{w_2}$). Let Ξ be the plane determined by $\widehat{w_2}$ and $\widehat{w_2}$ and $t_0 = \langle \widehat{w_2} | \widehat{w_3} \rangle$.

Lemma 41. $1 - |t_0| \geq \frac{\gamma_0^2}{18d}$.

Proof: Recall that $\|w_2 - w_3\| \geq \gamma_0$. Let the angle subtended at the origin by w_3 and w_2 be θ and $\angle \mathbf{0} w_3 w_2 = \alpha$. Let w_2' be some point on l closest to w_2 such that $\|w_3 - w_2\| = \gamma_0$. The perpendicular distance to l from the origin is at least $\frac{1}{\sqrt{d-1}}$ and the foot of the perpendicular lies at a distance of at most 2 from w_2. It follows that

$\sin(\alpha) \geq \frac{1}{3\sqrt{d-1}}$. Hence $\sin(\theta) \geq \frac{\gamma_0}{3\sqrt{d-1}}$. Hence $|\cos(\theta)| \leq \sqrt{1-\frac{\gamma_0^2}{9d}}$. The result follows because $\sqrt{1-x} \leq 1 - \frac{x}{2}$. □

Let $\widehat{w^{\perp}}_2$ be the unit vector in Ξ perpendicular to $\widehat{w}_2$. Let any unit vector in Ξ be $\widehat{w}(t) = \widehat{w}_2 t + \sqrt{1-t^2}\widehat{w^{\perp}}_2$. Let $w(t)$ be point at the which the line containing $\widehat{w}(t)$ intersects the line l.

Observation 42. *Since the line l is a perpendicular distance of at least $\frac{1}{\sqrt{d-1}}$ from the origin: $\|w(t_0+\epsilon) - w(t_0)\| \geq \frac{\epsilon}{\sqrt{d-1}}$.*

Notice that $w(t_0) = w_3$.

Notice that $\widehat{w}(t)$ lies along $\mathcal{B}' \cap \text{Affine}(h_2, \ldots, h_d, \mathbf{0})$.

Theorem 43. *Let Q, Ξ, $\mathbf{b}_i$, $\widehat{w}(t), \gamma_0$ and l be as above. Then*

$$\Pr_{\mathbf{b}_i \in Q,\ 1 \leq i \leq d-2,\ \mathcal{B}'_{\perp} \cap \text{Affine}(h_2,\ldots,h_d,\mathbf{0}) \in \Xi} [t \in [t_0, t_0+\epsilon]] \leq \frac{Cn^8 d^2}{\sigma^4 \gamma_0^2}\epsilon + 2n^{-2.9d+3.9},$$

where $n > d > 3$, $\sigma < 1.0$, $0 < \epsilon < \frac{\sigma^2\gamma_0^2}{54n^2d^2}$ and C is a constant.

Proof: Follows directly from Theorem 58 and Lemma 61. □

Theorem 44. *Let Q, Ξ, $\mathbf{b}_i$, $\widehat{w}(t), \gamma_0$ and l be as above. Then*

$$\Pr_{\mathbf{b}_i \in Q,\ 1 \leq i \leq d-2,\ \mathcal{B}'_{\perp} \cap \text{Affine}(h_2,\ldots,h_d,\mathbf{0}) \in \Xi} \left[|\langle \widehat{\mathbf{j}_2^{f_1}} | \widehat{e}_{f_1} \rangle| \leq \epsilon \right] \leq \frac{Cn^8 d^{2.5}}{\sigma^4 \gamma_0^3}\epsilon + 2n^{-2.9d+3.9}.$$

Proof: Follows from Lemma 40, Observation 42 and Theorem 43. □

Combining Theorems 44 and 37 we have

Theorem 45.

$$\Pr_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}} \left[|\langle \widehat{e}_{f_1} | \widehat{j^{f_1}}_2 \rangle| \leq \epsilon | E(u_g) \wedge S(e_f) \right] \leq \frac{Cn^{32}d^{13}}{\sigma^{16}}\epsilon + \frac{1}{d} + 4n^{-2.9d+3.9},$$

where C is some constant, $\sigma < 1.0$ and $n > d > 3$.

Putting Theorems 45 and 37 together we have:

Theorem 46.

$$\Pr_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[|\,||e_f||\,\langle\widehat{e}_{f_1}|\widehat{jf_1}_2\rangle| \le \epsilon^2 |E(u_g)\wedge S(e_f)\right] \;\le\; \frac{Cn^8d^{3.5}}{\sigma^4}\,\epsilon + \frac{Cn^{32}d^{13}}{\sigma^{16}}\,\epsilon + \frac{1}{d} + 6n^{-2.9d+3.9},$$

where C is some constant, $\sigma < 1.0$ and $n > d > 3$.

From this it follows that

Theorem 47.

$$\mathop{\mathbf{E}}_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[|\,||e_f||\,\langle\widehat{e}_{f_1}|\widehat{jf_1}_2\rangle|\;\;|E(u_g)\wedge S(e_f)\right] \ge \frac{C\sigma^{16}}{n^{64}d^{28}},$$

where C is some constant, $\sigma < 1.0$ and $n > d > 3$.

3.9 Technical Details: The Effect of Perturbations

We prove our main technical result in this section. The following Lemmas from [ST04] will prove useful.

Lemma 48 ([ST04])**.** *Let $k \ge 0$ and t be a non-negative random variable with density proportional to $\mu(t)t^k$ such that for some $t_0 > 0$, $\frac{\max_{0\le t\le t_0}\mu(t)}{\min_{0\le t\le t_0}\mu(t)} \le c$, then, $\mathbf{Pr}[t \le \epsilon] \le c\left(\frac{\epsilon}{t_0}\right)^{k+1}$.*

Lemma 49 ([ST04])**.** *Let $\mathbf{x}$ be a d-dimensional Gaussian vector with standard deviation σ centered at $\overline{\mathbf{a}}$. Then:*

1. *For any point $\mathbf{p}$, $\mathbf{Pr}[\|\mathbf{x}-\mathbf{p}\| \le \epsilon] \le \left(\frac{\epsilon}{\sigma}\right)^d$.*
2. *For any plane H, of dimension h, $\mathbf{Pr}[dist(x,H) \le \epsilon] \le \left(\frac{\epsilon}{\sigma}\right)^{d-h}$.*

Lemma 50 ([ST04]). *Let* $\mathbf{x}$ *be a* d*-dimensional Gaussian vector of standard deviation* σ *centered at the origin. Then:*

$$\Pr_x [\|x\| \geq k\sigma] \leq \frac{k^{d-2} e^{\frac{-k^2+d}{2}}}{d^{\frac{d}{2}}}$$

Lemma 51 ([ST04]). *Let* μ *be a Gaussian distribution of standard deviation* σ *centered at* $\overline{\mathbf{a}}$*. Let* $\mathbf{v}$ *be some vector and* r *a real. Then, the induced distribution* $\mu(\mathbf{x} \mid \mathbf{v}^T\mathbf{x} = r)$ *is a Gaussian distribution of standard deviation* σ *centered at the projection of* $\overline{\mathbf{a}}$ *onto the plane* $\{\mathbf{x} : \mathbf{v}^T\mathbf{x} = r\}$.

Lemma 52 ([ST04]). *Let* $\mathbf{x}$ *be a Gaussian random vector in* $\mathbb{R}^d$ *of standard deviation* σ *centered at the origin. Then, for* $n \geq 3$, $\mathbf{Pr}[\|\mathbf{x}\| \geq 3\sqrt{d \log n}\sigma] \leq n^{-2.9d}$. *More generally*

$$\mathbf{Pr}[\|\mathbf{x}\| \geq \alpha\sqrt{d}\sigma] \leq e^{-\alpha^2 \frac{d}{4}},$$

where $\alpha > 2$.

Lemma 53 ([ST04]). *Let* $\mu(\mathbf{x})$ *be a Gaussian distribution of standard deviation* σ *centered at a point* $\overline{\mathbf{a}}$*. Let* $k \geq 1$, *let* $\|\mathbf{x} - \overline{\mathbf{a}}\| \leq k$ *and* $\|x - y\| < \epsilon \leq k$. *Then,* $\frac{\mu(\mathbf{y})}{\mu(\mathbf{x})} \geq e^{\frac{-3k\epsilon}{2\sigma^2}}$.

Let $\boldsymbol{\omega}$ and $\mathbf{q}$ be unit vectors in d-dimensions and let $c = \langle \boldsymbol{\omega} | \mathbf{q} \rangle$. Then, $\boldsymbol{\omega}$ can be represented in the following form: $\boldsymbol{\omega} = (c, \boldsymbol{\psi}\sqrt{1 - c^2})$ where $\boldsymbol{\psi}$ is the unit vector obtained by projecting $\boldsymbol{\omega}$ onto the $(d-1)$-plane perpendicular to $\mathbf{q}$.

Lemma 54 ([ST04]). *The Jacobian of the change of variables from* $\boldsymbol{\omega}$ *to* $(c, \boldsymbol{\psi})$ *is given by*

$$\left| det \left(\frac{\partial(\boldsymbol{\omega})}{\partial(c, \boldsymbol{\psi})} \right) \right| = (1 - c^2)^{\frac{d-3}{2}}.$$

As before, we let $B' = \{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}\}$ be Gaussian vectors, with standard deviation σ, in n dimensions with $\mathbf{b}_i$ centered at $\overline{\mathbf{b}}_i$. Let Q be some fixed $(d-1)$-dimensional plane

through the origin and for any point (or vector) $\mathbf{p}$ let $\mathbf{p}^Q$ denote the image of $\mathbf{p}$ on Q under orthogonal projection.

Consider the case where $\mathbf{b}_i \in Q,\ 1 \leq i \leq d-2$. In this case, $Span(B')$ is a $(d-2)$-plane in Q. We let $\boldsymbol{\omega} \in Q$ be the unit vector perpendicular to this $(d-2)$-plane and denote the $n-d+1$ space perpendicular to Q (and containing the origin) by $Q_\perp$.

We fix some (arbitrary) coordinate system for $Span(B')$ and let $C' = \{\mathbf{c}_1, \ldots, \mathbf{c}_{d-2}\}$ be the coordinates of the points of B' (within $Span(B')$) under this coordinatization. It is clear that every point $\mathbf{b}_i$ can be specified by $\boldsymbol{\omega}$ and $\mathbf{c}_i$ (recall that $\mathbf{b}_i \in Q$). The Jacobian of this change of variables was given by Blaschke [ST04, San02].

Theorem 55. *Let $\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}$, $\boldsymbol{\omega}$ and $\mathbf{c}_1, \ldots, \mathbf{c}_{d-2}$ be as described above. Then*

$$d\mathbf{b}_1 \ldots d\mathbf{b}_{d-2} = (d-3)! Vol(\Delta(\mathbf{0}, \mathbf{c}_1, \ldots, \mathbf{c}_{d-2})) d\boldsymbol{\omega} d\mathbf{c}_1 \ldots d\mathbf{c}_{d-2},$$

where $\Delta(\ldots)$ denotes the simplex determined by the arguments.

Recall that the initial points $\mathbf{a}_1, \ldots, \mathbf{a}_n$ were centered at points $\overline{\mathbf{a}}_i$ such that $\|\overline{\mathbf{a}}_i\| \leq 1$. It follows that $\|\overline{\mathbf{b}}_i\| \leq n$ and so $\|\overline{\mathbf{b}}_i^Q\| \leq n$. From Lemma 51 it follows that the induced distribution of $\mathbf{b}_i$ conditioned on the event that $\mathbf{b}_i \in Q$ is a Gaussian with standard deviation σ centered at $\overline{\mathbf{b}}_i^Q$. Similarly from Lemma 52 it follows that

$$\mathop{\mathbf{Pr}}_{\mathbf{b}_i \in Q}\left[\|\mathbf{b}_i - \overline{\mathbf{b}}_i^Q\| \geq 3\sqrt{d \log n}\sigma\right] \leq n^{-2.9(d-1)}.$$

Hence we have:

Lemma 56.

$$\mathop{\mathbf{Pr}}_{\mathbf{b}_i \in Q}\left[\|\mathbf{b}_i\| \geq 2n\right] \leq n^{-2.9(d-1)}.$$

Hence $\|\mathbf{b}_i\| \leq 2n$ for all $1 \leq i \leq d-2$ with an "error" probability of $n^{-2.9d+3.9}$.

Let $\mathbf{q} \in Q$ be some fixed unit vector and let Λ be some fixed 2-plane containing $\mathbf{q}$. Let $\mu(\mathbf{p})$ be a Gaussian distribution of standard deviation σ centered at $\mathbf{p}$.

Theorem 57. *Let $c_0 \in [-1, 1]$ and Q, $\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}$, and $\boldsymbol{\omega}$ be as above. Let $\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}$ be restricted to vary in such a manner that:*

1. $\mathbf{b}_i \in Q,\ 1 \leq i \leq d-2$.

2. $\boldsymbol{\omega} \in \Lambda$.

Then,

$$\Pr_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2} \in Q, \boldsymbol{\omega} \in \Lambda} [\langle \boldsymbol{\omega} | \mathbf{q} \rangle \in [c_0, c_0 + \epsilon]] \leq C \frac{\epsilon}{\epsilon_0},$$

where $n > d > 3$, $\sigma < 1.0$, $0 < \epsilon < \frac{\sigma^2(1-|c_0|)}{3n^2 d} = \epsilon_0$ and C is some constant.

Proof: First we use the change of variables described in Theorem 55 and specify the variables $\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}$ in terms of $\boldsymbol{\omega}, \mathbf{c}_1, \ldots, \mathbf{c}_{d-2}$. The variables have density proportional to

$$\mu(\overline{\mathbf{b}}_1) \ldots \mu(\overline{\mathbf{b}}_{d-2})(d-3)! Vol(\Delta(\mathbf{0}, \mathbf{c}_1, \ldots, \mathbf{c}_{d-2})).$$

We now fix $\mathbf{c}_1, \ldots, \mathbf{c}_{d-2}$ so that the only varying quantity is $\boldsymbol{\omega}$ which is permitted to vary within Λ. We now use the change of variables given by Lemma 54. The density is now proportional to

$$\mu(\overline{\mathbf{b}}_1) \ldots \mu(\overline{\mathbf{b}}_{d-2})(1-c^2)^{\frac{d-4}{2}}, \tag{3.3}$$

where $c = \langle \boldsymbol{\omega} | \mathbf{q} \rangle$. Notice that the other terms (like the volume of the simplex) are constants since they are functions of fixed variables and can be ignored. Let $\Lambda^{\perp}$ be the $(d-3)$-space (within Q) perpendicular to Λ and containing the origin. Let $\mathbf{b}_i = \mathbf{q}_i + \mathbf{t}_i$ where $\mathbf{q}_i$ is the projection of $\mathbf{b}_i$ on Λ. Notice that since we only permit $\boldsymbol{\omega}$ to vary inside Λ, the values of $\mathbf{q}_i,\ \|\mathbf{t}_i\|$ and $\Lambda^{\perp}$ are all fixed and $\mathbf{t}_i \in \Lambda$. Let $\mathbf{q}^{\perp}$ be a vector within Λ perpendicular to $\mathbf{q}$. Then we have, $\boldsymbol{\omega} = \mathbf{q}c + \mathbf{q}^{\perp}\sqrt{1-c^2}$ and $\mathbf{t}_i = \|\mathbf{t}_i\|(\mathbf{q}^{\perp}c - \mathbf{q}\sqrt{1-c^2})$. Since $\|\mathbf{b}_i\| \leq 2n\ \forall i$ it follows that $\|\mathbf{t}_i\| \leq 2n\ \forall i$.

Since we seek to apply Lemma 48, we now compute the change in density as ϵ varies in the range $[0, \epsilon_0]$.

$$\begin{aligned} \frac{1-(c_0+\epsilon)^2}{1-c_0^2} &= 1-\epsilon\left(\frac{\epsilon+2c_0}{1-c_0^2}\right) \geq 1-\frac{3\epsilon}{1-c_0^2} \\ \implies \left(\frac{1-(c_0+\epsilon)^2}{1-c_0^2}\right)^{\frac{d-4}{2}} &\geq \left(1-\frac{3\epsilon}{1-c_0^2}\right)^d \geq 1-\frac{3\epsilon d}{1-c_0^2} \geq 1-\frac{\sigma^2}{n^2}. \quad (3.4) \end{aligned}$$

As for the change in $\mathbf{t}_i$ we have:

$$\delta\mathbf{t}_i = \|\mathbf{t}_i\|\left(\mathbf{q}^\perp\epsilon + \mathbf{q}\left(\sqrt{1-c_0^2}-\sqrt{1-(c_0+\epsilon)^2}\right)\right).$$

where

$$\sqrt{1-c_0^2}\left(1-\sqrt{\frac{1-(c_0+\epsilon)^2}{1-c_0^2}}\right) \leq \sqrt{1-c_0^2}\frac{3\epsilon}{1-c_0^2} = \frac{3\epsilon}{\sqrt{1-c_0^2}} \leq \frac{\sigma^2}{n^2 d}.$$

Hence we have $\|\delta\mathbf{t}_i\| \leq \|\mathbf{t}_i\|\frac{\sqrt{2}\sigma^2}{n^2 d}$. Using Lemma 53, the variation in $\mu(\overline{\mathbf{b}}_i)$ is lower bounded by $e^{-\frac{6n\|\mathbf{t}_i\|\frac{\sqrt{2}\sigma^2}{n^2 d}}{2\sigma^2}} = e^{-\frac{3\sqrt{2}}{d}}$, since the $\|\mathbf{t}_i\| \leq n$. Hence the total variation in $\mu(\overline{\mathbf{b}}_1)\ldots\mu(\overline{\mathbf{b}}_{d-2})$ is lower bounded by $e^{-3\sqrt{2}}$. Putting this and the result of equation 3.4 together, it follows that there exists a constant C such that the variation in density as ϵ varies in the range $[0, \epsilon_0]$ is at most C. We can now apply Lemma 48 and the result follows. □

Theorem 57 shows that we can avoid a "bad" point c_0 with "high" probability. Notice that this result gets weaker and weaker as $|c_0| \to 1$. But this will be sufficient for our purposes.

We now obtain a slightly more involved form of Theorem 57 which will be more convenient for our purposes.

Recall $Span(B')^\perp$ is a $(n-d+2)$-dimensional plane through the origin and so it intersects any $(d-1)$ subspace in a line. Now let T be some $(d-1)$-dimensional plane through the origin and let $\mathbf{t}_1$ be some (fixed) unit vector in T. Let $\mathbf{t}_2$ be the unit vector

along the intersection of $Span(B')^{\perp}$ and T. Now let $f : T \rightarrow Q$ be the map obtained by projecting T onto Q and for any unit vector $\mathbf{t} \in T$ let $\widehat{f(\mathbf{t})}$ be the unit vector along its projection. Now, let Ξ be some 2-plane containing $\mathbf{t}_1$ and $\mathbf{t}_3 \in \Xi$ be a unit vector perpendicular to $\mathbf{t}_1$. We now condition the $\mathbf{b}_i$ so that $\mathbf{t}_2$ lies on Ξ.

Since we wish to extend the result in Theorem 57, we let $\mathbf{q} = \widehat{f(\mathbf{t}_1)}$ and let the 2-plane Ξ be mapped to the 2-plane Λ under f. It is easy to see that $\widehat{f(\mathbf{t}_2)} = \boldsymbol{\omega}$. Every unit vector $\mathbf{t} \in \Xi$ can be represented by $(t, \sqrt{1-t^2})$ where $t = \langle \mathbf{t} | \mathbf{t}_1 \rangle$. Under the mapping f this representation gets mapped to $(c, \sqrt{1-c^2})$ on Λ where $c = \langle \widehat{f(\mathbf{t})} | \mathbf{q} \rangle$. Let $\delta = \langle \widehat{f(\mathbf{t}_3)} | \mathbf{q} \rangle$. We now obtain the Jacobian of f restricted to Ξ.

We have

$$\begin{aligned} c &= \frac{t + \delta\sqrt{1-t^2}}{\sqrt{1 + 2\delta t\sqrt{1-t^2}}}, \text{ and} \\ \frac{dc}{dt} &= \frac{1-\delta^2}{\left(1 + 2\delta t\sqrt{1-t^2}\right)}. \end{aligned} \tag{3.5}$$

Theorem 58. *Let $t_0 \in [-1, 1]$ and Q, $\mathbf{b}_i$, $\boldsymbol{\omega}$, $\mathbf{t}_i, \Xi$ and δ be as above. Let $\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}$ be restricted to vary in such a manner that*

1. $\mathbf{b}_i \in Q,\ 1 \leq i \leq d-2$.

2. $\mathbf{t}_2 \in \Xi$.

Then,

$$\underset{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2} \in Q, \mathbf{t}_2 \in \Xi}{\mathbf{Pr}} \left[\langle \mathbf{t}_2 | \mathbf{t}_1 \rangle \in [t_0, t_0 + \epsilon]\right] \leq \frac{4C}{(1-|\delta|)^2} \frac{\epsilon}{\epsilon_0}, \tag{3.6}$$

where $n > d > 3$, $\sigma < 1.0$, $0 < \epsilon < \frac{\sigma^2(1-|t_0|)}{3n^2 d} = \epsilon_0$ and C is some constant.

Note that the difference between Theorems 57 and 58 really is in Condition 2.

Proof: We proceed exactly as in the proof of Theorem 57 with one difference: we first use the map $f : \Xi \rightarrow \Lambda$ to change the coordinate from t to c. The Jacobian is given

by equation 3.5. So we have

$$c_0 = \frac{t_0 + \delta\sqrt{1-t_0^2}}{\sqrt{1 + 2\delta t_0\sqrt{1-t_0^2}}},$$

from which it follows that

$$(1-|t_0|)\frac{(1-|\delta|)}{2} \leq 1 - |c_0|.$$

This explains a factor of $\frac{2}{1-|\delta|}$ in the RHS of equation 3.6.

The density is now proportional to

$$\mu(\overline{\mathbf{b}}_1)\ldots\mu(\overline{\mathbf{b}}_{d-2})(1-c^2)^{\frac{d-4}{2}}\frac{1-\delta^2}{\left(1+2\delta t\sqrt{1-t^2}\right)}, \tag{3.7}$$

which is the counterpart of equation 3.3.

We have

$$\frac{max_{t\in[-1,1]}\frac{dc}{dt}}{min_{t\in[-1,1]}\frac{dc}{dt}} = \frac{1+|\delta|}{1-|\delta|} \leq \frac{2}{1-|\delta|}.$$

This explains the other $\frac{2}{1-|\delta|}$ in the RHS of equation 3.6. The rest follows from the proof of Theorem 57. □

Projection on Random $(d-2)$-planes

Let $\mathbf{q}$ be a unit vector in n dimensions and $B' = \{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}\}$, and $Q = Span(B')$ be as before. Let $\mathbf{q}^Q$ be the image of $\mathbf{q}$ on Q.

Theorem 59.

$$\Pr_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[\|\mathbf{q}^Q\| \leq \epsilon\right] \leq \left(\frac{2n\epsilon}{\sigma}\right)^{d-2} + n^{-2.9d+3.9}.$$

Proof: We wish to show that projecting $\mathbf{q}$ onto a random $(d-2)$-space does not change its length significantly. This is equivalent to showing that there exists a unit vector in the $(d-2)$-space which is such that projecting it onto $\mathbf{q}$ does not change its length significantly. This is what we do.

Let $\widehat{\mathbf{b}}_i$ be the unit vector along $\mathbf{b}_i$ and $\widehat{\mathbf{q}^Q}$ the unit vector along $\mathbf{q}^Q$. Let $\kappa_0 = |\langle \mathbf{q} | \widehat{\mathbf{q}^Q} \rangle|$ and $\kappa_i = |\langle \mathbf{q} | \widehat{\mathbf{b}}_i \rangle|,\ 1 \leq i \leq d-2$. It is clear that $\kappa_0 \geq \max_{1 \leq i \leq d-2}(\kappa_i)$. Notice that $\kappa_0 = \|\mathbf{q}^Q\|$.

Let $\mathbf{b}_i^q$ be the projection of vector $\mathbf{b}_i$ onto $\mathbf{q}$. It follows that $\mathbf{b}_i^q$ is a Gaussian with standard deviation σ centered at $\overline{\mathbf{b}}_i^q$. Since this is a one dimensional Gaussian it follows that $\underset{\mathbf{b}_i}{\mathbf{Pr}} \left[\|\mathbf{b}_i^q\| \leq \epsilon \right] \leq \frac{\epsilon}{\sigma}$.

From lemma 56 we may assume that $\|\mathbf{b}_i\| \leq 2n$ with an "error" probability of $n^{-2.9d+2.9}$. Hence it follows that $\underset{\mathbf{b}_i}{\mathbf{Pr}} \left[\kappa_i \leq \epsilon \right] \leq \frac{2n\epsilon}{\sigma}$ and since the $\mathbf{b}_i$ are all independent,

$$\underset{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}}{\mathbf{Pr}} \left[\kappa_1 \leq \epsilon \wedge \ldots \wedge \kappa_{d-2} \leq \epsilon \right] \leq \left(\frac{2n\epsilon}{\sigma} \right)^{d-2}.$$

Hence we have

$$\underset{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}}{\mathbf{Pr}} \left[\kappa_0 \leq \epsilon \right] \leq \left(\frac{2n\epsilon}{\sigma} \right)^{d-2} + n^{-2.9d+3.9}$$

and the result follows. □

Theorem 60. *Let $\mathbf{q}_1$ and $\mathbf{q}_2$ be unit vectors in n dimensions and let B' and Q be as before. Let $\mathbf{q}_1^Q$ and $\mathbf{q}_2^Q$ be their projections on Q and $\widehat{\mathbf{q}^Q}_1$ and $\widehat{\mathbf{q}^Q}_2$ denote the unit vectors in these directions. Then*

$$\underset{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}}{\mathbf{Pr}} \left[\frac{\|\widehat{\mathbf{q}^Q}_1 - \widehat{\mathbf{q}^Q}_2\|}{\|\mathbf{q}_1 - \mathbf{q}_2\|} \leq \epsilon \right] \leq \left(\frac{2n\epsilon}{\sigma} \right)^{2(d-2)} + \left(\frac{2n\epsilon}{\sigma} \right)^{d-3} + n^{-2.9d+3.9}.$$

In other words projecting the vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ on Q does not change the angle between them significantly (with high probability).

Proof: Let $\Lambda =\in Span(\mathbf{q}_1, \mathbf{q}_2)$. As before we assume that $\|\mathbf{b}_i\| \leq 2n$ with an "error" probability of $n^{-2.9d+3.9}$. For any unit vector $\widehat{\mathbf{b}}_i$, let m_i denote the magnitude of its projection on Λ and $\widehat{\mathbf{b}^\Lambda}_i$ denote the unit vector along this projection. Using the same reasoning as in the proof of Theorem 59, it is easy to see that we can find a unit vector, say

$\widehat{\mathbf{b}}_j$ which is such that $m_j \geq \epsilon$ with a probability of at least $1 - \left(\frac{2n\epsilon}{\sigma}\right)^{2(d-2)}$. Without loss of generality we assume that this vector is $\widehat{\mathbf{b}}_1$. Let $\mathbf{q}_3 \in \Lambda$ be a unit vector perpendicular to $\widehat{\mathbf{b}^{\Lambda}}_1$ and let L be the $(n-2)$-dimensional plane through the origin perpendicular to $\widehat{\mathbf{b}}_1$ and $\widehat{\mathbf{b}^{\Lambda}}_1$. It follows that $\mathbf{q}_3$ lies in L.

We now project the points $\mathbf{b}_2, \ldots, \mathbf{b}_{d-2}$ on L and apply Theorem 59. Let the new unit vectors (in L) be given by $\widehat{\mathbf{b}^L}_i,\ 2 \leq i \leq d-2$. From Theorem 59 it follows that there exist a unit vector $\widehat{\mathbf{b}^L}_j$ such that $|\langle \mathbf{q}_3 | \widehat{\mathbf{b}^L}_j \rangle| \geq \epsilon$ with probability at least $1 - \left(\frac{2n\epsilon}{\sigma}\right)^{d-3}$. Without loss of generality, we assume that this vector is $\widehat{\mathbf{b}^L}_2$.

Let $c_1 = \langle \widehat{\mathbf{b}^{\Lambda}}_1 | \widehat{\mathbf{b}_1} \rangle$ and $c_2 = \langle \widehat{\mathbf{b}^L}_2 | \mathbf{q}_3 \rangle$. Let $\{\widehat{\mathbf{b}_1}, \widehat{\mathbf{b}^L}_2, \mathbf{l}_3, \ldots, \mathbf{l}_{d-2}\}$ be a basis of $Span(B')$.

Let $\mathbf{q}_1 = \widehat{\mathbf{b}^{\Lambda}}_1 \cos(\theta_1) + \mathbf{q}_3 \sin(\theta_1)$ and $\mathbf{q}_2 = \widehat{\mathbf{b}^{\Lambda}}_1 \cos(\theta_2) + \mathbf{q}_3 \sin(\theta_2)$.

Then the projection of $\mathbf{q}_1$ and $\mathbf{q}_2$ onto $Span(B')$ can be written in the form

$$\begin{aligned}
\mathbf{q}_1^Q &= c_1 \cos(\theta_1) \widehat{\mathbf{b}}_1 + c_2 \sin(\theta_1) \widehat{\mathbf{b}^L}_2 + \langle \mathbf{q}_1 | \mathbf{l}_3 \rangle \mathbf{l}_3 + \ldots, \\
\mathbf{q}_2^Q &= c_1 \cos(\theta_2) \widehat{\mathbf{b}}_1 + c_2 \sin(\theta_2) \widehat{\mathbf{b}^L}_2 + \langle \mathbf{q}_2 | \mathbf{l}_3 \rangle \mathbf{l}_3 + \ldots, \\
\|\widehat{\mathbf{q}^Q}_1 - \widehat{\mathbf{q}^Q}_2\| &\geq \sqrt{c_1^2 \left(\cos(\theta_1) - \cos(\theta_2)\right)^2 + c_2^2 \left(\sin(\theta_1) - \sin(\theta_2)\right)^2},
\end{aligned}$$

where the last inequality follows because $\|\mathbf{q}_1^Q\|, \|\mathbf{q}_2^Q\| \leq 1$ as projection cannot increase the length of a vector.

Hence we have

$$\Pr_{\mathbf{b}_1, \ldots, \mathbf{b}_{d-2}} \left[\frac{\|\widehat{\mathbf{q}^Q}_1 - \widehat{\mathbf{q}^Q}_2\|}{\|\mathbf{q}_1 - \mathbf{q}_2\|} \leq \epsilon \right] \leq \left(\frac{2n\epsilon}{\sigma}\right)^{2(d-2)} + \left(\frac{2n\epsilon}{\sigma}\right)^{d-3} + n^{-2.9d+3.9}.$$

□

Lemma 61. *Let $\mathbf{q}_1$ and $\mathbf{q}_3$ be mutually perpendicular unit vectors in n dimensions and let B' and Q be as before. Let $\mathbf{q}_i^Q$ $i = 1, 3$ be their projections on Q and $\widehat{\mathbf{q}^Q}_i,\ i = 1, 3$ denote the unit vectors in these directions. Then,*

$$\Pr_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[1-|\langle\widehat{\mathbf{q}_1^Q}|\widehat{\mathbf{q}_3^Q}\rangle|\leq\epsilon\right]\leq 4\left(\frac{2\sqrt{2}n\epsilon}{\sigma}\right)^{d-3}+n^{-2.9d+3.9}.$$

Proof: We have $\|\mathbf{q}_3-\mathbf{q}_1\|=\|\mathbf{q}_3-(-\mathbf{q}_1)\|=\sqrt{2}$.

From Theorem 60, it follows that

$$\Pr_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[\|\widehat{\mathbf{q}^Q}_3-\widehat{\mathbf{q}^Q}_1\|\leq\epsilon\vee\|\widehat{\mathbf{q}^Q}_3-(-\widehat{\mathbf{q}^Q}_1)\|\leq\epsilon\right]\leq 4\left(\frac{\sqrt{2}n\epsilon}{\sigma}\right)^{d-3}+n^{-2.9d+3.9}.$$

Hence it follows that:

$$\Pr_{\mathbf{b}_1,\ldots,\mathbf{b}_{d-2}}\left[1-|\langle\widehat{\mathbf{q}^Q}_1|\widehat{\mathbf{q}^Q}_3\rangle|\leq\frac{\epsilon}{2}\right]\leq 4\left(\frac{\sqrt{2}n\epsilon}{\sigma}\right)^{d-3}+n^{-2.9d+3.9}.$$

□

3.10 Conclusions

We have shown that the shadow size of a smoothed polyhedron satisfies partial concentration bounds. There are two improvements possible. First, the bound can be improved by fine tuning the parameters in our theorems. We believe that the bound on the shadow size can be reduced to much smaller powers of n, d and σ^{-1}.

Second, the concentration bound can be extended to all variables. We believe a bound of the form

$$\Pr_{\mathbf{a}_1,\ldots,\mathbf{a}_n}\left[|\mathcal{C}\cap\Pi|\geq Poly(n,d,\sigma^{-1})\right]\leq n^{-d},$$

where $Poly(n,d,\sigma^{-1})$ is some polynomial should be possible.

Chapter 4

LOW RANK DECOMPOSITIONS FOR NOISY DATA

4.1 To Buy Or Not To Buy

Modern commerce has brought a deluge of options to today's consumer. Where one was restricted to what was locally available, be it books in the neighborhood bookstore or movies at the local theater, a few years ago, the power of Internet has now brought figuratively and for all practical purposes, literally, unlimited choice. Indeed it is possible and in fact quite convenient to buy our books or rent our movies from stores located in other parts of the country or even in other parts of the world. This vast increase in options has made it quite difficult for the average consumer to decide *what* exactly to buy or rent. For instance, given a choice of over 50000 movies on NETFLIX, word of mouth or even reading movie reviews is not of much help in deciding which one to rent out.

In light of these developments, modern retailers have become interested in *recommendation systems* which provide personalized recommendations to users based on their preferences. Such systems are used by may e-commerce leaders like AMAZON and NET-

FLIX.

4.2 Overview of Recommendation Systems

Recommendation systems use one of two main approaches: *Content Filtering* or *Collaborative Filtering*. The former creates a profile for each user or product to characterize its nature. For instance a user buying a Baby Einstein Bendy Ball from AMAZON may well be recommended a Munchkin Mozart Magic Cube as they are both aimed at the same target demographic: babies. The major drawback of this technique is that it is hard to scale up as it requires each product to be categorized manually. For instance, the system needs to know that both products mentioned above are intended for babies and it is not hard to see that this kind of (rather specific and detailed) information may be hard to obtain when the inventory of products (or users) is very large.

The other type of recommendation system, *Collaborative Filtering* relies not on specific information about products or users but simply on past user behavior. It stores a list of all users and products they have purchased in the past and uses this to make recommendations. For instance, if many users who purchased a Baby Einstein Bendy Ball also purchased a Munchkin Mozart Magic Cube in the past, then any user purchasing the former would be recommended the latter. Note that the system does not have any prior information about the similarity of these two products. The recommendation is based on past user behavior alone. The most obvious advantages of this system are that (a) it is *domain independent* and (b) it is more easily scalable.

Collaborative Filtering

The two main kinds of Collaborative Filtering are *Neighborhood Methods* and *Latent Factor Models*. The former depends on computing relationships between products (or users). The preference for a product by an user is determined by his preference for similar or *neighboring* products. A product's *neighbors* are those products that are rated similarly by other users. For instance, consider the three Matrix movies: The Matrix, The Matrix Reloaded and The Matrix Revolutions. These three can be identified as *neighbors* because many users who like (or dislike) one also tend to like (or dislike) the others. This similarity in the response of the users can be used to *cluster* these movies together. So to determine if a user would like The Matrix Revolutions we just look at his ratings (if any) for the other two movies. Of course, instead of looking at neighboring products one may look at neighboring users and the preferences of a user can be inferred by looking at the preferences of users *similar* to him. Note that these measures of *similarity* are extracted from past user behavior alone and not from detailed product descriptions (as would be the case in Content Filtering).

Latent Factor Models

Latent Factor Models are an alternative approach that tries to model user preferences and product characteristics by assuming that there are a small number of factors determining them. For instance, the *space* of all *movies* can be modeled as a combination of a few (say around 50 to 100) factors like *Action, Adventure, Animation, Comedies, Horror* etc. This set of factors can be considered as a *basis* for *movie-space* and any movie can be expressed as a tuple of weights $m = (m_i, \ldots, m_k)$ where m_i measures the *content* (or *strength*) of the i^{th} factor. For instance, the movie Rambo would receive a large weight for the *Action* factor and a small weight for *Comedy* or *Romance* while Ratatouille would

receive large weights for *Animation* and *Comedy* and low weight for *Action*.

A similar tuple $u = (u_1, \ldots, u_k)$ can be assigned to any given user as well where u_i captures the users *interest* or *preference* for the i^{th} factor. For instance, a user who likes action movies and dislikes comedies would have a large weight for *Action* and a low weight for *Comedy*.

Once this assignment is complete, to find the preference of user u for movie m, we just compute the dot product of u and m. Hence, in this model every movie and user is a k-tuple (where k is small and independent of the number of movies and users). It follows that the preference (or *rating* matrix) $\mathbf{R}$ where $\mathbf{R}_{i,j}$ is the *rating* by user i of movie j is given by $\mathbf{R} = \mathbf{U}\mathbf{M}^{\mathrm{T}}$ where $\mathbf{U}$ is a $n \times k$ matrix with row i corresponding to the i^{th} user and $\mathbf{M}$ is a $m \times k$ matrix with row j corresponding to j^{th} movie and n and m are the total number of users and movies respectively.

While working with real world data, the idea that only a few factors explain all user preferences is only an approximation at best so the equation above is more accurately written as $\mathbf{R} \approx \mathbf{U}\mathbf{M}^{\mathrm{T}}$

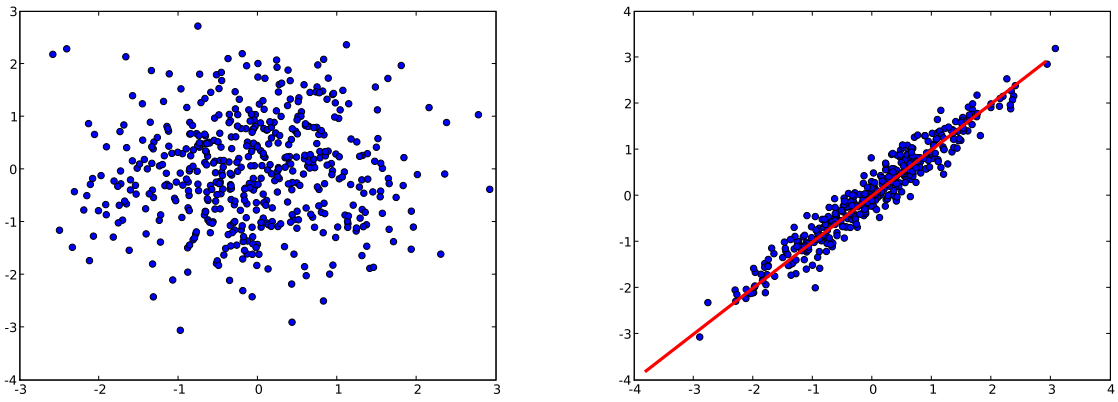
So the task of Collaborative Filtering using Latent Factor Models boils down to the following *Matrix Factorization* problem:

Matrix Factorization 1. *Given a partial rating matrix* $\mathbf{R}$ *(*i.e., *not all entries of* $\mathbf{R}$ *are known) with* n *rows and* m *columns find* factor *matrices* $\mathbf{U}$ *(*$n \times k$*) and* $\mathbf{M}$ *(*$m \times k$*) such that a "good" approximation of* $\mathbf{R}$ *can be obtained by* $\mathbf{R} \approx \mathbf{U}\mathbf{M}^{\mathrm{T}}$.

The main appeal of this technique is that the value of k (the number of factors) is typically much smaller than n or m. A factorization of this form with a "low" value of k is called a *low rank approximation (LRA)* of $\mathbf{R}$. This is also known as *Principal Component Analysis*.

A Geometric View of LRA

We give a brief intuitive description of LRA in this section. First consider the case where all entries of $\mathbf{R}$ are known and let $\mathbf{R}$ be a $n \times 2$ matrix. Since each row has two columns we can consider each row to be a point in two dimensions. Figure 4.1 shows two different 500×2 matrices. The points corresponding to the one with rank 1 (4.1b) lie (approximately) on a lower dimensional linear manifold (in this case a line). Hence for this matrix a good LRA (with $k = 1$) can be found (and is shown as a solid line) while such an approximation does not exist[1] for the matrix in (a).



(a) (b)

Figure 4.1: Rows of two 500×2 matrices plotted as points in the plane. The matrix in (a) has rank 2 and the one in (b) has rank (approximately) 1. The rank 1 LRA of (b) is shown as a solid line.

The intuition from this case carries over to higher dimensions. Let $\mathbf{R}$ be a $n \times m$ matrix. The rows (columns) of $\mathbf{R}$ can be thought of as points in m (n) dimensions. The main argument behind LRA is that if $\mathbf{R}$ were a rating matrix then these points lie on or close to some linear k dimensional manifold. How "good" the LRA is depends on how well this assumption is satisfied.

[1]More precisely, a "good" approximation does not exist. Any rank 1 approximation for (a) will have large error.

Now consider the case where not all entries of $\mathbf{R}$ are known. For instance, if in Figure 4.1b some entries are missing, it may still be possible to reconstruct the matrix once we have retrieved the line from the known entries. Indeed, given one entry of a row and the line, the other entry can be obtained by assuming that the point lies on the line. This, in a nutshell, is the idea behind using LRA for Collaborative Filtering.

4.3 Overview of Matrix Factorization

Given any matrix $\mathbf{R}$ and factors $\mathbf{U}$ and $\mathbf{M}$ we need some measure of how "good" the approximation is. We use the Frobenius norm of the error matrix.

Definition 62. *Given any $n \times m$ matrix $A = [a_{i,j}]$, the* Frobenius *norm of A is*

$$Fr(A) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i,j}^2.$$

For a given $\mathbf{U}$ and $\mathbf{M}$, the error matrix is given by $\mathbf{R} - \mathbf{U}\mathbf{M}^T$ and the error function $E(\mathbf{U}, \mathbf{M})$ is its Frobenius norm.

Let $\mathcal{A}^0 = \{(i,j) | 1 \leq i \leq n, 1 \leq j \leq m\}$ and $\mathcal{A} \subseteq \mathcal{A}^0$ be the subset of $\mathbf{R}$ that is known and let $E_{\mathcal{A}}(\mathbf{U}, \mathbf{M})$ be the error function restricted to $\mathcal{A}$ (*i.e.,* terms of $E(\mathbf{U}, \mathbf{M})$ that correspond to unknown entries of $\mathbf{R}$ are dropped).

First, notice that if all entries of $\mathbf{R}$ are known then the LRA of a given rank with lowest error can be obtained from the *Singular Value Decomposition (SVD)* of $\mathbf{R}$, [SJ03]. Let $\mathbf{R} = \mathbf{X}\Sigma\mathbf{Y}^T$ be the SVD of $\mathbf{R}$ then the LRA of rank k with lowest (Frobenius) error can be obtained by simply taking the columns of $\mathbf{X}, \mathbf{Y}$ and Σ corresponding to the largest k singular values. Hence $E_{\mathcal{A}^0}(\mathbf{U}, \mathbf{M})$ can be minimized by first computing the SVD. Of course, this cannot be done if some entries of $\mathbf{R}$ are unknown as the SVD cannot be computed in this case.

When not all entries of $\mathbf{R}$ are known, a natural generalization of the above approach would be to find the $\mathbf{M}$ and $\mathbf{U}$ that minimize $E_{\mathcal{A}}(\mathbf{U}, \mathbf{M})$. While dropping the unknown entries from the error function might seem like a trivial change, it has rather significant consequences. The properties of this problem[2] were investigated in [SJ03, CT09]. They showed the following:

1. Finding the $\mathbf{R}$ with minimum rank such that some (fixed) entries of $\mathbf{R}$ have known value is *NP*-hard.

2. The error function $E_{\mathcal{A}}(\mathbf{U}, \mathbf{M})$ has only one local minima (which is also the global minimum) if $\mathcal{A} = \mathcal{A}^0$ but this function can have multiple local minima otherwise .

3. When $\mathcal{A} = \mathcal{A}^0$ the optimal $\mathbf{U}$ and $\mathbf{M}$ have a greedy substructure. In other words, a solution of rank k is contained in a solution of rank $k + 1$. This is lost when $\mathcal{A} \neq \mathcal{A}^0$.

Result 2 is particularly significant because it implies that finding the (global) minimum of $E_{\mathcal{A}}(\mathbf{U}, \mathbf{M})$ is not easy. This difficulty is exacerbated by the presence of noise. In fact, in real world data sets, noise is a big enough factor that minimizing $E_{\mathcal{A}}(\mathbf{U}, \mathbf{M})$ can lead to errors in reconstructing $\mathbf{R}$. Regularization factors need to be taken into account to deal with noise. We explain more in the next few sections.

4.4 The NETFLIX PRIZE

On October 2, 2006, NETFLIX, an online movie rental company started a contest to improve the internal recommendation system it used[3]. It released a huge dataset and asked

[2] More precisely, a closely related problem, the *Weighted* LRA, was investigated in [SJ03].
[3] See `http://www.netflixprize.com/`

contestants to improve upon the accuracy of its own internal algorithms by 10%. The contest ended on September 18, 2009 and led to many new results in the field of Collaborative Filtering.

The data set contains over 100 million *ratings* by over $480,000$ users of over $17,780$ movies. A *rating* is an integer in the range $[1,5]$ where 1 indicates that the user did not like the movie and 5 indicates that he did. The data is very *sparse* in the sense that the ratings are available for very few $(user,\ movie)$ pairs. In fact, only about 1.1% of such ratings are known. While the average user rated 200 movies and the average movie was rated by over 5000 users, the data has high variance and some movies have as few as 3 ratings and some users have rated more than $10,000$ movies.

The challenge is to predict (some of) the unknown ratings by using the ones that are known.

The data also contains a quiz/test set, around 2.8 million $(user,\ movie)$ pairs[4] whose rating NETFLIX did not release. Of the total, half form a *quiz* set and the other half form a *test* set. The exact split was also not described by NETFLIX. Any prediction of these ratings was compared against the actual ratings (which NETFLIX had) and the *RMSE* of the prediction was computed. The challenge was to get a prediction that would have an RMSE of $\mathbf{0.8572}$ (or lower) on the test set.

Many results dealing with this problem have been published, some of the main ones are [Kor09b, Kor08a, Kor08b, BK07, SM07] and the three papers that describe the final solution that won the contest [Kor09a, TJ09, PC09].

A Brief Overview of the Techniques

Ensemble Techniques, *i.e.,* techniques that take the output of different algorithms and

[4]The data also contains dates which we ignore since we don't use this in any way

combine them (for example by computing an average) have proven to be the most successful in this contest. Indeed, the final solution was an ensemble of three different teams, *Big Chaos [TJ09], BellKor [Kor09a]* and *Pragmatic Theory [PC09]* each of which was an ensemble of many techniques.

As for the individual techniques, many different ones have been tried. The main ones are Neighborhood models and Latent Factor models (see Section 4.2) and many different flavors of these two have been published. Apart from these, models have been developed for the time variation of the ratings, the choice of movies rated by a given user (*i.e.,* the actual rating is ignored, only the fact that a given user rated a given movie is used), various measures of "similarity" between movies/users *etc.*

4.5 Our Results

We propose an improvement to the algorithm used for Matrix Factorization. We describe the basic model in Section 4.6 and its main shortcoming, instability in presence of noise and ill-conditioning in Section 4.7. While many variants of this model exist in practice, we believe that this shortcoming and the solution we propose are both widely applicable.

We show the results produced by our algorithm in Section 4.9.

4.6 Matrix Factorization with Regularization

Finding the optimal $\mathbf{U}$ and $\mathbf{M}$ to minimize $E_{\mathcal{A}}(\mathbf{U}, \mathbf{M})$ would run into difficulty with real world data due to the presence of noise and also due to users (or movies) with very few ratings. For instance, consider a user who has rated just one movie. Clearly depending on the ratings information to find his "preferences" (*i.e.,* the row corresponding to him in $\mathbf{U}$) will not work since there is so little data. To avoid *overfitting* (*i.e.,* being led astray by

the noise in the data) we need some kind of a "prior" view about users and movies to deal with rows/columns with too little information.

A common way of dealing with this is to add penalty terms to $E_{\mathcal{A}}(\mathbf{M}, \mathbf{U})$ that depend on the Frobenius norms of $\mathbf{M}$ and $\mathbf{U}$. The actual "error" function we use is

$$\mathcal{E}_{\mathcal{A}}^{\mathbf{R}}(\mathbf{U}, \mathbf{M}) = \sum_{(i,j)\in\mathcal{A}} (\mathbf{R}_{i,j} - \mathbf{U}_i \mathbf{M}_j^T)^2 + \sum_{i=1}^{n} \lambda_i \|\mathbf{U}_i\|^2 + \sum_{j=1}^{m} \gamma_j \|\mathbf{M}_j\|^2, \tag{4.1}$$

where $\mathbf{X}_k$ denotes row k for any matrix $\mathbf{X}$ and $\| \ldots \|^2$ denotes the Frobenius norm (Definition 62) and λ_i and γ_j are positive weights whose values are determined experimentally.

A formal derivation of this error function as the MAP solution of a model with Gaussian noise was developed in [SM07].

Two main algorithms for minimizing $\mathcal{E}_{\mathcal{A}}^{\mathbf{R}}(\mathbf{U}, \mathbf{M})$ exist. One based on *stochastic gradient descent* [TPNT08] and the other based on solving multiple *alternating least square* problems. We use the latter and describe this in more detail below.

Alternating Least Squares

The function $\mathcal{E}_{\mathcal{A}}^{\mathbf{R}}(\mathbf{U}, \mathbf{M})$ is not convex when both $\mathbf{U}$ and $\mathbf{M}$ are unknown and a closed form solution of the minimum certainly cannot be obtained. But when one of $\mathbf{U}$ or $\mathbf{M}$ is *fixed*, the function becomes quadratic (in the unknown) and a closed form expression for the optimum value of the other parameter can easily be found.

The *Alternating Least Squares* technique for finding the minimum of $\mathcal{E}_{\mathcal{A}}^{\mathbf{R}}(\mathbf{U}, \mathbf{M})$ uses this above idea and proceeds in the following manner:

1. First, let $\mathbf{M} = \mathbf{M}^{(0)}$ where $\mathbf{M}^{(0)}$ is a random matrix of size $m \times k$ and let $i = 0$.

2. Repeat the following steps until some convergence condition is met:

a) Let $\mathbf{U}^{(i)} = \underset{\mathbf{U}}{\operatorname{argmin}}\, \mathcal{E}_{\mathcal{A}}^{\mathbf{R}}(\mathbf{U}, \mathbf{M}^{(i)})$. Notice that this $\mathcal{E}_{\mathcal{A}}^{\mathbf{R}}(\mathbf{U}, \mathbf{M}^{(i)})$ is a quadratic function of $\mathbf{U}$. So a closed form solution for the optimum $\mathbf{U}$ can be found.

b) Let $\mathbf{M}^{(i+1)} = \underset{\mathbf{M}}{\operatorname{argmin}}\, \mathcal{E}_{\mathcal{A}}^{\mathbf{R}}(\mathbf{U}^{(i)}, \mathbf{M})$.

c) Let $i = i + 1$.

Notice that the above technique is not *guaranteed* to find the global optimum. In fact, we will argue empirically in the next section that it does not find a global optimum *most* of the time.

Using the above approach we can find the optimum solution of rank k. While some authors have fixed the value of k in advance and used the above algorithm to find the best solution of rank k, we use above algorithm to find the best rank 1 solution and use this to find the best rank k solution in a *greedy* manner. After finding the first $\mathbf{U}$ and $\mathbf{M}$ of rank 1 we subtract out $\mathbf{U}\mathbf{M}^T$ from[5] $\mathbf{R}$ and run the same algorithm on the residue. We repeat this k times to obtain a solution of rank k.

While this approach has the obvious advantages of being easier to code and having more flexibility since that value of k need not be fixed in advance, it has an additional advantage in that it can be easily modified to deal with the numerical instability described in Section 4.7.

Adding Biases

While directly trying to factorize $\mathbf{R}$ using the approach described in the previous section will work, we add a minor modification that helps improve the solution obtained.

The basic idea behind the LRA approach is of course that a *good* low rank approximation exists. It follows that if by some modification of the data we can make this

[5] We restrict this of course to the *known* entries of $\mathbf{R}$

approximation *better* then the LRA will work better as well. A simple modification is the removal of any *biases* from the data.

Intuitively, consider two users u_1 and u_2 such that u_1 tends to give *low* ratings to all movies even the ones he likes and u_2 tends to give *high* ratings to all movies even the ones he dislikes. So even if both happen to like the same movie, their rating could be different. The same problem could exist for movies as well.

While many different ways of eliminating this bias have been tried, we do this by make the rows and columns have zero mean as described below.

Let μ and ν be vectors of length n and m respectively such that the matrix $\mathbf{R}'$ where $\mathbf{R}'_{i,j} = \mathbf{R}_{i,j} - (\mu_i + \nu_j)$ be such that the rows and columns of $\mathbf{R}'$ add up to 0. Notice that all operations are restricted to known entries of $\mathbf{R}$. Also, notice that since the number of unknowns is the same as the number of variables, a unique solution exists and can be found by solving a matrix equation.

We alternate between finding factors of rank 1 and making the rows and columns have mean zero. This way we build up a LRA of the matrix in a greedy manner.

The complete technique is shown in Algorithm 1. Note that many different variants of LRA are used in practice and the techniques used by other authors may differ from the one shown in Algorithm 1. But we believe that the drawback mentioned in Section 4.7 is shared by all of them. Finally, notice that while Algorithm 1 takes k as input, this is not really necessary. The optimum value of k can well be determined based on the the vectors $\mu^{(t)}, \nu^{(t)}, \mathbf{U}^{(t)}$, and $\mathbf{M}^{(t)}$ say by computing the prediction error on a test set, see Section 4.7.

Input: The Ratings Matrix $\mathbf{R}$ and some integer k.
Output: The *Biased Low Rank Approximation* of $\mathbf{R}$ with rank k.
/* The output consists of $4k$ vectors $\mu^{(t)}, \nu^{(t)}, \mathbf{U}^{(t)}, \mathbf{M}^{(t)},\ 1 \leq t \leq k$. */
$\mathbf{R}^{(1)} \leftarrow \mathbf{R}$;
for *t=1* **to** k **do**

$(\mathbf{R}'^{(t)}, \mu^{(t)}, \nu^{(t)}) \leftarrow$ zeroMeanRowsAndCols($\mathbf{R}^{(t)}$);
/* $\mu^{(t)}$ of length n and $\nu^{(t)}$ of length m are such that $\mathbf{R}'^{(t)}$ where $\mathbf{R}'^{(t)}_{i,j} = \mathbf{R}^{(t)}_{i,j} - (\mu_i^{(t)} + \nu_j^{(t)})$ has row and column mean zero. Recall that all operations are restricted to the known entries of $\mathbf{R}$. See Section 4.6. */

$(\mathbf{U}^{(t)}, \mathbf{M}^{(t)}) \leftarrow$ bestRank1Approx($\mathbf{R}'^{(t)}$) // See Algorithm 2

$\mathbf{R}^{(t+1)} \leftarrow \mathbf{R}'^{(t)} - \left(\mathbf{U}^{(t)}\ \left(\mathbf{M}^{(t)}\right)^T\right)$;
/* Here $\mathbf{U}^{(t)}$ and $\mathbf{M}^{(t)}$ are treated as col vectors and all operations are restricted, as before, to the known entries of $\mathbf{R}$. Recall that $\mathbf{U}^{(t)}$ has size n, $\mathbf{M}^{(t)}$ has size m and $\mathbf{R}$ is a $n \times m$ matrix. */

end

Algorithm 1: Biased Low Rank Approximation

Input: The Ratings Matrix $\mathbf{R}'$ modified as described in Algorithm 1.
Output: Vectors $\mathbf{U}$ and $\mathbf{M}$ such that $\mathcal{E}_{\mathcal{A}}^{\mathbf{R}'}(\mathbf{U}, \mathbf{M})$ is a (local) minimum.

$\mathbf{M}^{(0)} \leftarrow$ randomVector(m) // Random Vector of length m.

$i \leftarrow 0$;

/* Keep iterating until the decrease in the function is less than some pre-set threshold. */
while *Progress towards minimum more than threshold* **do**

$\mathbf{U}^{(i)} \leftarrow \underset{\mathbf{U}}{\operatorname{argmin}}\ \mathcal{E}_{\mathcal{A}}^{\mathbf{R}'}(\mathbf{U}, \mathbf{M}^{(i)})$ // See Section 4.6.

$\mathbf{M}^{(i+1)} \leftarrow \underset{\mathbf{M}}{\operatorname{argmin}}\ \mathcal{E}_{\mathcal{A}}^{\mathbf{R}'}(\mathbf{U}^{(i)}, \mathbf{M})$;

$i \leftarrow i + 1$;

end
return $(\mathbf{U}^{(i-1)}, \mathbf{M}^{(i)})$

Algorithm 2: Best Rank 1 Approximation.

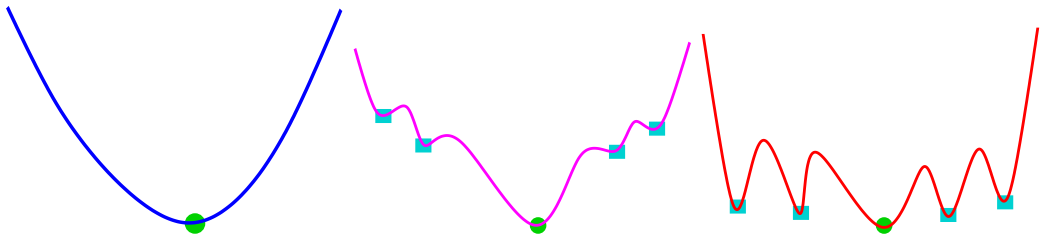
4.7 Of Minima, Local and Otherwise

We discuss the main problem with Algorithm 2 in this section. Recall that the Alternating Least Squares technique used here is not guaranteed to find the global minimum. While other more powerful (say *Conjugate Gradient* [She94] based) algorithms may well find the global minimum, we believe the more pertinent question here is if the global minimum matters at all.

Indeed, we claim that it does not. Let $\min_{(\mathbf{U},\mathbf{M})} \mathcal{E}_{\mathcal{A}}^{\mathbf{R}'}(\mathbf{U},\mathbf{M})$ be the global minimum of $\mathcal{E}_{\mathcal{A}}^{\mathbf{R}'}$. Then,

Claim 63. *The function* $\mathcal{E}_{\mathcal{A}}^{\mathbf{R}'}(\mathbf{U},\mathbf{M})$ *is very* ill-conditioned *in the sense that there are many, quite* different*, local minima* $(\mathbf{U}^{(i)},\mathbf{M}^{(i)}), 1 \leq i \leq r$*, which are such that the*

$$\mathcal{E}_{\mathcal{A}}^{\mathbf{R}'}(\mathbf{U}^{(i)},\mathbf{M}^{(i)}) \approx \min_{(\mathbf{U},\mathbf{M})} \mathcal{E}_{\mathcal{A}}^{\mathbf{R}'}(\mathbf{U},\mathbf{M}), 1 \leq i \leq r.$$



(a) (b) (c)

Figure 4.2: Three functions with different kind of minima. The global minimum in each case is marked with a circle and other minima, if present, are marked with squares. The function in (a) has only one minima which is also the global minimum. In this case, finding the minimum is "easy" (relatively speaking) and a simple gradient descent will do the job. In (b), the global minimum is clearly "better" (*i.e.,* has a much lower function value) that the local minima but it may not be easy to find it. In this case, it may help to invest in a more computationally expensive algorithm to find the global minimum. In (c) the local minima are "as good as" the global minimum and so there is not much to be gained by finding it. This is the case described in Claim 63.

In other words, there are many possible contenders for *optimal* $\mathbf{U}$ and $\mathbf{M}$. A simple illustration of this case (for a one dimensional function) is shown in Figure 4.2. Under

these circumstances we claim that investing in a more powerful/computationally intensive algorithm is not of much help. Indeed, the Alternating Least Squares method described in Algorithm 2 is quite sufficient to find these local minima. To deal with the multiplicity, we propose a simple averaging based approach that works well in practice.

In Section 4.8, we give some empirical evidence in support of Claim 63.

Since there are multiple local minima each of which is "as good " as the global one, it may seem that just arbitrarily picking one of them might be sufficient. Indeed, to the best of our knowledge all authors working on LRA have done exactly this. But we show in Section 4.9 that accuracy of LRA can be significantly increased by dealing explicitly with this multiplicity.

4.8 Ill-Conditioning in Real World Data

In this section we first provide some empirical evidence for Claim 63. For convenience, we use a small subset of the NETFLIX data.

The details of the data set is given in Table 4.1. Among all users with 50 ratings or more $40,000$ were first randomly sampled. Of these many rows (users) and columns (movies) were dropped so that the density of known ratings was close to that of the original data set. We also ensured that after this was done, every user had at least 17 ratings.

This data set was further split into two parts a *Test* set which contains exactly 10 ratings from each user (chosen at random) and a *Training* data set which contains everything else[6].

The idea is to train the models using the Training set and check their efficacy on the Test set. Notice that the Test and Training data are disjoint, *i.e.,* no rating is present in

[6]Note that the original NETFLIX data set was partitioned in a similar manner. See `www.netflixprize.com`

both.

	Number of Users (Rows)	26760
	Number of Movies (Cols)	8845
Training	Number of (known) Ratings	5705874
Training	Density of Ratings	0.0241067
Training	(Min, Average, Max) Ratings per User	(7, 213.2, 4371)
Training	(Min, Average, Max) Ratings per Movie	(27, 645.09, 12581)
Test	Ratings	267600 (10 per user)
Test	(Min, Average, Max) Ratings per Movie	(1, 31.8, 1417)

Table 4.1: The characteristics of the data set used to show evidence of multiple local minima.

Given two sets of vector pairs $L_1 = (\mathbf{U}^{(1)}, \mathbf{M}^{(1)})$ and $L_2 = (\mathbf{U}^{(2)}, \mathbf{M}^{(2)})$ corresponding to two local minima we need a way of determining if they are *different*. There are many ways of doing this. For instance, since these are vectors, we could compute the angle between $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ (and between the two $\mathbf{M}^{(i)}$) after normalizing them. But since our real interest is the performance on the Test & Training data sets, we do the following:

Definition 64.

$$\textbf{TrainingDiff}(L_1, L_2) = \sqrt{\frac{1}{T} \sum_{R_{i,j} \in \textit{ Training Data}} \left(\mathbf{U}_i^{(1)} \mathbf{M}_j^{(1)} - \mathbf{U}_i^{(2)} \mathbf{M}_j^{(2)}\right)^2}$$

where $\mathbf{X}_j^{(i)}$ *is the* j^{th} *element of vector* $\mathbf{X}^{(i)}$ *and* T *is total number of ratings in the Training data set. In other words* **TrainingDiff** *is the RMSE difference in approximation of the training data.* **TestDiff** *is defined in a similar manner on the Test data set.*

To have a baseline for judging the values of $TestDiff$ and $TrainingDiff$, recall that the goal of the NETFLIX contest was to obtain an RMSE of $\mathbf{0.8572}$ on the Test set in the original data set. Also, as we describe in Section 4.9, our techniques enable us to obtain a RMSE less than $\mathbf{0.8900}$. So a $TestDiff$ or $TrainingDiff$ value of 0.01 or more can be considered "*high*" where as a value of around 0.0001 is considered "*low*".

For ease of comparison all numerical function values shown are normalized. In other words we show $\sqrt{\frac{1}{T}\mathcal{E}_{\mathcal{A}}^{\mathbf{R}}}$ instead of just $\mathcal{E}_{\mathcal{A}}^{\mathbf{R}}$ where T is number of ratings in the Training set.

In each table below we show the differences between four different local minima. The values above the diagonal are on the Training data ($TrainingDiff$) and the ones below are on Test data ($TestDiff$). The vectors being compared are the local minima obtained by Algorithm 2 for different values of t in Algorithm 1. In the terminology of Algorithm 1, the minima are the output of Algorithm 2 run four different times (with random start points) for different values of t. Notice that the effects of the multiple local minima are negligible for low values of t but become quite significant as t increases.

Minima for $t = 1$.					
Min	1	2	3	4	Func. Value
	$TrainingDiff$				
1	**0**	0.0000233977	0.00014336	0.000103278	0.885207
2	0.0000222078	**0**	0.000140455	0.000113472	0.885207
3	0.000125653	0.000123273	**0**	0.000175208	0.885207
4	0.0000916749	0.000103449	0.00015439	**0**	0.885207
	$TestDiff$				

Table 4.2: Local minima for $t = 1$. The (normalized) function value is the same and the differences between the minima are all of the order of 0.0001 and so can be considered "low".

4.9 Stability in the Presence of Ill-Conditioning

In this section we describe a simple way of dealing with the Ill-Conditioning detailed above.

The first question is of course if one needs to explicitly deal with this at all. Indeed, to the best of our knowledge, this phenomenon has not been addressed in practice. But

Minima for $t = 2$.					
Min	1	2	3	4	Func. Value
	$TrainingDiff$				
1	**0**	0.000197014	0.00465571	0.00518868	0.868197
2	0.000181879	**0**	0.00453304	0.00502271	0.868197
3	0.00415142	0.00403386	**0**	0.00234874	0.868197
4	0.00463414	0.00447895	0.00204568	**0**	0.868198
	$TestDiff$				

Minima for $t = 3$.					
Min	1	2	3	4	Func. Value
	$TrainingDiff$				
1	**0**	0.00611607	0.00685102	0.0137033	0.857469
2	0.00559149	**0**	0.00624819	0.00821366	0.857468
3	0.00685371	0.00566217	**0**	0.00936824	0.857469
4	0.0130176	0.00796279	0.00837535	**0**	0.85747
	$TestDiff$				

Minima for $t = 5$.					
Min	1	2	3	4	Func. Value
	$TrainingDiff$				
1	**0**	0.00620409	0.00272794	0.104782	0.840448
2	0.00551914	**0**	0.00351568	0.101353	0.840444
3	0.00244566	0.0031107	**0**	0.10329	0.840446
4	0.0923895	0.0894177	0.0910847	**0**	0.840469
	$TestDiff$				

Table 4.3: Local minima for t=$2, 3$ and 5. Notice that the (normalized) function value is almost the same but the differences between the minima increase with t.

we show (empirically) that addressing this can improve accuracy of the LRA.

Our approach is rather simple. We compute multiple local minima and average them out. This modified version of Algorithm 1 is given in Algorithm 3. We do not modify Algorithm 2. Notice that different invocations of Algorithm 2 in Algorithm 3 differ only in the start vector $M^{(0)}$ which is randomly assigned. This random start point results in the local minima $\left(\mathbf{U}^{(t,l)}, \mathbf{M}^{(t,l)}\right)$ being different.

Minima for $t = 10$.					
Min	1	2	3	4	Func. Value
	$TrainingDiff$				
1	**0**	0.092726	0.0602415	0.0158364	0.810196
2	0.0882396	**0**	0.0715442	0.0925134	0.810197
3	0.057659	0.071235	**0**	0.0579986	0.810221
4	0.0151036	0.0880427	0.0553639	**0**	0.810201
	$TestDiff$				

Table 4.4: Local minima for $t = 10$. The (normalized) function value is almost the same but the differences between the minima are quite "high".

Finally, note that while Algorithm 3 takes k and r as input they can also be determined dynamically at run time.

Input: The Ratings Matrix $\mathbf{R}$ and integers k and r.
Output: The *Stable* Biased Low Rank Approximation of $\mathbf{R}$ with rank k.
/* The output consists of $2k(r+1)$ vectors in total of which $2k$ are vectors $\mu^{(t)}, \nu^{(t)},\ 1 \leq t \leq k$ and $2rk$ are vectors $\mathbf{U}^{(t,l)}, \mathbf{M}^{(t,l)},\ 1 \leq t \leq k,\ 1 \leq l \leq r$. */
$\mathbf{R}^{(1)} \leftarrow \mathbf{R}$;
for *t=1* **to** k **do**

$(\mathbf{R}'^{(t)}, \mu^{(t)}, \nu^{(t)}) \leftarrow$ zeroMeanRowsAndCols($\mathbf{R}^{(t)}$);
/* $\mu^{(t)}$ of length n and $\nu^{(t)}$ of length m are such that $\mathbf{R}'^{(t)}$ where $\mathbf{R}'^{(t)}_{i,j} = \mathbf{R}^{(t)}_{i,j} - (\mu^{(t)}_i + \nu^{(t)}_j)$ has row and column mean zero. Recall that all operations are restricted to the known entries of $\mathbf{R}$. See Section 4.6. */

for *l=1* **to** r **do**
$(\mathbf{U}^{(t,l)}, \mathbf{M}^{(t,l)}) \leftarrow$ bestRank1Approx($\mathbf{R}'^{(t)}$) // See Algorithm 2
end

$\mathbf{R}^{(t+1)} \leftarrow \mathbf{R}'^{(t)} - \frac{1}{r}\left(\sum_{l=1}^{r}\left(\mathbf{U}^{(t,l)}\ \left(\mathbf{M}^{(t,l)}\right)^T\right)\right)$;
/* Here $\mathbf{U}^{(t,l)}$ and $\mathbf{M}^{(t,l)}$ are treated as col vectors and all operations are restricted, as before, to the known entries of $\mathbf{R}$. Recall that $\mathbf{U}^{(t,l)}$ has size n, $\mathbf{M}^{(t,l)}$ has size m and $\mathbf{R}$ is a $n \times m$ matrix. */
end

Algorithm 3: Stable Low Rank Approximation

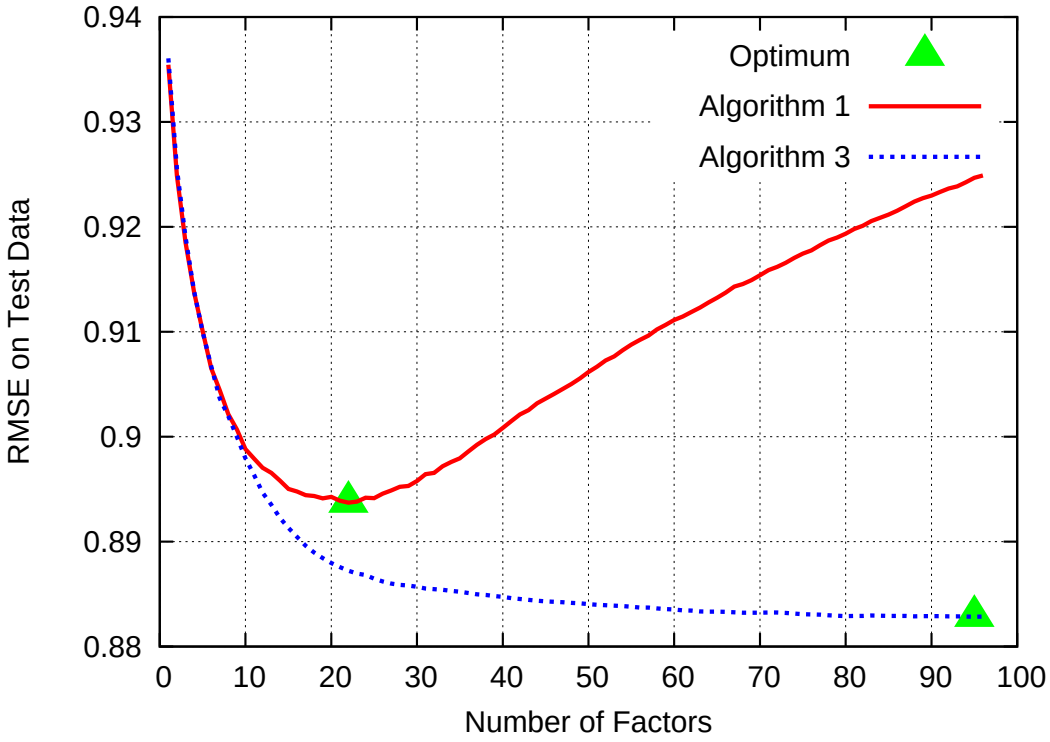


Figure 4.3: The RMSE of the output of Algorithm 1 and Algorithm 3 on the Test set. Notice that the errors are comparable for small number of factors but that Algorithm 3 is more stable and has lower error as the number of factors increases. The lowest error Algorithm 1 is 0.89370 with 22 factors while Algorithm 3 attains an error of 0.882839 with 95 factors.

We compare the results of Algorithm 1 and Algorithm 3 in Figure 4.3 where the RMSE (over the Test set) of the output of each is shown. While the errors are comparable for few factors, Algorithm 3 is clearly more stable and manages to obtain an error that is lower by around 0.01, as the number of factors increases.

4.10 Conclusions

We explored the effects of Ill-Conditioning in real world data on the stability of LRA. We showed that multiple local minima, each of which is "as good as" the global one, exist

and came up with a simple technique to deal with this multiplicity.

Our technique is more stable and leads to lower prediction errors. It would be interesting to combine this with the other Ensemble techniques that have been published.

Chapter 5

CONVEXITY IN THE TOPOLOGICAL AFFINE PLANE

5.1 Basic Definitions

The notion of a *topological affine plane* (TAP) $\mathcal{A}$ is most simply defined by means of one of its standard models (see, e.g., [Gru72]): Consider a circle C_∞ in the Euclidean plane; its interior, $\operatorname{int} C_\infty$, constitutes the set $|\mathcal{A}|$ of points of $\mathcal{A}$. For each pair of antipodal points on C_∞, the interior of a simple Jordan arc joining the points (and not meeting C_∞ anywhere else except at its endpoints) is called a *pseudoline*. Suppose we are given, for each pair x, y of points of $|\mathcal{A}|$, a unique pseudoline $\overleftrightarrow{xy}$ containing x and y and depending continuously on x and y in the Hausdorff metric. (Recall that two point sets lie within distance d of each other in the Hausdorff metric if each point of either lies within distance d of some point of the other.) Suppose further that any two of these pseudolines meet (and necessarily cross) at exactly one point, or else that they share their endpoints on C_∞, and that their intersection depends continuously on the two pseudolines. (It can be shown that this last condition follows as a consequence of the previous continuity assumption; see

[Sal68] for a more thorough discussion.) Thus any finite collection of these pseudolines forms what is commonly known as an *arrangement*; the TAP $\mathcal{A}$ can thus be thought of as a two-parameter "continuous arrangement" of pseudolines. If we work in the closed disk $|\mathcal{A}| \cup C_\infty$, rather than in the open disk $|\mathcal{A}|$, and identify antipodal points in C_∞, the same definitions give what is known as a *topological projective plane*.

In $\mathcal{A}$, the terms *pseudoray*, *pseudohalfspace*, *pseudotriangle*, etc. all have the obvious meaning.

We fix a parametrization of the circle C_∞ by a variable θ running from 0 to 2π. If a pseudoline in $\mathcal{A}$ is directed, we may then speak of its *direction* as the value of θ corresponding to its terminal point. The *angle* from one directed pseudoline to another is the length of the counterclockwise arc of C_∞ from the first endpoint to the second, and if x, y, z are three points in $|\mathcal{A}|$, $\angle xyz$ means the angle from $\overrightarrow{yx}$ to $\overrightarrow{yz}$. Two pseudolines are called *parallel* if their endpoints on C_∞ coincide. Given a direction θ, we may therefore speak of a *parallel sweep* having direction θ: choose any pseudoline *not* in direction θ, and join each point on it with the point on C_∞ having direction θ; by the continuity assumption, this gives a continuously varying family of parallel pseudolines. In the same way, we can speak of continuously *rotating* a pseudoline, or alternatively a pseudoray, about a point $x \in |\mathcal{A}|$.

If a directed pseudoline l meets a set Y and Y lies in the closed right pseudohalfplane determined by l, we call l a *left tangent* to Y; similarly for a *right tangent*.

The following facts about arrangements and topological affine planes will be used in the sequel.

1. Every arrangement of eight or fewer pseudolines in a TAP $\mathcal{A}$ is *stretchable*, i.e., there is a homeomorphism of $|\mathcal{A}|$ with the Euclidean plane that maps the pseudolines to straight lines [GP80]; this is not the case, in general, with arrangements of nine or more

pseudolines. (This fact is not used here in an essential way, but we invoke it occasionally to make the situation easier to depict.)

2. There is a homeomorphism from the Euclidean plane to $|\mathcal{A}|$ taking each line to an arc of a circle passing through two antipodal points of C_∞. (The inverse of the mapping

$$(x, y) \mapsto (\frac{x}{1 - x^2 - y^2}, \frac{y}{1 - x^2 - y^2})$$

does this, for example.) This shows that the Euclidean plane is, in particular, a TAP.

3. Every arrangement of pseudolines can be extended to a TAP [GPWZ94]; see also [Kal00].

For basic facts about pseudoline arrangements, see [Goo04] or [Gru72].

Given any two points $x, y \in |\mathcal{A}|$, we can speak unambiguously of the *pseudosegment* $\overline{xy}$. As in the Euclidean plane, we may therefore call a set $Y \subset |\mathcal{A}|$ *convex* if $x, y \in Y \Rightarrow \overline{xy} \subset Y$. Trivially, the intersection of convex sets is convex; this enables us to define, as usual, the convex hull $\operatorname{conv} S$ of a set S as the smallest convex set containing S, i.e., the intersection of all the convex sets containing S.

This notion of convexity enjoys the same basic properties with respect to the underlying set of pseudolines that defines our TAP as ordinary convexity does with respect to straight lines:

Proposition 65.

1. $A \subset \operatorname{conv} A$
2. $A \subset B \Rightarrow \operatorname{conv} A \subset \operatorname{conv} B$
3. $\operatorname{conv} \operatorname{conv} A = \operatorname{conv} A$

We also have

Proposition 66. *If $\mathcal{A}$ and $\mathcal{B}$ are TAPs and $f : |\mathcal{A}| \to |\mathcal{B}|$ is a homeomorphism that maps the pseudolines of $\mathcal{A}$ to those of $\mathcal{B}$, then*

1. X is convex in $\mathcal{A} \Rightarrow f(X)$ is convex in $\mathcal{B}$
2. $x \in \text{conv}\, X$ in $\mathcal{A} \Leftrightarrow f(x) \in \text{conv}\, f(X)$ in $\mathcal{B}$.

Proof: These both follow immediately from the definitions. □

We add that Cantwell [Can74] takes a synthetic-geometric approach to some of the same questions, and obtains several of the same results we do using different (and in some cases more difficult) proofs.

5.2 The Separation Theorem

We begin by establishing the following basic result.

Theorem 67. *Given disjoint compact convex sets X, Y in a topological affine plane, there is a pseudoline l of the plane that separates X from Y.*

The proof proceeds by a sequence of auxiliary results.

Lemma 68. *If Y is a compact convex set and $x \notin Y$, there is a pseudoline l through x that misses Y.*

Proof: Suppose every pseudoline through x met Y. Consider each pseudoray starting at x. Its line meets Y, either on the side of the pseudoray or on the opposite side, but not both (otherwise, by the convexity of Y, we would have $x \in Y$). This defines a function

$f_x : C_\infty \to \{+, -\}$ which, for the same reason, cannot have the same value on a pair of antipodal points of C_∞. Hence for some $\theta_0 \in [0, 2\pi)$, we must have $f_x(\theta) = +$ for a sequence of directions arbitrarily close to θ_0, and $f_x(\theta) = -$ for another such sequence. This means that there is a sequence of points y_i with $\overrightarrow{xy_i}$ in directions θ_i arbitrarily close to θ_0, and another such sequence y_i' with $\overrightarrow{xy_i'}$ in directions arbitrarily close to $\theta_0 + \pi$. By the compactness of Y and the continuity of the pseudoline determined by a pair of points, we therefore get points $y_0, y_0' \in Y$ with $\overrightarrow{xy_0}$ and $\overrightarrow{xy_0'}$ pointing in opposite directions, which again implies that $x \in Y$ by the convexity of Y, a contradiction. □

Lemma 69. *If Y is a compact convex set and $x \notin Y$, there is a unique left tangent pseudoray $\overrightarrow{xy}$ from x to Y.*

Proof: Rotate a pseudoray x_θ in direction $\theta = 0$ to 2π around x. Then x_θ meets Y for some $\theta = \theta_1$ and, by Lemma 68, misses it for some $\theta = \theta_2$. As we rotate counterclockwise from θ_1 to θ_2, we reach (by the compactness of Y) a final direction θ_0 in which there exists a pseudoray xy_0, before we lose this property. Then x_{θ_0} is clearly a left tangent pseudoray from x to Y.

The uniqueness follows from the fact that two pseudolines cannot meet twice. □

Lemma 70. *If X and Y are disjoint compact convex subsets of $|\mathcal{A}|$ with $x_i \in X$ and $y_i \in Y$ for $i = 1, 2$, then the pseudorays $\overrightarrow{x_1y_1}$ and $\overrightarrow{x_2y_2}$ cannot point in opposite directions.*

Proof: Stretching the four pseudolines gives the situation depicted in Figure 5.1. Since lines x_1y_1 and x_2y_2 are parallel, pseudosegments $\overline{x_1x_2}$ and $\overline{y_1y_2}$ would have to cross, with their intersection therefore lying in both X and Y, contradicting the disjointness of these sets. □

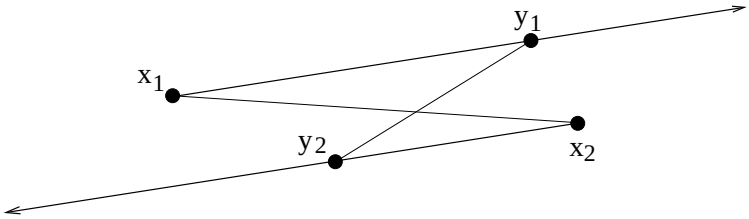


Figure 5.1: Opposite directions.

Lemma 71. *If a pseudoline l misses a compact convex set Y, and p is a point on l, we can rotate l slightly in both directions around p so that the resulting pseudolines still miss Y. The corresponding result holds if p is one of the endpoints of l on C_∞.*

Proof: Let $\delta = \inf\{\mathrm{dist}(y, x) \mid y \in Y, x \in l\}$. Then $\delta > 0$ by the compactness of Y. If we rotate l in either direction around p so that the Hausdorff distance to l remains less than δ, the result follows. (The proof is unchanged if $x \in C_\infty$.) □

Lemma 72. *If Y is a compact convex set and $x \notin Y$, there is a pseudotriangle containing x in its interior, all of whose sides (extended) miss Y, such that Y is contained in a region bounded by only two sides of the pseudotriangle, suitably extended.*

Proof: By Lemma 68, there is a pseudoline l through x missing Y. We may assume, without loss of generality, that l is directed so that Y lies on its right. Choose points u and v on l so that $v < x < u$ (along l), as in Figure 5.2. By Lemma 71, we can rotate l slightly around u, in the counterclockwise direction, so that the resulting directed pseudoline l_u still has Y on its right, and similarly we can rotate l slightly around v, in the clockwise direction, so that the resulting directed pseudoline l_v still has Y on its right. Let m be a pseudoline parallel to l and lying to its left. Let $p = l_u \cap l_v$, $q = l_u \cap m$, and $r = l_v \cap m$. Then pqr is the desired pseudotriangle. □

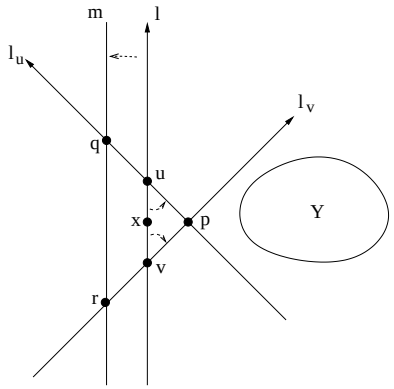


Figure 5.2: Modifying a pseudoline missing Y to a pseudotriangle.

Lemma 73. *Let Y be a compact convex set and l a pseudoline missing Y. For $x \in l$, let $f(x)$ be the direction of the pseudoray from x that is left tangent to Y. Then f is a monotone mapping from l to C_∞. The same conclusion holds if l passes through Y, provided f is restricted to the points of l lying on one side of Y. In both cases, the mapping f is continuous.*

Proof: Consider first the case where l misses the set Y altogether. Direct l so that Y is on its left, and direct C_∞ so that θ increases in the counterclockwise direction. We claim that $f : l \to C_\infty$ is then a monotone increasing function.

Suppose $x_1 < x_2$ on l, and let l_i be the left tangent from x_i to Y for $i = 1, 2$. Stretching l, l_1, and l_2 (Fact 1 above), we obtain the situation shown in Figure 5.3(a). If l_1 did not cross l_2 to the left of l, the portion of l_1 to the left of l would lie entirely below l_2, hence l_1 could not be tangent to Y. Therefore $f(x_1) < f(x_2)$.

A similar argument works in the case where l passes through Y and where $x_1 < x_2$ on the same side of Y, as in Figure 5.3(b) (where $x_1 < x_2$ are both below Y; the corresponding argument works if they are both above).

Finally, suppose $f(x_0) = \theta_0$, and suppose $\theta_1 < \theta_0 < \theta_2$, as in Figure 5.3(c). For $i = 1, 2$, let x_i be the intersecton of l with the (unique) right tangent l_i from θ_i to Y. (This

makes l_i the *left* tangent from x_i to Y.) Then by the above, $f(x)$ lies between θ_1 and θ_2 for every x between x_1 and x_2 on l, so that $f : l \to C_\infty$ is continuous.

□

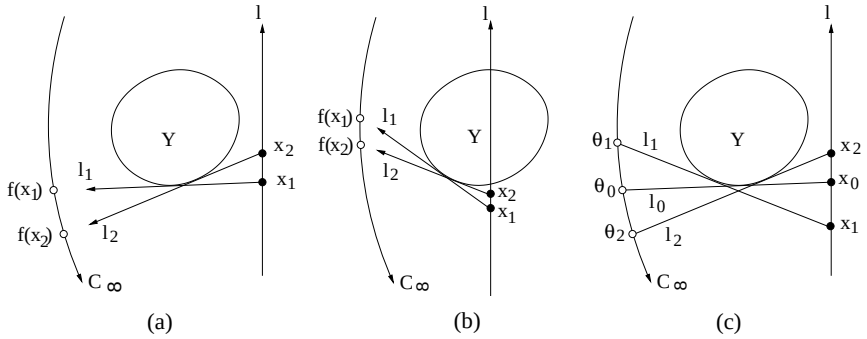


Figure 5.3: Monotonicity of tangent direction along a line.

Theorem 74. *Let Y be a compact convex set. For $x \notin Y$, let $f(x)$ be the direction of the pseudoray from x that is left tangent to Y. Then f is a continuous mapping from the complement of Y to C_∞.*

Proof: Suppose $x \notin Y$. By Lemma 72, there is a pseudotriangle pqr containing x whose sides, $\overleftrightarrow{pq}$, $\overleftrightarrow{pr}$, and $\overleftrightarrow{qr}$, suitably extended, miss Y, with Y contained in the region bounded by only (say) $\overleftrightarrow{pr}$ and $\overleftrightarrow{qr}$, as in Figure 5.4(a). Suppose $f(x) = \theta_0$, and suppose $\theta_1 < \theta_0 < \theta_2$. We must show that there is an open neighborhood of x that is mapped into (θ_1, θ_2) by f.

Consider the pseudosegments $\overline{xp}$, $\overline{xq}$, and $\overline{xr}$. For a real parameter t going from 0 to 1, move a point $p(t)$ monotonically along segment $\overline{xp}$ from p to x, and — for each value of t — consider the parallel translate of pseudoline $\overleftrightarrow{pq}$ passing through $p(t)$. The latter intersects segment $\overline{xq}$ at some point: call it $q(t)$. Now the parallel translate of pseudoline $\overleftrightarrow{pr}$ and that of pseudoline $\overleftrightarrow{qr}$ passing through $p(t)$ and $q(t)$ respectively each meet the pseudosegment $\overline{xr}$ (but not necessarily at the same point: that would require

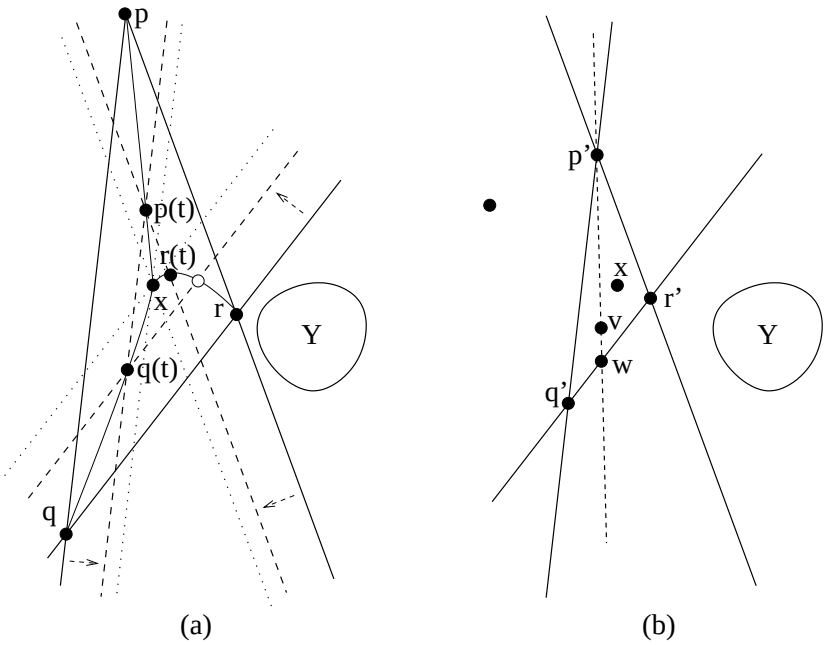


Figure 5.4: Continuity of the left tangent to a fixed set Y.

Desargues's theorem, which is generally false in a TAP!), say at points $r_1(t)$ and $r_2(t)$, respectively. Let $r(t)$ be the one of the these two points which lies closer to x along $\overline{xr}$, as in Figure 5.4(a).

Then by construction, and by the continuity both of a pseudoline as a function of a point pair and of the intersection of two pseudolines as a function of the pseudolines, for t sufficiently close to 1, the points $p(t), q(t), r(t)$ will reach positions p', q', r' with the following properties, as in Figure 5.4(b) :

1. $f(p'), f(q'), f(r') \in (\theta_1, \theta_2)$
2. x belongs to pseudotriangle $p'q'r'$
3. $p'q'$, $p'r'$, and $q'r'$ all miss Y, and Y lies in the region bounded by only $p'r'$ and $q'r'$ (see Figure 5.4(b)).

Then given any point v inside the pseudotriangle $p'q'r'$, there is a pseudoline containing v that passes through vertex p', crosses pseudosegment $\overleftrightarrow{q'r'}$ at a point w between vertices q' and r', and misses the region containing Y, hence misses Y itself. By Lemma 73, we have, first, $f(w) \in (\theta_1, \theta_2)$, and then, finally, $f(v) \in (\theta_1, \theta_2)$. □

We can now complete the proof of Theorem 67.

Proof: For each $x \in X$, consider the (unique) directed left tangent pseudoray l_x from x to Y (see Figure 5.5); it exists, by Lemma 69. This tangent has a direction θ_x. Since, by Lemma 70, the existence of such a pseudoray precludes the existence of another such pointing in the opposite direction, it follows that there is a direction θ_0, $0 < \theta_0 \leq 2\pi$, such that no l_x has direction θ_0. Let

$$\Theta = \{\theta \ , \ \theta_0 - 2\pi \leq \theta < \theta_0 \mid l_x \text{ has direction } \theta \text{ for some } x \in X\},$$

and let $\theta' = \sup \Theta$. Then, by the compactness of X and the continuity of θ_x as a function of x (Theorem 74), there is an $x_0 \in X$ such that l_{x_0} has direction θ'. It follows that l_{x_0} is a right tangent to X, since the existence of a point of X to its right would (as a consequence of Lemma 73) contradict the fact that no direction in Θ exceeds θ'. This shows that X and Y have a "right-left XY tangent" l, i.e., a common tangent that meets X on the right before it meets Y on the left.

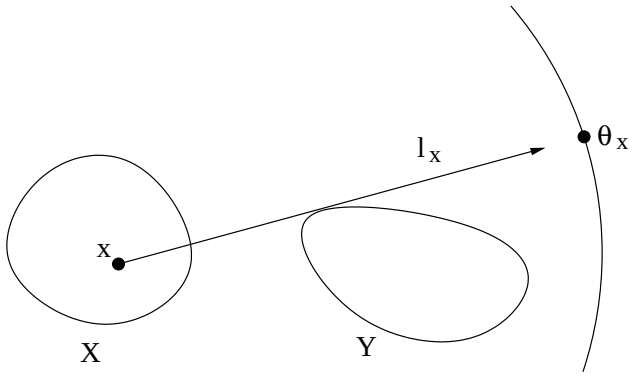


Figure 5.5: The left tangent from point x to set Y.

Finally, to show the existence of a strict separator for X and Y, choose any point $p \in l$ strictly between $l \cap X$ and $l \cap Y$, as in Figure 5.6; such a point must exist by the compactness of X and Y. Rotating l slightly in the counterclockwise direction about p will then produce a strict separator, by Lemma 71. □

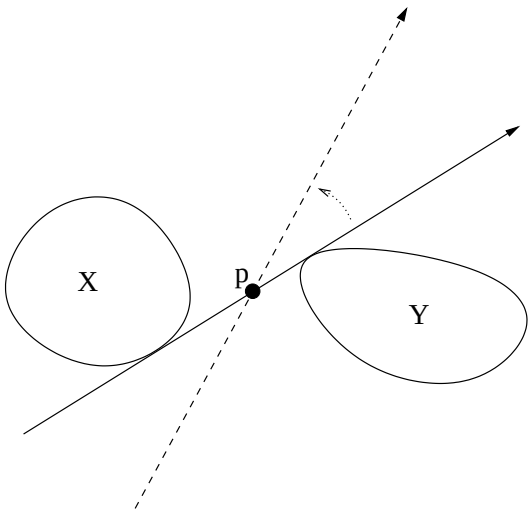


Figure 5.6: A strict separator.

5.3 Some Theorems of Combinatorial Convexity

In this section we generalize a number of standard results in combinatorial convexity to topological affine planes. An excellent single source for the original theorems is [DGK63].

We begin with an extension of Radon's theorem to a TAP.

Theorem 75. *Any set S of cardinality at least 4 in a TAP $\mathcal{A}$ can be partitioned into two subsets whose convex hulls meet.*

Proof: It is enough to prove this if $|S| = 4$. In that case, Fact 1 above allows us to straighten the arrangement of the at most 6 pseudolines joining the four points of S in pairs, and then an application of the standard version of Radon's theorem in $\mathbb{R}^2$ immediately yields the result, by Proposition 66(2) □

The usual proof of Helly's theorem from Radon's works in our situation:

Theorem 76. *If $\{X_i\}_{i\in I}$ is a family of at least three convex sets in a TAP $\mathcal{A}$ with either I finite or each set X_i compact, and if every three of the sets have a nonempty intersection, then all the sets meet.*

Proof: Suppose first that the index set I is finite, say $I = \{1, \ldots, n\}$. We prove the result by induction. It clearly holds for $n = 3$. Supposing $n > 3$, we let $\Xi_j = \{X_i \,|\, i \neq j\}$. By induction hypothesis, the sets belonging to each family Ξ_j, $j = 1, \ldots, n$, contain a point x_j. For each j, we have $x_j \in X_i$ for all $i \neq j$. Since $n \geq 4$, we can apply Radon's theorem to the points $x_1, \ldots, x_n$: they can be partitioned into two sets $\{x_1, \ldots, x_k\} \cup \{x_{k+1}, \ldots, x_n\}$ such that there is a point $x_0 \in \text{conv}\,\{x_1, \ldots, x_k\} \cap \text{conv}\,\{x_{k+1}, \ldots, x_n\}$. But then by Proposition 65, we have $x_0 \in X_i$ for $i = k+1, \ldots, n$ and (respectively) $x_0 \in X_i$ for $i = 1, \ldots, k$, so that the conclusion follows.

If I is infinite, but each X_i is compact, the result follows from the finite case, since any collection of compact sets has the finite intersection property (if every finite collection has a point in common, so does the entire set). □

The generalization of Carathéodory's theorem to TAPs requires a slightly more subtle argument:

Lemma 77. *If x, y_1, y_2, y_3 are four distinct points in $|\mathcal{A}|$ such that the three angles $\angle y_1xy_2$, $\angle y_2xy_3$, and $\angle y_3xy_1$ are all at most π, then $x \in$ conv (y_1, y_2, y_3), and conversely.*

Proof: The situation is as shown in Figure 5.7(a). If, say, $\angle y_1 x y_2 = \pi$, then y_1, x, y_2 are copseudolinear, in that order, so we are done. We may therefore assume that each of the three angles is less than π. Suppose we had $x \notin$ conv (y_1, y_2, y_3). Then by Lemma 68, there would be a pseudoline l through x missing conv (y_1, y_2, y_3), as in Figure 5.7(b), but then one of the three angles in question ($\angle y_3 x y_1$ in the figure) could not be less than π.

For the converse, we need only observe that if, say, $\angle y_3 x y_1 > \pi$, as in Figure 5.7(b), then there is a pseudoline l through x with all of the points y_i lying on the same side of l (in the figure, start with $y_1 x$ and rotate it slightly around x), so that x could not belong to conv (y_1, y_2, y_3). □

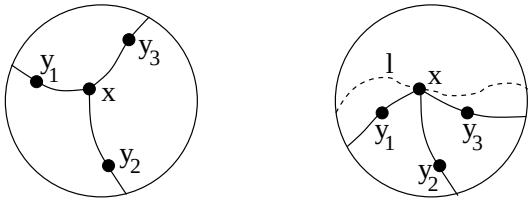


Figure 5.7: $x \in \text{conv}(y_1, y_2, y_3)$.

Theorem 78. *If X is a set of points in a TAP and $y \in$ conv X, then there are points $x_1, x_2, x_3 \in X$ such that $y \in$ conv $\{x_1, x_2, x_3\}$.*

Proof: We assume that $y \notin X$ as the result would be trivially true otherwise. Since $y \in \text{conv} X$ we must have $X \neq \emptyset$. In fact, we can assume that X contains more than three points as as the result follows immediately otherwise. Choose any point $x \in X$.

Let

$$\Theta_L = \{\theta, 0 \leq \theta < \pi \mid \text{there is some } z \in X \text{ with } z \text{ to the left of } \overrightarrow{xy} \text{ and } \angle zyx = \theta\}$$

and

$$\Theta_R = \{\theta, 0 \leq \theta < \pi \mid \text{there is some } z \in X \text{ with } z \text{ to the right of } \overrightarrow{xy} \text{ and } \angle xyz = \theta\}.$$

Let $\theta_L = \sup \Theta_L$ and $\theta_R = \sup \Theta_R$.

If $\theta_L + \theta_R > \pi$, we are done, by Lemma 77: see Figure 5.8.

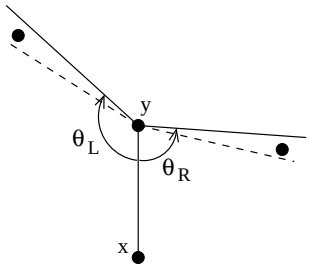


Figure 5.8: $\theta_L + \theta_R > \pi$.

If $\theta_L + \theta_R < \pi$, then by Lemma 77 (converse) we have a contradiction to the assumption that $y \in \text{conv}\, X$ (see Figure 5.9: the convex set shown there contains X but not y).

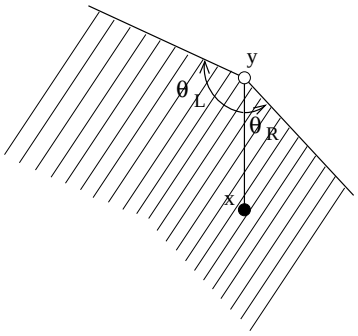


Figure 5.9: $\theta_L + \theta_R < \pi$.

The only remaining case is where $\theta_L + \theta_R = \pi$. Notice first that if θ_L or θ_R is *itself equal to* π, say the former, then either we are done if there is a point $z \in X$ to the right of $\overrightarrow{xy}$, as in Figure 5.10(a), or else we have a contradiction to the fact that $y \in \text{conv}\, X$, as in Figure 5.10(b), where the convex set shown contains X but not y.

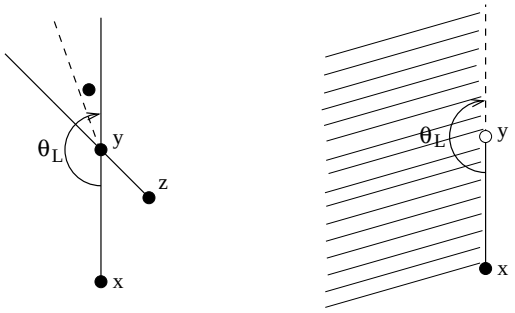


Figure 5.10: $\theta_L = \pi$.

So suppose finally that $\theta_L + \theta_R = \pi$, with both θ_L and $\theta_R > 0$. Let l be the line formed by the two rays in those directions. If there are points $z_1, z_2 \in X$ on line l lying on opposite sides of y, we are done. If not, say there is no point $z \in X$ on one side of y on l, as in Figure 5.11. Then y is not contained in the convex set shown, but this set contains X, which again contradicts the hypothesis. □

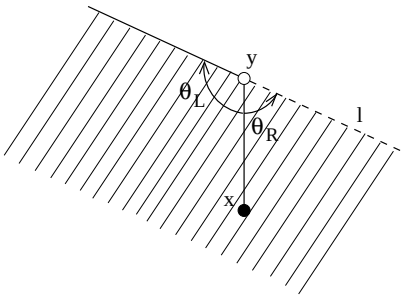


Figure 5.11: $\theta_L + \theta_R = \pi$, $\theta_L > 0$, $\theta_R > 0$.

Notice that a similar argument gives us the existence of a supporting pseudoline at any boundary point of a convex set in a TAP:

Theorem 79. *If X is a convex set in a TAP $\mathcal{A}$, and x is a boundary point of X, then there is a pseudoline l through x with no point of X on one side of it.*

Proof: We may suppose X contains points other than x. Let y be such a point, and consider the pseudoline $\overleftrightarrow{xy}$. If there were a point $z \in X \cap \overleftrightarrow{xy}$ lying on the other side of x from y, as in Figure 5.12(a), then there could not be points of X on both sides of $\overleftrightarrow{xy}$, since otherwise x would be an interior point of X; hence in this case we are done. Let us suppose, then that there is no point of $X \cap \overleftrightarrow{xy}$ lying on the other side of x from y, as in Figure 5.12(b). Then, as in the proof of Theorem 78, let

$$\Theta_L = \{\theta, 0 \leq \theta < \pi \mid \text{there is some } z \in X \text{ with } z \text{ to the left of } \overrightarrow{yx} \text{ and } \angle zxy = \theta\}$$

and

$$\Theta_L = \{\theta, 0 \leq \theta < \pi \mid \text{there is some } z \in X \text{ with } z \text{ to the right of } \overrightarrow{yx} \text{ and } \angle yxz = \theta\}.$$

Let $\theta_L = \sup \Theta_L$ and $\theta_R = \sup \Theta_R$. We cannot have $\theta_L + \theta_R > \pi$, since that would make x an interior point of X. It follows that the pseudoline through x in direction (say) θ_L is a supporting pseudoline for X. □

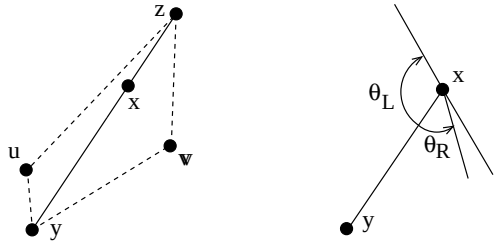


Figure 5.12: Existence of a supporting pseudoline.

As a corollary of Carathéodory's theorem, we obtain yet another basic fact about convex sets in the setting of a TAP:

Corollary 80. *If X is compact, then conv X is also compact.*

Proof: Since X is bounded (i.e., bounded away from C_∞), a parallel sweep in three suitable directions yields a pseudotriangle T containing X. But the interior of T is convex; this shows that $\operatorname{conv} X$, which is inside T, is also bounded.

To show that $\operatorname{conv} X$ is closed, we argue as follows. Suppose $\lim_{n\to\infty} x_n = x$ and $x_n \in \operatorname{conv} X$ for every n. By Theorem 78, for every n, there are points $x_n^1, x_n^2, x_n^3 \in X$ such that $x_n \in \operatorname{conv}(x_n^1, x_n^2, x_n^3)$. Since all the points x_n^1 belong to the compact set X, there is a subsequence $(x_{n_i}^1)$ of (x_n^1) converging to a point x^1 of X, then there is a subsequence $(x_{n_{i_j}}^2)$ of $(x_{n_i}^2)$ converging to a point x^2 of X, and finally there is a subsequence $(x_{n_{i_{j_k}}}^3)$ of $(x_{n_{i_j}}^3)$ converging to a point x^3 of X. Since $x_n \in \operatorname{conv}(x_{n_{i_{j_k}}}^1, x_{n_{i_{j_k}}}^2, x_{n_{i_{j_k}}}^3)$ for every n, we must have $x \in \operatorname{conv}(x^1, x^2, x^3)$, otherwise by the separation theorem for a point and a convex set in a TAP, there would be a pseudoline strictly separating x from conv (x^1, x^2, x^3), and we could not have $x_n \in \operatorname{conv}(x_{n_{i_{j_k}}}^1, x_{n_{i_{j_k}}}^2, x_{n_{i_{j_k}}}^3)$ for all n. □

A second corollary is a generalization of Kirchberger's theorem to TAPs:

Theorem 81. *If X and Y are compact sets of points in a TAP $\mathcal{A}$ with the property that given any four points in their union there is a pseudoline of $\mathcal{A}$ separating those in X from those in Y, then there is a pseudoline of $\mathcal{A}$ separating all of X from all of Y.*

Proof: Suppose not. Since $\operatorname{conv} X$ and $\operatorname{conv} Y$ are compact by Corollary 80, it follows from Theorem 67 that there is a point $z \in \operatorname{conv} X \cap \operatorname{conv} Y$. By Theorem 78, there are points $x_1, x_2, x_3 \in X$ and $y_1, y_2, y_3 \in Y$ such that $z \in \operatorname{conv}(x_1, x_2, x_3)$ and $z \in \operatorname{conv}(y_1, y_2, y_3)$. For convenience, straighten the six pseudolines $\overleftrightarrow{x_1x_2}, \overleftrightarrow{x_1x_3}, \overleftrightarrow{x_2x_3}$, $\overleftrightarrow{y_1y_2}, \overleftrightarrow{y_1y_3}, \overleftrightarrow{y_2y_3}$. This yields two intersecting triangles with vertices x_1, x_2, x_3 and y_1, y_2, y_3 respectively. If any x_i were contained in triangle $y_1y_2y_3$ or any y_i in triangle $x_1x_2x_3$, as in Figure 5.13(a), this would contradict the hypothesis. Since the interiors of the two triangles meet, however, it follows that a side of one must intersect a side of the other, as

in Figure 5.13(b), again contradicting the hypothesis. □

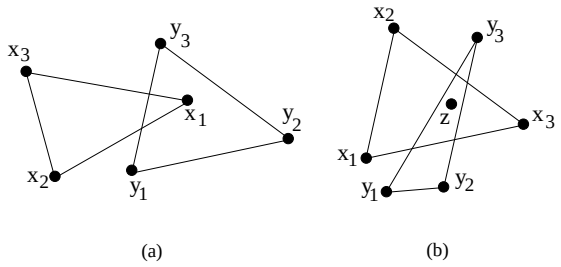


Figure 5.13: Kirchberger's theorem.

Yet another corollary is the following result, which allows us to define extreme points of convex sets in a TAP, and to generalize to TAPs the planar case of the Minkowski theorem on extreme points (see, e.g., [Sch93], p.276).

Proposition 82. *If X is a convex set in a TAP $\mathcal{A}$ and $x \in X$, the following are equivalent:*

1. $X \setminus \{x\}$ is convex

2. There are no points $y, z \in X \setminus \{x\}$ such that x lies between y and z.

Proof: $(1) \Rightarrow (2)$ is clear, since if x were between y and z then $X \setminus \{x\}$ would not be convex.

Conversely, suppose $X \setminus \{x\}$ is not convex. Since X *is*, we have $x \in \text{conv}(X \setminus \{x\})$. Therefore, by Theorem 78, there are points $w, y, z \in X \setminus \{x\}$ such that $x \in \text{conv}(w, y, z)$. If x lies in the boundary of the pseudotriangle conv (w, y, z), then $x \in$ (say) conv (y, z), and we are done. Otherwise, x lies in the interior of the pseudotriangle, as in Figure 5.14.

Consider any pseudoline l through x. It meets the sides of the pseudotriangle in two points — either one of w, y, z plus another point v on one of the sides of the pseudotriangle, or else in two points u, v lying on different sides of the pseudotriangle, say u on $\overline{wy}$ and v on $\overline{wz}$. In each case, however, we get a contradiction to condition (2). □

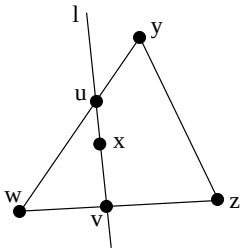


Figure 5.14: Minkowski's theorem on extreme points.

If the situation in Proposition 82 holds, we say that x is an *extreme point* of the convex set X.

Theorem 83. *A compact convex set X in a TAP $\mathcal{A}$ is the convex hull of its extreme points.*

Proof: Take a point $x \in X$. Choose a pseudoline l through x. Let l meet the boundary ∂X in two points, y and z, lying on opposite sides of x. Consider a supporting pseudoline to X at y; it exists by Theorem 79. The pseudosegment of ∂X containing y has extreme endpoints p and q (one or both of which may be y itself). Similarly for z: extreme endpoints r and s. Then $x \in \text{conv}\,(p, q, r, s)$. □

As a final corollary to Carathéodory's theorem we get the so-called *anti-exchange principle* for convex sets in a TAP:

Theorem 84. *If S is a compact convex set in a TAP $\mathcal{A}$ with $x, y \notin S$ but $y \in$ conv $(S \cup \{x\})$ and $x \in$ conv$(S \cup \{y\})$, then $y = x$.*

Proof: It follows from Theorem 78 plus the hypothesis $x, y \notin S$ that there are points $s_1, s_2, s_3, s_4 \in S$ such that $y \in \text{conv}\,(s_1, s_2, x)$ and $x \in \text{conv}\,(s_3, s_4, y)$. Let l be a pseudoline (strictly) separating x from S. It follows that l (strictly) separates y from S as well, since otherwise the entire set conv $(S \cup \{y\})$ would lie in the (closed) halfplane determined by l that does not contain x, so that x could not lie in conv $(S \cup \{y\})$.

As a consequence of Fact 1 above, we may assume that the seven pseudolines consisting of l plus the sides of the pseudotriangles $s_1 s_2 x$ and $s_3 s_4 y$ are straight, and that $\mathcal{A}$ is the Euclidean plane. Assuming $y \neq x$, this gives the situation depicted in Figure 5.15. Since x is in triangle $s_3 s_4 y$, the distance from x to l must be less than the distance from y to l. But the reverse is true for the corresponding reason, and this gives a contradiction. □

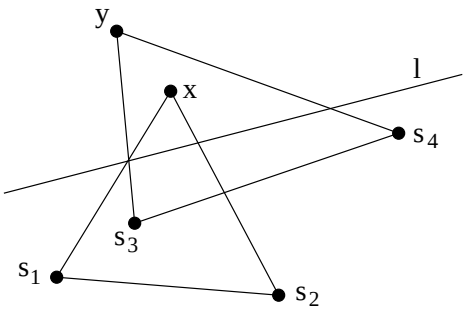


Figure 5.15: Anti-exchange.

We remark that Theorem 84 can also be proven without recourse to the Euclidean plane, by simply considering the order of points along the directed pseudoline $\overrightarrow{xy}$.

5.4 Conclusions

The reader may wonder why we have extended the standard convexity theorems only in dimension 2. What about topological affine d-spaces? The reason is simply that there are none, besides the standard Euclidean ones. It has been known for well over a century (see [Hil10], for example) that as soon as the dimension is 3 or more, the standard axioms of geometry imply Desargues's theorem; it follows from this that the space in question is isomorphic to the usual Euclidean d-space [CK78]. Thus it is only in dimension $d = 2$, where nonstretchable pseudoline arrangements proliferate, that non-Euclidean topological affine d-spaces can exist.

BIBLIOGRAPHY

[AFG09] Patrizio Angelini, Fabrizio Frati, and Luca Grilli. An algorithm to construct greedy drawings of triangulations. In *Graph Drawing: 16th International Symposium, GD 2008, Heraklion, Crete, Greece, September 21-24, 2008. Revised Papers*, pages 26–37, Berlin, Heidelberg, 2009. Springer-Verlag.

[AZ99] Nina Amenta and Günter Ziegler. Deformed products and maximal shadows of polytopes. *Advances in Discrete and Computational Geometry, Contemporary Mathematics*, 223:57 – 90, 1999.

[BFM07] Nicolas Bonichon, Stefan Felsner, and Mohamed Mosbah. Convex drawings of 3-connected planar graphs. *Algorithmica*, 47:399–420, 2007.

[Bix02] Robert E. Bixby. Solving real-world linear programs: A decade and more of progress. *Operations Research*, 50(1):3–15, 2002.

[BK07] Robert M. Bell and Yehuda Koren. Scalable collaborative filtering with jointly derived neighborhood interpolation weights. In *ICDM*, pages 43–52. IEEE Computer Society, 2007.

[BMSU99] Prosenjit Bose, Pat Morin, Ivan Stojmenovic, and Jorge Urrutia. Routing with guaranteed delivery in ad hoc wireless networks. In *DIALM '99: Pro-*

ceedings of the 3rd international workshop on Discrete algorithms and methods for mobile computing and communications, pages 48–55, New York, NY, USA, 1999. ACM Press.

[Bor80] Karl Heinz Borgwardt. *The Simplex Method: A Probabilistic Analysis*. Number 1 in Algorithms and Combinatorics. Spring-Verlag, 1980.

[Bre00] E. Brehm. 3-orientations and schnyder 3-tree decompositions. *Diplomarbeit, Freie Universitat, Berlin*, 2000.

[Can74] J. Cantwell. Geometric convexity, i. *Bulletin of the Institute of Mathematics Academia Sinica*, 2:289 – 307, 1974.

[CK78] J. Cantwell and D. C. Kay. Geometric convexity, iii: Embedding. *Transactions of the American Mathematical Society*, 246:211 – 230, 1978.

[CT09] Emmanuel J. Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *CoRR*, abs/0903.1476, 2009.

[dFPP88] Hubert de Fraysseix, János Pach, and Richard Pollack. Small sets supporting fáry embeddings of planar graphs. In *STOC '88: Proceedings of the twentieth annual ACM symposium on Theory of computing*, pages 426–433, New York, NY, USA, 1988. ACM Press.

[DGH+06] Raghavan Dhandapani, Jacob E. Goodman, Andreas Holmsen, Richard Pollack, and Shakhar Smorodinsky. Convexity in topological affine planes. *Accepted to Discrete and Computational Geometry*, 2006.

[DGK63] L. Danzer, B. Grunbaum, and V. Klee. Helly's theorem and its relatives. In Victor Klee, editor, *Convexity*, pages 101 – 180. American Mathematical Society, 1963.

[Dha08] Raghavan Dhandapani. Greedy drawings of triangulations. In *SODA '08: Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 102–111, Philadelphia, PA, USA, 2008. Society for Industrial and Applied Mathematics.

[EG09] David Eppstein and Michael T. Goodrich. Succinct greedy graph drawing in the hyperbolic plane. In *Graph Drawing: 16th International Symposium, GD 2008, Heraklion, Crete, Greece, September 21-24, 2008. Revised Papers*, pages 14–25, Berlin, Heidelberg, 2009. Springer-Verlag.

[Fel01] S. Felsner. Convex drawings of planar graphs and the order dimension of 3-polytopes. *Order*, 18:19 – 37, 2001.

[Fel04] S. Felsner. *Geometric Graphs and Arrangements*. Vieweg Verlag, 2004.

[FPS05] Eric Fusy, Dominique Poulalhon, and Gilles Schaeffer. Dissections and trees, with applications to optimal mesh encoding and to random sampling. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 690–699, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.

[GGH+01] Jie Gao, Leonidas J. Guibas, John Hershberger, Li Zhang, and An Zhu. Geometric spanner for routing in mobile networks. In *MobiHoc '01: Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing*, pages 45–55, New York, NY, USA, 2001. ACM Press.

[Gol94] D Goldfarb. On the complexity of the simplex algorithm. *Advances in optimization and numerical analysis: Proc. 6th Workshop on Optimization and Numerical Analysis*, pages 25 – 38, 1994.

[Goo04] J. E. Goodman. *Handbook of Discrete and Computational Geometry*, chapter Pseudoline arrangements, pages 97 – 128. CRC Press, 2004.

[GP80] J. E. Goodman and R. Pollack. Proof of Grunbaum's conjecture on the stretchability of certain arrangements of pseudolines. *Journal of Combinatorial Theory Series A*, 29:385 – 390, 1980.

[GPWZ94] J. E. Goodman, R. Pollack, R. Wenger, and T. Zamfirescu. Arrangements and topological planes. *American Mathematical Monthly*, 101:866 – 878, 1994.

[Gru72] B. Grunbaum. *Arrangements and Spreads*. American Mathematical Society, 1972.

[Hil10] D. Hilbert. *Foundations of Geometry*. Open Court, Chicago, 2 edition, 1910.

[HT74] J. Hopcroft and R. Tarjan. Efficient planarity testing. *Journal of the ACM*, 21(4):549 – 568, 1974.

[Kal00] F. B. Kalhoff. Oriented rank three matroids and projective planes. *European Journal of Combinatorics*, 21:347 – 365, 2000.

[Kan92] G. Kant. Drawing planar graphs using the lmc-ordering. In *Foundations of Computer Science, Proceedings., 33rd Annual Symposium on*, pages 101 – 110, 1992.

[KK00] Brad Karp and H. T. Kung. GPSR: Greedy perimeter stateless routing for wireless networks. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 243–254, New York, NY, USA, 2000. ACM Press.

[KKM29] B. Knaster, C. Kuratowski, and C. Mazurkiewicz. Ein Beweis des Fixpunktsatzes fur n-dimensionale Simplexe. *Fundamenta Mathematicae*, 14:132 – 137, 1929.

[Kle06a] Robert Kleinberg. Personal communication. 2006.

[Kle06b] Robert Kleinberg. Geographic routing in hyperbolic space. In *Workshop on Parallelism in Algorithms and Architectures*. University of Maryland, College Park, May 12, 2006.

[KM72] V. Klee and G. J. Minty. How good is the simplex algorithm? In O. Shisha, editor, *Inequalities III*, pages 159 – 173. Academic Press, New York, NY, 1972.

[Kor08a] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Ying Li, Bing Liu, and Sunita Sarawagi, editors, *KDD*, pages 426–434. ACM, 2008.

[Kor08b] Yehuda Koren. Tutorial on recent progress in collaborative filtering. In Pearl Pu, Derek G. Bridge, Bamshad Mobasher, and Francesco Ricci, editors, *RecSys*, pages 333–334. ACM, 2008.

[Kor09a] Yehuda Koren. The bellkor solution to the netflix grand prize. `http://www.netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf`, 2009.

[Kor09b] Yehuda Koren. Collaborative filtering with temporal dynamics. In John F. Elder IV, Françoise Fogelman-Soulié, Peter A. Flach, and Mohammed Javeed Zaki, editors, *KDD*, pages 447–456. ACM, 2009.

[KS06] Jonathan Kelner and Daniel Spielman. A randomized polynomial-time simplex algorithm for linear programming. In *38th ACM Symposium on Theory of Computing Seattle, Washington, USA*, 2006.

[KWZZ03] Fabian Kuhn, Roger Wattenhofer, Yan Zhang, and Aaron Zollinger. Geometric ad-hoc routing: of theory and practice. In *PODC '03: Proceedings of the twenty-second annual symposium on Principles of distributed computing*, pages 63–72, New York, NY, USA, 2003. ACM Press.

[LLW88] N. Linial, L. Lovasz, and A. Wigderson. Rubber bands, convex embeddings and graph connectivity. *Combinatorica*, 8(1):91 – 102, 1988.

[May06] Petar Maymounkov. Greedy embeddings, trees, and euclidean vs. lobachevsky geometry. *Submitted*, 2006.

[MG06] Jirí Matouek and Bernd Gärtner. *Understanding and Using Linear Programming (Universitext)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[ML08] Ankur Moitra and Tom Leighton. Some results on greedy embeddings in metric spaces. In *FOCS '08: Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 337–346, Washington, DC, USA, 2008. IEEE Computer Society.

[NR04] Takao Nishizeki and Dr Md Saidur Rahman. *Planar Graph Drawing*. World Scientific Publishing, 2004.

[PC09] Martin Piotte and Martin Chabbert. The pragmatic theory solution to the netflix grand prize. `http://www.netflixprize.com/assets/GrandPrize2009_BPC_PragmaticTheory.pdf`, 2009.

[PR05] Christos H. Papadimitriou and David Ratajczak. On a conjecture related to geometric routing. *Theor. Comput. Sci.*, 344(1):3–14, 2005.

[Rot05] Gunter Rote. Strictly convex drawings of planar graphs. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 728–734, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.

[RPSS03] Ananth Rao, Christos Papadimitriou, Scott Shenker, and Ion Stoica. Geographic routing without location information. In *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking*, pages 96–108, New York, NY, USA, 2003. ACM Press.

[Sal68] H. R. Salzmann. *Advances in Mathematics*, chapter Topological planes, pages 1 – 60. Number 2. 1968.

[San02] Luis A. Santalo. *Integral Geometry and Geometric Probability*. Cambridge University Press, 2002.

[Sch90] Walter Schnyder. Embedding planar graphs on the grid. In *SODA '90: Proceedings of the first annual ACM-SIAM symposium on Discrete algorithms*, pages 138–148, Philadelphia, PA, USA, 1990. Society for Industrial and Applied Mathematics.

[Sch93] R. Schneider. *Handbook of Convex Geometry*, chapter Convex surfaces, curvature and surface area measures, pages 273 – 299. North-Holland, Amsterdam, 1993.

[Sha87] Ron Shamir. The efficiency of the simplex method: a survey. *Management Science*, 33(3):301–334, 1987.

[She94] Jonathan R Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. Technical report, Carnegie Mellon University, Pittsburgh, PA, USA, 1994.

[SJ03] N. Srebro and T. Jaakkola. Weighted low-rank approximations. In *Proceedings of the Twentieth International Conference on Machine Learning*, 2003.

[SM07] Ruslan Salakhutdinov and Andriy Mnih. Probabilistic matrix factorization. In John C. Platt, Daphne Koller, Yoram Singer, and Sam T. Roweis, editors, *NIPS*. MIT Press, 2007.

[ST04] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *J. ACM*, 51(3):385–463, 2004.

[TBET98] Ioannis G. Tollis, Giuseppe Di Battista, Peter Eades, and Roberto Tamassia. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall, 1998.

[TJ09] Andreas Toscher and Michael Jahrer. The bigchaos solution to the netfix grand prize. `http://www.netflixprize.com/assets/GrandPrize2009_BPC_BigChaos.pdf`, 2009.

[Tod91] Michael J. Todd. Probabilistic models for linear programming. *Mathematics of Operations Research*, 16(4):671–693, 1991.

[TPNT08] Gábor Takács, István Pilászy, Bottyán Németh, and Domonkos Tikk. Investigation of various matrix factorization methods for large recommender systems. In *2nd Netflix-KDD Workshop*, Las Vegas, NV, USA, August 24, 2008.

[Tut56] W. T. Tutte. A theorem on planar graphs. *Trans. Amer. Math. Soc.*, 82:99–116, 1956.

[Tut60] W.T. Tutte. Convex representations of graphs. In *Proceedings of the London Mathematical Society*, pages 304 – 320, 1960.

[Ver06] Roman Vershynin. Beyond hirsch conjecture: Walks on random polytopes and smoothed complexity of the simplex method. *FOCS*, 0:133–142, 2006.