# WEAK SEPARATION OF SETS

*by*
*Michael E. Houle*

School of Computer Science
McGill University
Montréal, Canada

July 1989

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

# Abstract

Consider the following fundamental problem: given two sets $R$ and $G$ of objects positioned in $d$-dimensional Euclidean space, does there exist a surface of some specific type which separates the objects of $R$ from the objects of $G$?

Much attention has been given to this problem, for many classes of objects and separating surfaces. However, very few satisfactory alternatives exist when the objects are not separable by any of the surfaces of the chosen class. In this thesis, a new combinatorial measure of separability is proposed, based on the largest subset of the objects in $R \cup G$ that may be separated using surfaces drawn from a certain class. The combinatorial and algorithmic questions arising from this *weak separation measure* are the main focus of the thesis. The strong relationship between the separable subsets of point sets and faces of hyperplane arrangements is investigated, and a variety of algorithms are presented for finding linear and spherical separators for point sets and sets of hyperspheres.

# Résumé

Considérons ce problème fondamental: étant donné deux ensembles $R$ et $G$ d'objets situés dans l'espace euclidien de dimension $d$, existe-t-il une surface d'une catégorie particulière qui sépare les objets de $R$ des objets de $G$?

Beaucoup d'attention a été accordée à la recherche des surfaces séparatrices variées pour plusieurs classes d'objets. Cependant, très peu d'alternatives satisfaisantes existent lorsqu'aucune des surfaces de la classe choisie ne peut séparer les objets. Dans cette thèse, une nouvelle mesure de "séparabilité" est proposée, fondée sur le plus grand nombre possible d'objets de $R \cup G$ qui admettent une surface séparatrice extraite d'une classe spécifique. Les problèmes de nature combinatoire et algorithmique provenant de la *mesure de séparation faible* sont le principal sujet de cette thèse. Le rapport fort qui existe entre les sous-ensembles séparables d'ensembles de points et les configurations d'hyperplans est étudié, et plusieurs algorithmes sont présentés pour obtenir les hyperplans séparateurs et hypersphères séparatrices d'ensembles de points et d'hypersphères.

# Statement of Originality

Except for the background material in Section 2.2, the hyperplane arrangement construction algorithm of Section 6.2, and the two-dimensional topological line sweep algorithm in Section 6.5, all elements of this thesis should be considered original contributions to knowledge. Any other theorems or algorithms appearing in this thesis that are the work of others are clearly indicated in the text. Furthermore, no assistance outside that acknowledged in the preface has been received.

# Preface

One does not write a Ph.D. thesis without being able to compile a list of positive and negative influences upon the final product. In the interest of fairness, I will devote equal time to describing both types of influences in this preface.

One of the negative influences over this thesis has been the Montréal Canadiens. Every fall and winter, they can always be relied upon to siphon away two evenings per week, reaching a glorious peak of five days per week in the spring during the playoffs. Another black hole into which much of my effort has fallen is the morass of gaming, computer or otherwise. Although I suspect that my sanity was preserved during times of great stress by the availability of these diversions, a great many work hours were lost thereby. Other perhaps more wholesome distractions that nonetheless each contributed in some small way to the delaying of my degree include weekend hockey or soccer, recreational reading, sunshine in general, and language lessons.

On the positive side, first of all, I would like to thank my family for supporting me in my endeavours, even though they still don't quite understand what it is that I do. David Avis, Naji Mouawad, Gilles Pesant, Jean-Marc Robert, David Samuel, Tom Shermer, and Rafe Wenger have all helped greatly either by proofreading the various technical reports that eventually constituted this thesis, or through interesting and fruitful discussion pertaining to topics of the thesis, or both. I also thank Jean-Marc for the many discussions concerning joint work that did not make it into this thesis, yet which paved the way for the results of Chapter 7. Mike "der Mouse" Parker is particularly appreciated for his zeal in hunting down the last (?) few bugs in this thesis. The presence of Peter Egyed, Hossam ElGindy, Minoru Ishii, Alain Leblanc, Alexis Maciel, Minou Mansouri, and Marek Teichmann also contributed to a very

productive research environment.

I would also like to thank the technical staff and the secretaries for their assistance in fighting off the various disasters that inevitably arise whenever deadlines draw near.

The financial support I received from NSERC, McGill University, and the Québec Government was very gratefully received.

Lastly, but certainly not least, I especially thank my advisor, Godfried Toussaint, not only for his good advice over the past five years, or his willingness to let me work at my own pace, but also for introducing me to computational geometry at a time when I was very close to abandoning computer science altogether.

*To Arton and Nathaniel,*
*for leaving no plane unperturbed*

# Contents

**Abstract** ii

**Résumé** iii

**Statement of Originality** iv

**Preface** v

**1 Introduction** 1

**2 Geometric Preliminaries and Definitions** 4

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

2.2 Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

2.3 Weak Separation Definitions . . . . . . . . . . . . . . . . . . . . . . . 12

**3 Theorems on the Existence of Separators** 15

3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

3.2 Separation Using Hyperplanes . . . . . . . . . . . . . . . . . . . . . . 18

3.3 Separation Using Hyperspheres . . . . . . . . . . . . . . . . . . . . . 21

**4 Separation and Duality** 26

4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

4.2 Dual Transforms for Linear Separation . . . . . . . . . . . . . . . . . 27

4.3 Linear Separation and Arrangements . . . . . . . . . . . . . . . . . . 31

4.4 Dual Transforms for Spherical Separation . . . . . . . . . . . . . . . . 35

4.5 Spherical Separation and Arrangements . . . . . . . . . . . . . . . . . . 41

**5 Separable Subsets and Components** **46**

5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46

5.2 Upper Bounds for Separable Components . . . . . . . . . . . . . . . . . 48

5.3 Counting the Number of Separable Subsets . . . . . . . . . . . . . . . 57

**6 Weak Separation Algorithms** **67**

6.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67

6.2 Constructing Homogeneous Hyperplane Arrangements . . . . . . . . . 68

6.3 Constructing Half-Space Arrangements . . . . . . . . . . . . . . . . . 75

6.4 Finding Weak Separators . . . . . . . . . . . . . . . . . . . . . . . . 78

6.5 Space Reduction Using the Topological Sweep . . . . . . . . . . . . . 83

**7 Wide Linear Separation Algorithms** **92**

7.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 92

7.2 Wide Strong Linear Separation . . . . . . . . . . . . . . . . . . . . . 94

7.3 Wide Weak Linear Separation . . . . . . . . . . . . . . . . . . . . . . 99

7.4 Linear Separation of Hyperspheres . . . . . . . . . . . . . . . . . . . 101

**8 Conclusion** **106**

# List of Figures


2.1 A partitioning by surface S . . . . . . . . . . . . . . . . . . . . . . . . 13

3.1 Convex sets in $E^2$ with exactly one subfamily of 3 sets non-intersecting 17

3.2 The stereographic projection of a hyperplane onto a hypersphere . . . 22

3.3 The stereographic projection of a $(d-1)$-dimensional sphere . . . . . . 23

3.4 The only spherical separator is degenerate . . . . . . . . . . . . . . . 24

4.1 A half-space of $E^d$ in the cartesian setting of $E^{d+1}$ . . . . . . . . . . 30

4.2 Mapping of Voronoi sites in $E^d$ to hyperplanes in $E^{d+1}$ . . . . . . . . 36

4.3 A sequence of hyperspheres tending toward a hyperplane in $E^d$ . . . . 39

5.1 A 5-set and an 8-set . . . . . . . . . . . . . . . . . . . . . . . . . . . 47

5.2 Construction for the proof of Lemma 5.6 . . . . . . . . . . . . . . . . . 53

5.3 Construction for the proof of Lemma 5.7 . . . . . . . . . . . . . . . . . 54

5.4 A closed cone in a homogeneous half-space arrangement . . . . . . . . 58

5.5 Construction for proof of Lemma 5.15 . . . . . . . . . . . . . . . . . . 61

6.1 A line arrangement and its incidence graph . . . . . . . . . . . . . . . 69

6.2 A node of the incidence graph of Figure 6.1 . . . . . . . . . . . . . . . 70

6.3 A zone in an arrangement of lines . . . . . . . . . . . . . . . . . . . . 72

6.4 Updating differential values when inserting hyperplane $h$ . . . . . . . 76

6.5 Examples of cuts in an arrangement of lines . . . . . . . . . . . . . . . 84

6.6 Finding a plane avoiding all edges of $\mathcal{A}(H_>)$ . . . . . . . . . . . . . . 87

7.1 A widest strong separator - unweighted case . . . . . . . . . . . . . . . 93

# Chapter 1

# Introduction

Consider the following fundamental problem: assume that we are given two sets of objects positioned in Euclidean space, one consisting of red objects and the other consisting of green objects. Does there exist a surface of some specific type which separates the red objects from the green objects? This question may be posed with various different classes of objects and separators, in spaces of any dimension.

Most of what is already known about separation involves the separator classes of hyperplanes and hyperspheres, and the object class of finite point sets. Theorems providing conditions for the existence of linear separators have been known for some time, notably those of Kirchberger [Kir03], who showed that the exact linear separability of two sets of points depends on the separability of all subsets of their union of a certain fixed size. More recently, the same was shown for spherical separability by Lay [Lay71]. The relationship between linear separability and convexity theory has also been well studied [DGK63,Val64]. One of the simpler proofs of Kirchberger's theorem relies heavily on the famous combinatorial convexity result of Helly [Hel23]. Stoer and Witzgall [SW70] have shown that two sets of points are linearly separable if and only if their convex hulls are separable.

Much attention has been given to the problem of finding specific types of separating surfaces for sets of points. For example, in the setting of two-dimensional image processing, efficient algorithms to find circular separators for two sets of points can be used to recognize disks [Bha88,Fis86,OKM86]. In the pattern recognition setting of two-category point classification, linear separators are often sought for use as discriminant functions. Many strategies for obtaining these separators exist (see [DH73]). However, it often happens that the two point sets cannot be separated by a simple hyperplane. In these instances, higher-order surfaces such as hyperspheres or hypercones are sometimes considered as candidate separators, or linear discriminant functions based on statistical considerations are employed [Cov65]. However, these approaches do not concern themselves with combinatorial alternatives to exact linear separators in the event that none can be found. One such alternative, explored in this thesis, is to use discriminant functions that correctly classify the greatest number of objects in the union of the two sets. Such surfaces, since they are not necessarily exact separators, shall be called *weak* separators of the sets. A formal definition of weak separators will be given later in the thesis.

One of the natural questions to ask, upon being told that a certain surface does not separate two sets of objects $R$ and $G$, is which subsets of $R$ and $G$ *are* separated? This question will be the motivation behind the combinatorial investigations of the thesis that will be conducted in Chapters 3 to 5. These chapters will be concerned with the theoretical aspects of weak separation of point sets by both hyperplanes and hyperspheres. Chapters 6 and 7 deal with the algorithmic aspects of weak separation.

Chapter 2 contains definitions, terminology, and other background information upon which the discussions of the thesis are based. The areas touched upon in this chapter include analytic geometry, elementary topology, hyperplane arrangements, and linear programming. Also in this chapter, new definitions involving separation and separators will be introduced.

In Chapter 3, results on the existence of exact separators are presented that are extensions of the first results of Kirchberger to finite families of arbitrary sets. This same treatment is applied to the case of exact spherical separation. The results of this chapter are contained in the paper "Theorems on the existence of separating sets", which has been accepted for publication in the journal *Discrete and Computational Geometry*.

In Chapter 4, the linearly and spherically separable subsets of two sets of points are characterized by establishing a correspondence between separable subsets and intersections of half-spaces in dual arrangements of half-spaces. The transforms used to develop this correspondence are extensions of transforms that have been used to solve many problems in computational geometry, notably in the constructions of convex hulls and higher-dimensional Voronoi diagrams [PS85,Ede87].

In Chapter 5, bounds on the worst-case size of certain fixed-size separable subsets are given, and their relationship to the theory of $k$-sets is shown. Also in this chapter, exact expressions are developed for the number of linearly and spherically separable subsets of two point sets, based on the relation between point sets and arrangements established in the previous chapter.

Chapter 6 is devoted to a variety of basic algorithms for finding weak linear and spherical separators of point sets. These algorithms are ultimately based upon hyperplane arrangement construction and sweeping techniques developed recently by Edelsbrunner, O'Rourke, and Seidel [EOS86], and Edelsbrunner and Guibas [EG86].

In Chapter 7, certain applications and extensions of the algorithms of the previous chapter are examined. In particular, algorithms are presented that determine "wide" linear separators of point sets; that is, separators that avoid the sets being separated by the greatest amount, according to a natural metric. These separators are in a sense of "higher quality" than those found using the more straightforward methods of the previous chapter. This method of finding wide linear separators of point sets will be shown to be applicable to the problem of finding a linear separator of two sets of hyperspheres.

Finally, in Chapter 8, some open problems relating to separation are discussed.

# Chapter 2

# Geometric Preliminaries and Definitions

## 2.1 Introduction

The topics covered in this thesis fall into the category of discrete and computational geometry. This discipline straddles the boundary between mathematics and theoretical computer science, and as such encompasses many subfields.

One important text of great relevance to this thesis is "Algorithms in Combinatorial Geometry" by Herbert Edelsbrunner [Ede87]. In it, the author explores in depth the relationship between the combinatorial structure of arrangements of hyperplanes, and their many applications in computational geometry. Much of the background material assumed by this thesis is covered in this text, including the theoretical and algorithmic aspects of hyperplane arrangements and Voronoi diagrams, geometric transforms, and the theory of $k$-sets. A more introductory (and more general) reference for computational geometry is the book "Computational Geometry" by Preparata and Shamos [PS85].

A solid reference for Euclidean and projective geometry is Borsuk's "Analytic Geometry" [Bor69], in which (among others) the topics of duality, homogeneous coordinates, and vector algebra are explored. Other recommended books on transformations are "Geometric Transformations" by Iaglom [Iag62], and "A Survey of Geometry" by

Eves [Eve72]. For convexity theory, there are many good texts. A good reference for convexity theory is Valentine's "Convex Sets" [Val64], which includes the theorems of Kirchberger concerning the linear separability of sets. Many theorems on convex sets are also to be found in "Helly's Theorem and its Relatives", by Danzer, Grünbaum, and Klee [DGK63]. Two well-known works of Grünbaum, "Convex Polytopes" [Grü67] and "Arrangements and Spreads" [Grü72], provide a very thorough treatment of the combinatorial structure of convex polytopes and hyperplane arrangements.

An excellent book on linear programming is "Linear Programming", by Vašek Chvátal [Chv83], a text which may be considered both introductory and advanced. There is a great variety of texts available on topology. An advanced book is "Principles of Mathematical Analysis" by Rudin [Rud64]. For a more introductory text, I recommend Bartle's "The Elements of Real Analysis" [Bar76]. Two of the standard texts on graph theory are "Graph Theory" by Harary [Har69], and "Graph Theory with Applications" by Bondy and Murty [BM76].

The next section deals with the definitions and mathematical properties that is the background of this thesis. The topics discussed include coordinate systems, basic topological objects and properties, flats, hyperplane arrangements, polytopes, linear programming, and graph theory. None of the definitions of this section are new; the reader familiar with these areas should feel free to pass over any part or all of this section, and to later use it as a reference if necessary.

Section 2.3 contains new definitions pertaining to the separation of sets. The terminology introduced in this section forms the basis of discussion in the succeeding chapters.

## 2.2 Background

The setting for this thesis is the $d$-dimensional *Euclidean space* $E^d$, of which each point $p$ is represented by its *cartesian coordinates*, a real-valued $d$-dimensional vector $(p_1, p_2, \ldots, p_d)$. We will let $\boldsymbol{R}$ and $\boldsymbol{R}^d$ represent the set of real numbers and real valued $d$-dimensional vectors, respectively. Let $x = (x_1, x_2, \ldots, x_d)$ and $y = (y_1, y_2, \ldots, y_d)$

be two points in $E^d$. The *dot product* or *inner product* of $x$ and $y$, denoted $x \cdot y$, is $\sum_{i=1}^{d} x_i y_i$. The (Euclidean) *norm* of $x$, denoted $\|x\|$, is defined as $\sqrt{x \cdot x}$, and the (Euclidean) *distance* between $x$ and $y$ is defined to be $\|x - y\|$. If $\|x\| = 1$, $x$ is said to be a *unit vector*. If $X$ and $Y$ are subsets of $E^d$, the *minimum Euclidean distance* between $X$ and $Y$ is $\delta(X, Y) = \min\{\|x - y\| \mid x \in X, y \in Y\}$. The point $(0, 0, \ldots, 0)$ is called the *origin*, and shall be represented by the symbol $\mathbf{0}$. The $x_i$-*axis* is the set of all points with zero $x_k$-coordinate, for all $k \neq i$.

Another way of representing the points of $E^d$ is through the use of *homogeneous* coordinates. These coordinates are $(d+1)$-dimensional vectors of the form $(p_0, p_1, \ldots, p_d)$, where each $p_i \in \boldsymbol{R}$ and $p_0 \neq 0$. The point $p \in E^d$ with cartesian coordinates $(p_1, p_2, \ldots, p_d)$ is assigned to the set of vectors $(\lambda, \lambda p_1, \lambda p_2, \ldots, \lambda p_d)$, over all $\lambda \in \boldsymbol{R}$, $\lambda \neq 0$. Any of these vectors can be said to represent the point $p$. The point $q \in E^d$ with homogeneous coordinates $(q_0, q_1, \ldots, q_d)$, $q_0 \neq 0$, then corresponds to the vector $(\frac{q_1}{q_0}, \frac{q_2}{q_0}, \ldots, \frac{q_d}{q_0})$ in cartesian coordinates. When expressing points of $E^d$ in homogeneous coordinates, it will sometimes be convenient to restrict the first coordinate $q_0$ to be positive. The cartesian origin $\mathbf{0}$ of $E^d$ can be expressed using homogeneous coordinates as $(\lambda, 0, \ldots, 0)$, for any $\lambda \in \boldsymbol{R}$, $\lambda \neq 0$.

Many of the object and property definitions in this section will be presented using both cartesian and homogeneous coordinates. Demonstration of the equivalence of these definitions will be left to the reader. Before proceeding further, some basic topological concepts are required.

A *ball* centred at $x$ is the set of all points of $E^d$ whose distance from $x$ is strictly less than some fixed radius $r \in \boldsymbol{R}$. A subset of $E^d$ is called *open* if it is the union of some collection of balls. A *closed* set is one whose complement is open. The *interior* of $X$, denoted $int(X)$, is the union of all the open sets contained in $X$. The *closure* of $X$, denoted $cl(X)$, is the intersection of all closed sets containing $X$. The *boundary* of $X$, denoted $bd(X)$, is the set $cl(X) \setminus int(X)$. A set $X$ is *bounded* if it is a subset of some ball. $X$ is *disconnected* if it is a subset of the union of two disjoint open sets in $E^d$, each containing some point of $X$. Otherwise, $X$ is said to be *connected*. A *connected component* of $X$ is a connected subset of $X$ that is contained in no other connected subsets of $X$. In $E^d$, a set is *compact* if and only if it is closed and bounded.

Let $\mathcal{F}$ be a mapping with domain $X$ and range in set $Y$. If each range element of $\mathcal{F}$ is associated with a unique domain element, we say that $\mathcal{F}$ is *injective* or *one-to-one.* If the range of $\mathcal{F}$ contains every element of $Y$, we say that $\mathcal{F}$ is *surjective*, or *onto.* A mapping that is both injective and surjective is known as a *bijection.* If $x$ is a point and $X$ is a set of points, the *inverse* of $\mathcal{F}$ is defined as $\mathcal{F}^{-1}(x) = \{y | \mathcal{F}(y) = x\}$, and $\mathcal{F}^{-1}(X) = \{y | \mathcal{F}(y) \in X\}$. If $\mathcal{F}$ is a function, then $\mathcal{F}$ is *continuous* if $\mathcal{F}^{-1}(X)$ is an open set whenever $X$ is an open set. In Euclidean space, a function is continuous if and only if $\lim_{x \to a} \mathcal{F}(x) = \mathcal{F}(a)$.

Two sets are called *isometric* if there exists a bijection between them that preserves distances. Such a bijection is called an *isometry.* A subset of $E^d$ isometric to $E^k$ is called an *affine subspace* or *flat* of *dimension* $k$ (also known as a *k-flat*). A 0-flat is a *point*, a 1-flat is a *line*, and a 2-flat is a *plane.* A $(d-1)$-flat is called a *hyperplane.* By convention, the empty set is considered to be a $(-1)$-flat.

A set of points $X$ is called *collinear* if all the points lie on the same line. $X$ is *coplanar* if all the points lie on the same plane. If no $k + 2$ points are contained in the same $k$-flat, $X$ is said to be in *general position.*

A *hypersphere* centred at point $x$ and of radius $r$ is defined as the set of points $y$ such that $\|x - y\| = r$. A hypersphere of radius 1 is called a *unit hypersphere.* In $E^2$ and $E^3$, hyperspheres are known as *circles* and *spheres* respectively.

The *affine hull* *aff*$(X)$ of a subset $X$ of $E^d$ is the smallest flat containing $X$. The *dimension* of $X$ (denoted *dim*$(X)$) is the dimension of its affine hull. A set of $k+1$ points of $E^d$ is said to be *affinely independent* if the dimension of their affine hull is $k$. A point $x \in E^d$ in cartesian coordinates is said to be an *affine combination* of a set of points $U = \{u_0, u_1, \ldots, u_k\}$ if $x = \sum_{i=0}^{k} \alpha_i u_i$ and $\sum_{i=0}^{k} \alpha_i = 1$. If in addition each $\alpha_i$ is non-negative, then $x$ is called a *convex combination* of $U$.

If $x$ and $U$ are represented in homogeneous coordinates, then $x$ is an *affine combination* of $U$ if $x = \sum_{i=0}^{k} \alpha_i u_i$ and $x_0 \neq 0$, where at least one of $\alpha_0, \alpha_1, \ldots, \alpha_k$ is not zero. If the first coordinate $u_{i0}$ of each $u_i$ is restricted to values greater than zero, and if in addition each $\alpha_i$ is non-negative, then $x$ is a *convex combination* of $U$.

A $k$-flat may be given as the set of affine combinations of $k+1$ affinely-independent points. Therefore, given such affinely-independent points, a $k$-flat may be parameterized by the set $\{\alpha_0, \alpha_1, \ldots, \alpha_k\}$ of the preceding definitions. If $u$ and $v$ are distinct points of $\boldsymbol{E}^d$, then the set of all affine combinations of $u$ and $v$ forms a line; the set of convex combinations of $u$ and $v$ (denoted $\overline{uv}$) is known as a (*closed*) *line segment.* The points $u$ and $v$ are called the *endpoints* of $\overline{uv}$. The *open* line segment joining $u$ and $v$ is the closed line segment minus its endpoints. A set $X$ is *relatively open* if it is the intersection of some open set with some flat. The *relative interior* of $X$ is the union of all the relatively open sets in *aff*$(X)$ that are contained in $X$.

A point set $X$ is called *convex* if the closed line segment joining any two points $u$ and $v$ of $X$ is entirely contained in $X$. It is easily shown that the common intersection of convex sets is itself convex. The *convex hull* of $X$, denoted by *conv*$(X)$, is the smallest convex set containing $X$. The boundary of the convex hull shall be denoted by $CH(X)$. In common practice, both *conv*$(X)$ and $CH(X)$ are called the "convex hull".

Using cartesian coordinates, a hyperplane may be expressed as the set of points $h = \{x \in \boldsymbol{E}^d | \, u \cdot x = c, u \in \boldsymbol{R}^d, c \in \boldsymbol{R}, u \neq 0\}$. The vector $u$ is called a *normal* vector for $h$. In homogeneous coordinates, a hyperplane may be expressed as the set of points $h = \{x \in \boldsymbol{E}^d | \, u \cdot x = 0, u \in \boldsymbol{R}^{d+1}, u \neq 0\}$. Sometimes, for the sake of brevity, these conditions on $u$ and $c$ are assumed but not stated.

Hyperplane $h$ splits $\boldsymbol{E}^d$ into two regions known as *half-spaces.* Using cartesian coordinates, the *closed* half-spaces defined by $h$ are parameterized as $\{x \in \boldsymbol{E}^d | \, u \cdot x \geq c\}$ and $\{x \in \boldsymbol{E}^d | \, u \cdot x \leq c\}$. Using homogeneous coordinates, these half-spaces become $\{x \in \boldsymbol{E}^d | \, u \cdot x \geq 0, x_0 > 0\}$ and $\{x \in \boldsymbol{E}^d | \, u \cdot x \leq 0, x_0 > 0\}$. Alternatively, if one wishes to consider tuples with $x_0 < 0$, it is easily seen that these expressions are equivalent to $\{x \in \boldsymbol{E}^d | \, x_0(u \cdot x) \geq 0\}$ and $\{x \in \boldsymbol{E}^d | \, x_0(u \cdot x) \leq 0\}$, respectively. The *open* half-spaces determined by $h$ are obtained from the closed half-spaces by eliminating the points of $h$. The hyperplane $h$ is said to *bound* its open and closed half-spaces.

A line $l$ in $\boldsymbol{E}^d$ not entirely contained in or parallel to a hyperplane $h$ is split by $h$ into two parts called *rays.* The intersection of $l$ and a closed half-space bounded by $h$ is known as a *closed* ray; with an open half-space, it is known as an *open* ray. The intersection of $l$ and $h$ is a point, called the *endpoint* of these rays. A ray is sometimes called a *half-line*.

Using cartesian coordinates, a closed ray may be parameterized as $u + \lambda v$, $\lambda \geq 0$, where $u$ and $v$ represent points of $\boldsymbol{E}^d$, $v \neq \mathbf{0}$. In homogeneous coordinates, the parameterization of a closed ray is $u + \lambda v$, $\lambda \geq 0$, where $u$ is a point of $\boldsymbol{E}^d$, and $v \in \boldsymbol{R}^{d+1}$, $v_0 = 0$, $v \neq \mathbf{0}$. In both the cartesian and homogeneous representations, adding the restriction $\lambda > 0$ gives an open ray. The endpoint of all these rays is the point $u$.

Let the points of $\boldsymbol{E}^d$ be represented using cartesian coordinates. A set $Y \subset \boldsymbol{E}^d$ is a *translation* of set $X$ if $Y = \{x + y \mid x \in X \subset \boldsymbol{E}^d\}$ for some fixed $y \in \boldsymbol{E}^d$. Two flats are *parallel* if one is a translate of the other. Two vectors $u$ and $v$ are said to be *orthogonal* to each other if $u \cdot v = 0$. Two intersecting flats $f_1$ and $f_2$ are *orthogonal* if, for all choices of $p_1$ in $f_1$ and $p_2$ in $f_2$, there exists a point $p$ in the intersection of $f_1$ and $f_2$ such that the vectors $p_1 - p$ and $p_2 - p$ are orthogonal. The *orthogonal projection* of a point $x$ onto a $k$-flat $f$ is the unique point $p$ of $f$ such that the the vectors $x - p$ and $p_f - p$ are orthogonal, for all points $p_f$ of $f$ distinct from $p$. Objects that are orthogonal to each other will also be referred to as being *normal* or *perpendicular* to each other.

A hyperplane $h$ is said to *separate* two sets $X$ and $Y$ if $X$ is contained in one closed half-space bounded by $h$, and $Y$ is contained in the other. $X$ and $Y$ are then said to be *separable*. If neither $X$ nor $Y$ intersect $h$, then $h$ *strictly separates* $X$ and $Y$, and the sets are *strictly separable*. Otherwise, if $X$ does intersect $h$, then $h$ is called a *supporting hyperplane* for $X$.

A set of hyperplanes $H = \{h_1, h_2, \ldots, h_n\}$ divides $\boldsymbol{E}^d$ into a set of connected convex regions called an *arrangement*, denoted $\mathcal{A}(H)$. Let $h_i{}^+$ and $h_i{}^-$ be the two open half-spaces bounded by hyperplane $h_i$. For a point $x$ in $\boldsymbol{E}^d$ we define

$$\nu_i(x) = \begin{cases} +1 & \text{if } x \in h_i{}^+, \\ 0 & \text{if } x \in h_i, \\ -1 & \text{if } x \in h_1{}^-, \end{cases} \quad \text{and}$$

for $1 \leq i \leq n$. The vector $\nu(x) = (\nu_1(x), \nu_2(x), \ldots, \nu_n(x))$ is called the *position vector* of $x$. If two points $x$ and $y$ have identical position vectors, then they are considered *equivalent*, and the equivalence classes defined by this relationship are called the *faces* of $\mathcal{A}(H)$. Since each point of face $f$ has the same position vector, we will sometimes refer to this vector as $\nu(f)$. A face $f$ is called a *k-face* if its affine hull is of dimension $k$. Faces of an arrangement are relatively open. A *vertex* is a 0-face, an *edge* is a 1-face, a *facet* is a $(d-1)$-face, and a *cell* is a $d$-face. A face $g$ is called a *subface* of face $f$ if the dimension of $g$ is one less than the dimension of $f$, and $g \subseteq cl(f)$. If so, we say that $f$ is a *superface* of $g$, and that $f$ and $g$ are *incident* upon each other. If $f$ and $g$ are two faces such that $g \subseteq cl(f)$, then $g$ is said to *bound* $f$.

If $\mathcal{A}(H)$ is an arrangement of $n \geq d$ hyperplanes, then $\mathcal{A}(H)$ is called *simple* if every $d$ hyperplanes of $H$ intersect in a unique point, and every $d+1$ hyperplanes have no common intersection. If $n < d$, then we say that $\mathcal{A}(H)$ is simple if the hyperplanes of $H$ intersect in a common $(d-n)$-flat. Equivalently, one could define a simple arrangement as one in which every $d-k$ hyperplanes intersect in a common $k$-flat, for $0 \leq k \leq d-1$. If $\mathcal{A}(H)$ is simple, then $H$ is said to be in *general position.*

A collection $\mathcal{C}$ of disjoint relatively open subsets of $E^d$ is called a *cell complex* if $E^d$ is the union of all sets in $\mathcal{C}$, and if the closure of any set in $\mathcal{C}$ is composed of the union of sets in $\mathcal{C}$. A cell complex in $E^2$ is called a (planar) *subdivision.* The collection of faces of a hyperplane arrangement form a cell complex in $E^d$. An arrangement in $E^2$ is often called a *line arrangement.*

The intersection of a finite number of closed half-spaces is known as a *polyhedral set.* If in addition it is bounded, it is called a (convex) *polytope.* Since all polytopes considered in this thesis will be convex, we will use the term "polytope" to mean "convex polytope". Alternatively, a polytope may be defined as the convex hull of a finite set of points. A *convex polygon* is a polytope of dimension 2, and a (convex) *polyhedron* is a polytope of dimension 3.

Given a set of distinct points $S = \{s_1, s_2, \ldots, s_n\}$ in $E^d$, the polyhedral sets of the form $v_i = \{x \in E^d \mid \|x - s_i\| \geq \|x - s_j\|, \forall s_j \in S\}$ determine a cell complex in $E^d$. This cell complex is called the $d$-dimensional *Voronoi diagram* of $S$. The points of $S$ are also called the *sites* of the Voronoi diagram.

If $P$ is a polytope, the set of *extreme points* (or *vertices*) of $P$ is the smallest set of points whose convex hull is $P$. A $k$-dimensional *simplex* is a $k$-dimensional polytope with $k+1$ vertices. A 1-dimensional simplex is called a *line segment*, a 2-dimensional simplex is a *triangle*, and a 3-dimensional simplex is a *tetrahedron*.

Since a bounded cell of an arrangement of hyperplanes is the interior of a polytope, it is natural that there be some overlap in the terminology used to describe these objects. Let polytope $P$ be the intersection of the finite set of closed half-spaces $H'$. Let $H$ be the set of bounding hyperplanes of $H'$. Then $f \subseteq P$ is a *face* of $P$ if $f$ is also a face of $\mathcal{A}(H)$. As with arrangements, we define *vertices*, *edges*, *subfaces*, and so on. However, if $P$ is of dimension $k$, we shall say that a *facet* of $P$ is a face of dimension $k-1$.

In Euclidean space, parallel hyperplanes do not intersect. This lack of intersection often results in the need to examine special cases for theorems and algorithms in $\boldsymbol{E}^d$. To avoid such problems, one resorts to the use of the *projective space* $\boldsymbol{P}^d$. For each line $l$ in $\boldsymbol{E}^d$, consider the set $L(l)$ of all lines parallel to $l$. We obtain $\boldsymbol{P}^d$ from $\boldsymbol{E}^d$ by the addition of a new point for each such set $L(l)$. These new points are called *improper* points, and the original points of $\boldsymbol{E}^d$ are called *proper*. Each line in a given set of parallel lines is extended to include the improper point corresponding to that set. These extended lines are called *projective* lines.

In projective space. distance has no meaning, and isometric mappings are impossible. Hence the definition of a flat in $\boldsymbol{E}^d$ cannot be carried over to $\boldsymbol{P}^d$. Instead, a *projective mapping* is a bijection of a set of points to a set of points which preserves projective lines. A set of points in $\boldsymbol{P}^d$ which is a projective mapping of $\boldsymbol{P}^k$ is known as a (projective) *k-flat*. Each $k$-flat $f$ in $\boldsymbol{E}^d$ has a corresponding projective $k$-flat, which is obtained by adding to it all the improper points that are projective extensions of lines contained in $f$. Every two projective $k$-flats intersect in a projective $(k-1)$-flat. Whether a flat is Euclidean or projective will be understood from the context.

The points of $\boldsymbol{P}^d$ are represented using homogeneous coordinates. Each point $p \in \boldsymbol{P}^d$ corresponds to the set of tuples $\lambda(p_0, p_1, \ldots, p_d)$, where $\lambda \in \boldsymbol{R}$, $\lambda \neq 0$, and $(p_0, p_1, \ldots, p_d) \neq (0, 0, \ldots, 0)$. If $p_0 \neq 0$, then the tuple $p$ corresponds to both a point of $\boldsymbol{P}^d$ and a point of $\boldsymbol{E}^d$. If $p_0 = 0$, then $p$ is an improper point associated with each of the lines of the form $l(q) = \{x \in \boldsymbol{E}^d | \, x = q + tp, t \in \boldsymbol{R}\}$, for all $q \in \boldsymbol{E}^d$.

Some terminology from linear programming is required. Let $x$ be a variable point of $E^d$. A *mathematical programming problem* is one where the goal is to choose some $x$ maximizing (or minimizing) some real-valued function, called an *objective function*, subject to certain *constraints* on $x$. The set of points of $E^d$ satisfying the problem constraints is called the *feasibility region* of the problem. If this region is empty, the problem is *infeasible*, otherwise the problem is *feasible*. A point of the feasibility region is called a *feasible solution* of the problem. If there exists a sequence of feasible solutions upon which the objective function diverges to infinity (if a maximizing problem) or negative infinity (if a minimizing problem), then the problem is considered *unbounded*. Otherwise, the problem is *bounded*. Unfortunately, this well-established terminology is somewhat ambiguous: to say that a problem is bounded is not the same as saying that a problem has a bounded feasibility region. If the objective function is linear in $x$, and the constraints describe halfspaces of $E^d$, the problem is called a *linear programming* problem.

We will require only the most basic definitions from graph theory. A *graph* is a collection of *nodes* (also called *vertices*) and *arcs* (also called *edges*), where each arc relates two nodes of the collection. Such an arc is said to be *incident* upon these two nodes. If two nodes are joined by some arc of the graph, they are called *adjacent*. Arcs are often represented by ordered or unordered pairs of nodes.

## 2.3 Weak Separation Definitions

Let $\mathcal{S}$ be a class of analytic surfaces in $E^d$, such that every surface $S \in \mathcal{S}$ is such that $E^d \setminus S$ consists of the two connected components $S_>$ and $S_<$. Let $\mathcal{R}$ and $\mathcal{G}$ be finite families of non-empty subsets of $E^d$, such that the sets of $\mathcal{R}$ are labeled *red*, and the sets of $\mathcal{G}$ are labeled *green*. The surface $S \in \mathcal{S}$ can be said to *partition* the family $\mathcal{R}$ into six disjoint subfamilies (see Figure 2.1):

$$\begin{aligned}
\mathcal{R}_= &= \{R \in \mathcal{R} \mid R \subseteq S\} \\
\mathcal{R}_> &= \{R \in \mathcal{R} \mid R \subseteq S_>\} \\
\mathcal{R}_< &= \{R \in \mathcal{R} \mid R \subseteq S_<\} \\
\mathcal{R}_\geq &= \{R \in \mathcal{R} \mid R \subseteq (S \cup S_>)\} \setminus \mathcal{R}_=
\end{aligned}$$

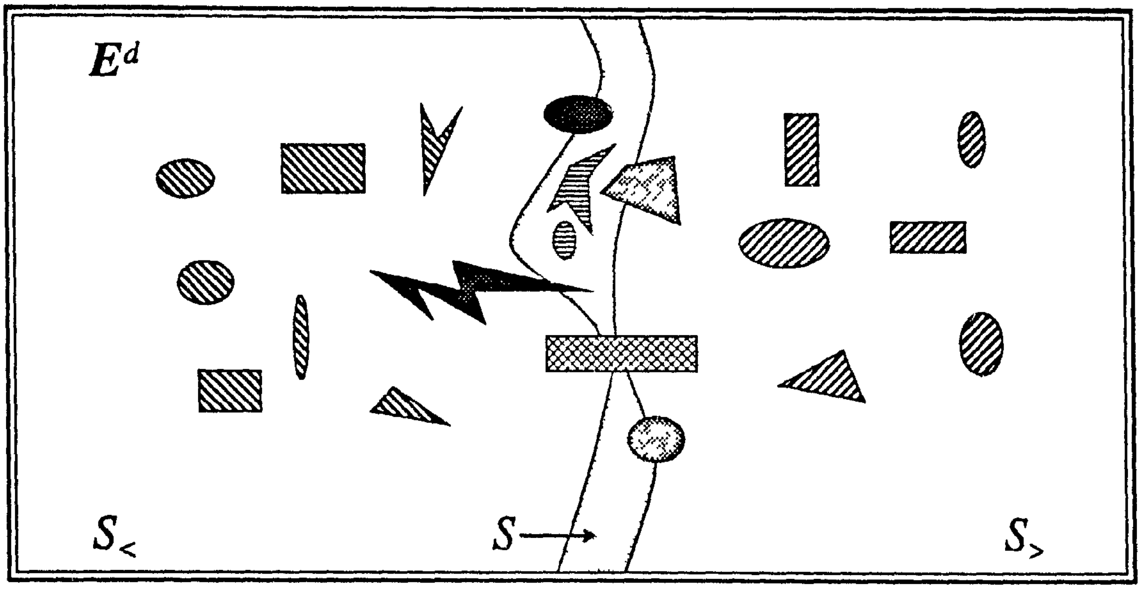


Figure 2.1: A partitioning by surface S

$$
\begin{aligned}
\mathcal{R}_{\leq} &= \{R \in \mathcal{R} \mid R \subseteq (S \cup S_<)\} \setminus \mathcal{R}_= \\
\mathcal{R}_0 &= \mathcal{R} \setminus (\mathcal{R}_= \cup \mathcal{R}_> \cup \mathcal{R}_< \cup \mathcal{R}_\geq \cup \mathcal{R}_\leq)
\end{aligned}
$$

Similarly, $\mathcal{G}$ is partitioned into the families $\mathcal{G}_=$, $\mathcal{G}_>$, $\mathcal{G}_<$, $\mathcal{G}_\geq$, $\mathcal{G}_\leq$, and $\mathcal{G}_0$.

Since the member sets of $\mathcal{R}_{\gg} = \mathcal{R}_= \cup \mathcal{R}_> \cup \mathcal{R}_\geq$ contain no points of $S_<$, and since the sets of $\mathcal{G}_{\ll} = \mathcal{G}_= \cup \mathcal{G}_< \cup \mathcal{G}_\leq$ contain no points of $S_>$, $\mathcal{R}_{\gg}$ and $\mathcal{G}_{\ll}$ are separated by $S$. Similarly, the families $\mathcal{R}_{\ll} = \mathcal{R}_= \cup \mathcal{R}_< \cup \mathcal{R}_\leq$ and $\mathcal{G}_{\gg} = \mathcal{G}_= \cup \mathcal{G}_> \cup \mathcal{G}_\geq$ are also separated by $S$. We shall call $\mathcal{R}_{\gg} \cup \mathcal{G}_{\ll}$ and $\mathcal{R}_{\ll} \cup \mathcal{G}_{\gg}$ the two *separable components* of $\mathcal{R}$ and $\mathcal{G}$ determined by $S$. These components are not necessarily disjoint, since the member sets of $\mathcal{R}_=$ and $\mathcal{G}_=$ are contained in both; nor do they account for all sets in $\mathcal{R} \cup \mathcal{G}$, as the members of subfamilies $\mathcal{R}_0$ and $\mathcal{G}_0$ are contained in neither component. Since the members of families $\mathcal{R}_>$ and $\mathcal{G}_<$ contain no points of $S$, these subfamilies of $\mathcal{R}$ and $\mathcal{G}$ are *strictly* separated by $S$. Accordingly, we call $\mathcal{R}_> \cup \mathcal{G}_<$ and $\mathcal{R}_< \cup \mathcal{G}_>$ the two *strictly* separable components of $\mathcal{R}$ and $\mathcal{G}$ with respect to $S$. Unlike non-strictly separable components, a pair of strictly separable components determined by a common separating surface must be disjoint.

We will say that the *size* of a separable component, be it strict or non-strict, is the number of non-empty member sets comprising that component. Let $\overline{\mathcal{C}}(\mathcal{R}, \mathcal{G})$ be the set of all separable components of $\mathcal{R}$ and $\mathcal{G}$, taken over all surfaces in $\mathcal{S}$. Then a component in $\overline{C}(\mathcal{R}, \mathcal{G})$ of greatest size will be called *maximal*, and one of smallest size will be called *minimal*. If $S$ is a surface determining a maximal component, then $S$ is said to be a *weak* (non-strict) separator of $\mathcal{R}$ and $\mathcal{G}$. If this maximal component contains all of $\mathcal{R}$ and $\mathcal{G}$, then $S$ completely separates $\mathcal{R}$ from $\mathcal{G}$, for which we say it is a *strong* (non-strict) separator. Analogously, we define *maximal* and *minimal strictly* separable components, and *weak* and *strong strict* separators.

The size of a maximal component gives rise to a measure of the separability of $\mathcal{R}$ and $\mathcal{G}$. Let $k$ be this size, and let $n$ be the number of member sets of $\mathcal{R} \cup \mathcal{G}$. Then the (strict or non-strict) *interpenetration* of $\mathcal{R}$ and $\mathcal{G}$ (with respect to the class of surfaces $\mathcal{S}$) is $n - k$; that is, the minimum number of sets of $\mathcal{R} \cup \mathcal{G}$ that need be eliminated to render the remaining sets separable. Interpenetration of zero indicates that the points are separable with respect to $\mathcal{S}$, and interpenetration approaching $n/2$ indicates that the sets are indistinguishable.

This thesis will be largely restricted to the investigation of separation of point sets with respect to the classes of hyperplanes and hyperspheres in $\boldsymbol{E}^d$. Although the class of separators will be clear from context, we will often distinguish these classes through the use of terms such as strict *linear* separators, weak *spherical* separation, and so on.

# Chapter 3

# Theorems on the Existence of Separators

## 3.1 Introduction

Two subsets $P$ and $Q$ of the $d$-dimensional Euclidean space $\boldsymbol{E}^d$ are said to be *(strictly) linearly separable* if there exists some hyperplane $h$ such that $P$ is contained in one of the two open half-spaces bounded by $h$, and $Q$ is contained in the other. In 1903, Paul Kirchberger published a fundamental theorem on the existence of strict linear separators for finite point sets in $\boldsymbol{E}^d$ [Kir03]:

**Theorem 3.1 (Kirchberger)** *Two finite subsets $P$ and $Q$ of $\boldsymbol{E}^d$ are strictly linearly separable if and only if, for each set $T$ consisting of at most $d+2$ points of $P \cup Q$, the sets $T \cap P$ and $T \cap Q$ are strictly linearly separable.*

A notion closely related to that of linear separability is spherical separability. Two subsets of $\boldsymbol{E}^d$ are said to be *(strictly) spherically separable* if there exists some hypersphere $s$ such that the interior of $s$ contains one subset and the exterior of $s$ contains the other. S. R. Lay extended Kirchberger's theorem to spherical separability in the following manner [Lay71]:

**Theorem 3.2 (Lay)** *Two finite subsets $P$ and $Q$ of $\boldsymbol{E}^d$ are strictly spherically separable if and only if, for each set $T$ consisting of at most $d+3$ points of $P \cup Q$, the sets $T \cap P$ and $T \cap Q$ are strictly spherically separable.*

One standard proof of Kirchberger's theorem, that of Rademacher and Schoenberg [RS50], employs the well-known theorem due to Helly concerning the existence of points in the common intersection of convex sets [Hel23,DGK63]. Whereas the original theorem of Helly is somewhat more general, we will require only the following restricted formulation:

**Theorem 3.3 (Helly)** *The members of a finite family $\mathcal{K}$ of convex subsets of $\boldsymbol{E}^d$ have a common intersection point if and only if, for each family $\mathcal{T}$ consisting of at most $d+1$ members of $\mathcal{K}$, the members of $\mathcal{T}$ have a common intersection point.*

These theorems are similar in that a "global" property of sets (linear separability, spherical separability, common intersection) is dependent upon the same property considered "locally" over subsets of bounded cardinality, these cardinalities being $d+2$ for Kirchberger's theorem, $d+3$ for Lay's, and $d+1$ for Helly's. It is not difficult to produce examples which demonstrate that the respective cardinalities cannot be decreased using the formulations given above. However, there is still a significant dissimilarity between Helly's theorem and the others. To illustrate this dissimilarity, let us consider an example. Let $\mathcal{K} = \{K_1, K_2, \ldots, K_n\}$ be a family of $n$ convex sets of $\boldsymbol{E}^d$, $n > d$, defined as follows (see Figure 3.1):

1. sets $K_1, K_2, \ldots, K_{d+1}$ are closed half-spaces whose bounding hyperplanes contain the $d+1$ facets of some $d$-dimensional simplex in $\boldsymbol{E}^d$,

2. these half-spaces do not contain the interior of this simplex, and

3. the remaining convex sets of $\mathcal{K}$, if any, are closed balls containing the simplex.

It is easily verified that the members of $\mathcal{K}$ have no point in common, yet with the exception of the subfamily $\{K_1, K_2, \ldots, K_{d+1}\}$, every subfamily consisting of at most $d+1$ members of $\mathcal{K}$ has a common point of intersection. If one were to test a family

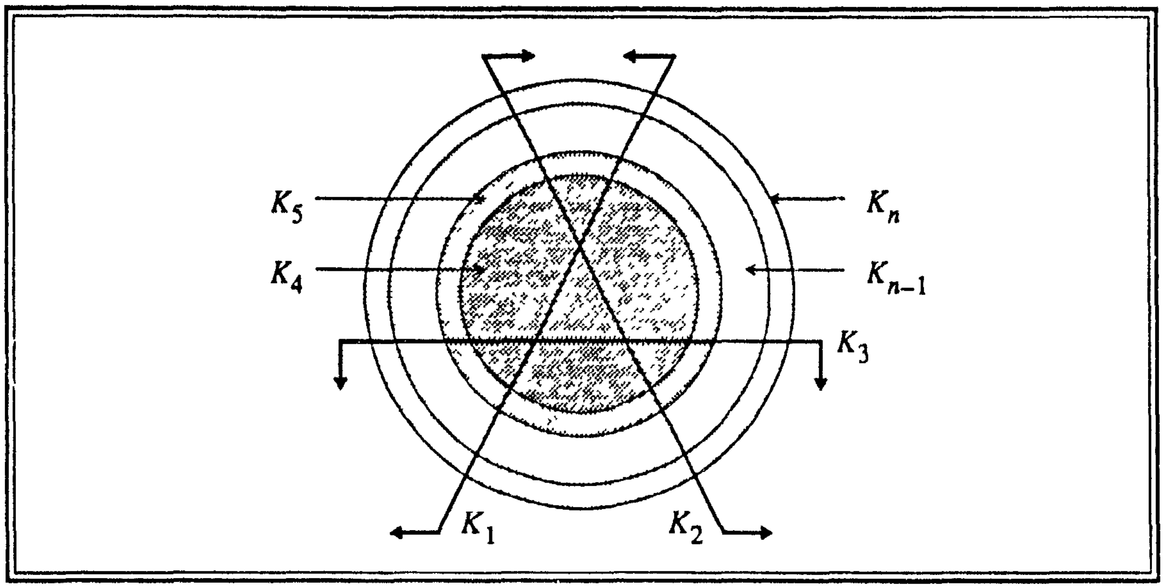


Figure 3.1: Convex sets in $E^2$ with exactly one subfamily of 3 sets non-intersecting

of convex sets for common intersection using Helly's theorem as a guide, one would expect to have to test all $\binom{n}{d+1}$ different subfamilies of cardinality $d+1$, in the worst case, before being able to make a decision. On the other hand, it is not hard to see that there are no examples of point sets $P$ and $Q$ of $E^1$, of combined cardinality $n > 3$, such that $P$ and $Q$ are not linearly separable but only one subset of $P \cup Q$ of cardinality three is not linearly separable. A similar situation exists in the setting of spherical separability. These observations suggest the possibility that Kirchberger's and Lay's theorems are not "optimal," in that fewer than $\binom{n}{d+2}$ subsets of $P \cup Q$ in $E^d$ need be tested for local linear separability in order to ascertain whether $P$ and $Q$ are themselves linearly separable, and fewer than $\binom{n}{d+3}$ subsets need be tested to ascertain whether $P$ and $Q$ are spherically separable. Indeed, this is reflected in the following refinement of Kirchberger's theorem, due to Watson [Wat73]:

**Theorem 3.4 (Watson)** *Let $P$ and $Q$ be disjoint finite sets of points in $E^d$, and let $x$ be any point in $P \cup Q$. $P$ and $Q$ are strictly linearly separable if and only if, for each set $T \subseteq P \cup Q$ consisting of at most $d+2$ points and containing $x$, the sets $T \cap P$ and $T \cap Q$ are strictly linearly separable.*

The main result of the next section is a generalization of Watson's refinement to finite families of arbitrary subsets of $E^d$. Two such families $\mathcal{R}$ and $\mathcal{G}$ shall be said to be (strictly) linearly separable if there exists some hyperplane $h$ such that the member sets of $\mathcal{R}$ are contained in one of the two open half-spaces bounded by $h$, and the member sets of $\mathcal{G}$ are contained in the other. In an analogous fashion, we may also define the spherical separability of finite families. Section 3.3 concerns itself with similar treatments of Lay's theorem.

## 3.2 Separation Using Hyperplanes

Let $h = \{x \in E^d |\, u \cdot x = 1\}$ be a hyperplane avoiding the origin, where $u \in E^d, u \neq 0$. Of the two open half-spaces delimited by $h$, we shall say that the half-space containing the origin, $h^+ = \{x \in E^d |\, u \cdot x < 1\}$, shall be called the *inner* half-space of $h$. Similarly, the other half-space, $h^- = \{x \in E^d |\, u \cdot x > 1\}$, shall be known as the *outer* half-space of $h$. Consider the point-hyperplane dual transform $\mathcal{D}$ that maps each point $p \in E^d$ ($p \neq 0$) into the hyperplane $\mathcal{D}(p) = \{x \in E^d |\, p \cdot x = 1\}$, and each origin-avoiding hyperplane $h = \{x \in E^d |\, u \cdot x = 1\}$ into the point $\mathcal{D}(h) = u$. The following observation is fairly straightforward:

**Observation 3.5** *Let $p$ be a point in $E^d$ other than the origin, and let $h$ be a hyperplane of $E^d$ avoiding the origin. If point $p$ is contained in hyperplane $h$, then point $\mathcal{D}(h)$ is contained in hyperplane $\mathcal{D}(p)$. Otherwise, if $p$ is contained in the inner (outer) half-space of $h$, then $\mathcal{D}(h)$ is contained in the inner (outer) half-space of $\mathcal{D}(p)$.*

Let $\mathcal{R}$ and $\mathcal{G}$ be finite families of subsets of the $d$-dimensional Euclidean space $E^d$, such that the members of $\mathcal{R}$ and $\mathcal{G}$ are coloured *red* and *green* respectively. We consider an augmentation $\mathcal{D}^*$ of the dual transform $\mathcal{D}$ that maps red sets $R \in \mathcal{R}$ into a collection of outer half-spaces $\mathcal{D}^*(R)$, and green sets $G \in \mathcal{G}$ into a collection of inner half-spaces $\mathcal{D}^*(G)$. That is, if $r$ is an element of some red set $R$, then the outer half-space $\mathcal{D}(r)^-$ is a member of $\mathcal{D}^*(R)$; the green case is defined analogously. Since $\mathcal{D}$ is undefined on the origin, we will say that a red point at the origin is mapped under $\mathcal{D}^*$ to the empty set $\emptyset$, and that a green point at the origin is mapped to the entire space $E^d$. The empty set and the space $E^d$ can be thought of as the outer and inner half-spaces of a hyperplane at infinity, respectively. Finally, if $P$ is a coloured set, we shall denote the common intersection of the half-spaces of $\mathcal{D}^*(P)$ as $\mathcal{I}(P)$. It should be noted that $\mathcal{I}(P)$ is necessarily convex (possibly empty), as it is the intersection of convex sets. The set $\mathcal{I}(P)$ has an interesting interpretation in light of separation:

**Lemma 3.6** *Let $P$ be a green (red) subset of $E^d$, and let $\mathcal{I}(P)$ be the common intersection of the members of $\mathcal{D}^*(P)$ as defined above. Then point $x \neq \mathbf{0}$ is contained in $\mathcal{I}(P)$ if and only if its dual hyperplane $\mathcal{D}(x)$ has all points of $P$ contained in its inner (outer) half-space.*

**Proof** Let $h^*$ be any member of $\mathcal{D}^*(P)$. By definition, $h^*$ is either an inner (outer) half-space of some hyperplane $h$ whose dual point $\mathcal{D}(h)$ is a point of $P$, or the entire space $E^d$ (empty set $\emptyset$). If $h^* = E^d$, then the point $\mathcal{D}^*(h^*)$ of $P$ is the origin, and is contained in the inner half-space of every hyperplane that is the dual under $\mathcal{D}$ of points of $\mathcal{I}(P) \setminus \{0\}$. (If $h^* = \emptyset$, the set $\mathcal{I}(P)$ is empty.) Otherwise, let $x \neq \mathbf{0}$ be a point of $E^d$ contained in $h^*$. Since $x$ is contained in the inner (outer) half-space of $h$, by Observation 3.5 we have green (red) point $\mathcal{D}(h)$ contained in the inner (outer) half-space of hyperplane $\mathcal{D}(x)$. Then $h^* \setminus \{0\}$ is precisely the set of all points of $E^d$ whose dual hyperplanes under $\mathcal{D}$ have inner (outer) half-spaces containing green (red) point $\mathcal{D}(h)$. Therefore $\mathcal{I}(P) \setminus \{0\}$ is the set of all points of $E^d$ whose dual hyperplanes under $\mathcal{D}$ have inner (outer) half-spaces containing $P$. □

**Theorem 3.7** *Let $\mathcal{R}$ and $\mathcal{G}$ be non-empty finite families of subsets of $E^d$, and let $P$ be any non-empty member of $\mathcal{R} \cup \mathcal{G}$. Then $\mathcal{R}$ and $\mathcal{G}$ are strictly linearly separable if and only if for each family $\mathcal{B}$ consisting of $d+1$ or fewer members of $\mathcal{R} \cup \mathcal{G}$, the families $(\mathcal{B} \cup \{P\}) \cap \mathcal{R}$ and $(\mathcal{B} \cup \{P\}) \cap \mathcal{G}$ are strictly linearly separable.*

**Proof** It suffices to prove the non-trivial implication. Without loss of generality, we assume that $P$ is a member of $\mathcal{G}$ and that the members of $\mathcal{R}$ and $\mathcal{G}$ are coloured red and green respectively. Also without loss of generality, we may translate the sets of $\mathcal{R}$ and $\mathcal{G}$ such that the set $P$ contains the origin. Let $\mathcal{B}$ be a set of $d+1$ or fewer members of $\mathcal{R} \cup \mathcal{G}$. By assumption, there exists a hyperplane $h$ that separates the families $\mathcal{B}_R = (\mathcal{B} \cup \{P\}) \cap \mathcal{R}$ and $\mathcal{B}_G = (\mathcal{B} \cup \{P\}) \cap \mathcal{G}$. Since $P$ contains the origin, hyperplane $h$ must avoid it, and $P$ is contained in the inner half-space of $h$. Thus all sets of $\mathcal{B}_G$ must be contained in the inner half-space of $h$, and all sets of $\mathcal{B}_R$ must be contained in the outer half-space. Lemma 3.6 then implies that if $B$ is a member of $\mathcal{B}_R \cup \mathcal{B}_G$, the point $\mathcal{D}(h)$ of $\boldsymbol{E}^d$ is contained in $\mathcal{I}(B)$, which in turn implies that the common intersection of these sets is non-empty. Since every such subset $\mathcal{B}$ of $d+1$ or fewer members of $\mathcal{R} \cup \mathcal{G}$ has this property, Helly's theorem implies that the common intersection $I$ of all sets of the form $\{\mathcal{I}(Q) | Q \in \mathcal{R} \cup \mathcal{G}\}$ is non-empty.

It should be noted that $I$ does not contain the origin: otherwise, since no outer half-space may contain the origin, the set $\mathcal{R}$ would be empty, violating the assumption. Let $x \neq 0$ be a point of $I$. Since $x$ is contained in $\mathcal{I}(Q)$ for each $Q \in \mathcal{R} \cup \mathcal{G}$, Lemma 3.6 again implies that each member $R$ of $\mathcal{R}$ is contained in the outer half-space of hyperplane $\mathcal{D}(x)$, and each member $G$ of $\mathcal{G}$ is contained in the inner half-space. Therefore the sets $\mathcal{R}$ and $\mathcal{G}$ are strictly linearly separable as required. □

The open half-spaces of a linear separator for families $\mathcal{R}$ and $\mathcal{G}$ may be labeled according to the family contained by each. In the context of Theorem 3.7, this labeling involves a degree of freedom that is eliminated by the choice of some distinguished set $P$ of $\mathcal{R} \cup \mathcal{G}$. In this sense, $P$ acts as a "focus" or a "reference" for the local tests of linear separability. The next theorem shows that one may refer to a distinguished *direction* instead of a distinguished set.

For simplicity of exposition, we assume that the distinguished direction is that of the positive $x_d$-axis, and will refer to it as the *vertical* direction. A hyperplane $h$ that does not contain a translate of the $x_d$-axis will be said to be *non-vertical.* The open half-spaces of $h$ can be described analytically as $h^+ = \{x \in \boldsymbol{E}^d | \, x_d > \sum_{i=1}^{d-1} u_i x_i + u_d\}$ and $h^- = \{x \in \boldsymbol{E}^d | \, x_d < \sum_{i=1}^{d-1} u_i x_i + u_d\}$. The halfspaces $h^+$ and $h^-$ will be called the *upper* and *lower* half-spaces of $h$, respectively. The points of $h^+$ will be said to be *above* $h$, and the points of $h^-$ will be said to be *below.*

**Theorem 3.8** *Let $\mathcal{R}$ and $\mathcal{G}$ be non-empty finite families of subsets of $\boldsymbol{E}^d$. Then $\mathcal{R}$ and $\mathcal{G}$ are strictly separable by a non-vertical hyperplane with $\mathcal{R}$ above the hyperplane and $\mathcal{G}$ below, if and only if for each family $\mathcal{B}$ consisting of $d+1$ or fewer members of $\mathcal{R} \cup \mathcal{G}$, the families $\mathcal{B} \cap \mathcal{R}$ and $\mathcal{B} \cap \mathcal{G}$ are strictly separable by a non-vertical hyperplane* *with $\mathcal{B} \cap \mathcal{R}$ above and $\mathcal{B} \cap \mathcal{G}$ below.*

**Proof** Let $\mathcal{B}$ be a family of $d+1$ or fewer members of $\mathcal{R} \cup \mathcal{G}$ as defined above, and let $h_b$ be a non-vertical hyperplane such that $\mathcal{B} \cap \mathcal{R}$ is above $h_b$ and $\mathcal{B} \cap \mathcal{G}$ is below. Also, let $P$ be the intersection of all upper half-spaces of hyperplanes $h_b$ over all finitely many choices of subfamily $\mathcal{B}$ of $\mathcal{R} \cup \mathcal{G}$. Note that $P$ cannot be empty. Then the families $(\mathcal{B} \cup \{P\}) \cap (\mathcal{R} \cup \{P\})$ and $(\mathcal{B} \cup \{P\}) \cap \mathcal{G}$ are linearly separable. Therefore the families $\mathcal{R} \cup \{P\}$ and $\mathcal{G}$ are strictly linearly separable by Theorem 3.7. But every vertical hyperplane intersects $P$, so the separator must be non-vertical. Finally, $P$ being above the linear separator implies the result. □

## 3.3 Separation Using Hyperspheres

In the proof of his theorem on spherical separability, Lay transforms an instance of a spherical separability problem in $E^d$ into a linear separability problem in $E^{d+1}$, by means of a stereographic projection. In this new setting, Lay applies Kirchberger's theorem directly to obtain his result. In this section, we will adapt Lay's proof in proving existence theorems for spherical separators similar to the linear separation theorems of the previous section.

Let $h$ be a hyperplane in $E^d$, and let $\Sigma$ be a hypersphere tangent to $h$ at point $p$. Let $p'$ be the point of $\Sigma$ antipodal to $p$. The *stereographic projection* $\tau$ of point $x \in h$ onto $\Sigma$ (based at $p'$) is defined as being the intersection of the line containing $x$ and $p'$ with $\Sigma \setminus \{p'\}$. (see Figure 3.2). This establishes a bijective correspondence between points of $h$ and points of $\Sigma \setminus \{p'\}$. Before presenting the theorems of this section, we shall present (without proof) some basic properties of stereographic projections. For additional information on stereographic projections and geometric transformations in general, the reader is refered to [Eve72,Iag62].

**Lemma 3.9** *Let $h$ be a hyperplane in $E^{d+1}$ and let $\Sigma$ be a $d$-dimensional hypersphere of unit radius tangent to $h$ at point $p$. Let $\tau$ be the stereographic projection of $h$ onto $\Sigma$ based at the point $p'$ antipodal to $p$ in $\Sigma$. Let $s$ be a $(d-1)$-dimensional sphere*

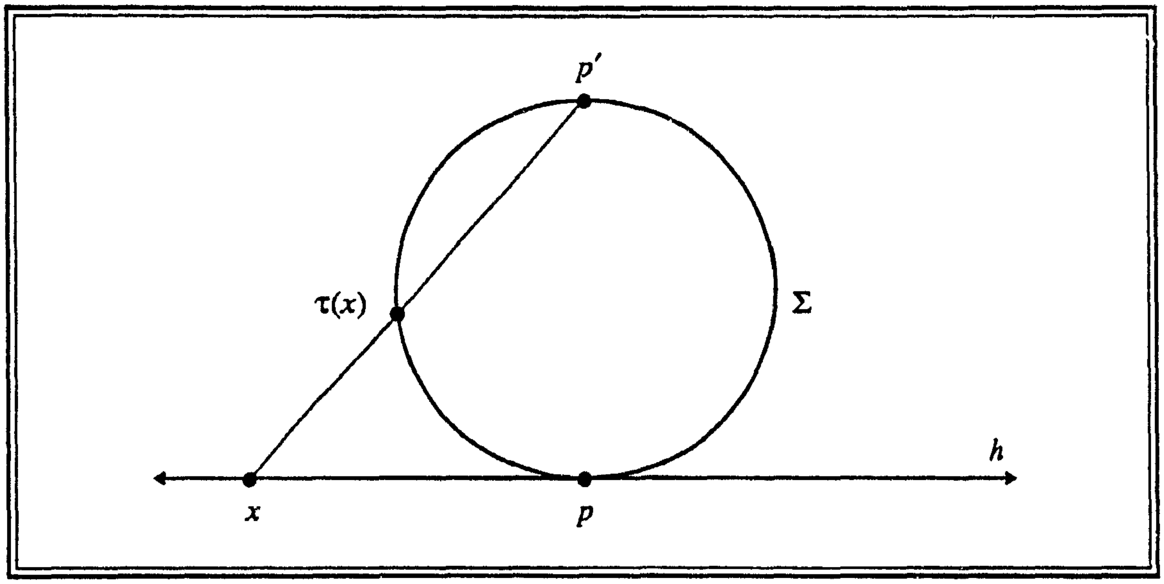


Figure 3.2: The stereographic projection of a hyperplane onto a hypersphere

*contained in* $h$*. and let* $s^+$ *and* $s^-$ *be its interior and exterior relative to* $h$*, respectively. Then*

1. *the projection* $\tau(s)$ *of* $s$ *onto* $\Sigma$ *is the intersection of* $\Sigma$ *and some hyperplane* $h_s$*, and*

2. *the projections* $\tau(s^+)$ *and* $\tau(s^-)$ *are each contained in different open halfspaces defined by* $h_s$*.*

See Figure 3.3 for an illustration of these relationships. A flat $f$ of dimension $d-1$ contained in $h$ may be viewed as a degenerate $(d-1)$-dimensional sphere centred at infinity with infinite radius. Obviously, the stereographic projection of $f$ is contained in the intersection of $\Sigma$ and a hyperplane passing through both $p'$ and $f$.

**Theorem 3.10** *Let* $\mathcal{R}$ *and* $\mathcal{G}$ *be non-empty finite families of subsets of* $\mathbf{E}^d$*, and let* $P$ *be any non-empty member of* $\mathcal{R} \cup \mathcal{G}$*. Then* $\mathcal{R}$ *and* $\mathcal{G}$ *are strictly separable by a (possibly degenerate) hypersphere if and only if for each family* $\mathcal{B}$ *consisting of* $d+2$ *or fewer members of* $\mathcal{R} \cup \mathcal{G}$*, the families* $(\mathcal{B} \cup \{P\}) \cap \mathcal{R}$ *and* $(\mathcal{B} \cup \{P\}) \cap \mathcal{G}$ *are strictly separable by a (possibly degenerate) hypersphere.*

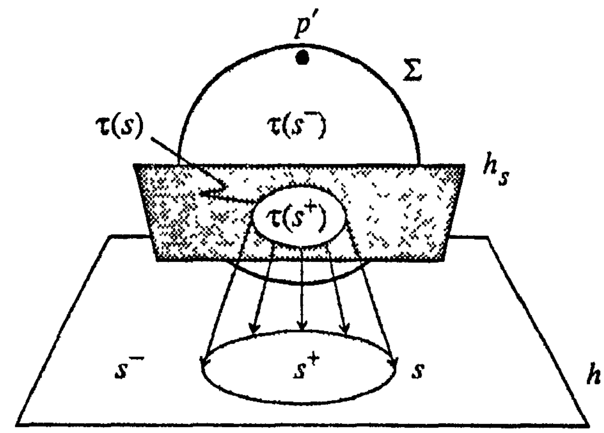


Figure 3.3: The stereographic projection of a $(d-1)$-dimensional sphere

**Proof** Let $E^d$ be embedded in some hyperplane $h$ of $E^{d+1}$, and let $\Sigma$ be a $d$-dimensional unit sphere tangent to $h$ at some arbitrary point $p$. Let $\tau$ be the stereographic projection of $h$ onto $\Sigma$ based at the point $p'$ antipodal to $p$. Let $\mathcal{B}$ be a family consisting of $d+2$ or fewer members of $\mathcal{R} \cup \mathcal{G}$ such that $\mathcal{B}_R = (\mathcal{B} \cup \{P\}) \cap \mathcal{R}$ and $\mathcal{B}_G = (\mathcal{B} \cup \{P\}) \cap \mathcal{G}$ are spherically separable in $h$ by some $(d-1)$-dimensional sphere $s$. If $h_s$ is a hyperplane containing $\tau(s)$, then Lemma 3.9 implies that the families $\tau(\mathcal{B}_R)$ and $\tau(\mathcal{B}_G)$ are strictly linearly separable by $h_s$. Therefore by Theorem 3.7, the families $\tau(\mathcal{R})$ and $\tau(\mathcal{G})$ are linearly separable.

Let $h'_s$ be a linear separator of $\tau(\mathcal{R})$ and $\tau(\mathcal{G})$ such that $h'_s$ intersects $\Sigma$ in some $(d-1)$-dimensional sphere $s'$. Since $\tau(\mathcal{R})$ and $\tau(\mathcal{G})$ are both non-empty, such a separator must exist. Then the (possibly degenerate) $(d-1)$-dimensional sphere $\tau^{-1}(s')$ strictly separates $\mathcal{R}$ and $\mathcal{G}$. □

Figure 3.4 gives an example of two families of sets in $E^2$ where every subfamily of five members is strictly spherically separable, but the only separator for the entire collection is degenerate. It should be noted that the closures of the triangles of Figure 3.4 intersect the separator $h$, but the triangles themselves do not. In the formulation of the previous theorem, if we restrict the members of $\mathcal{R}$ and $\mathcal{G}$ to be

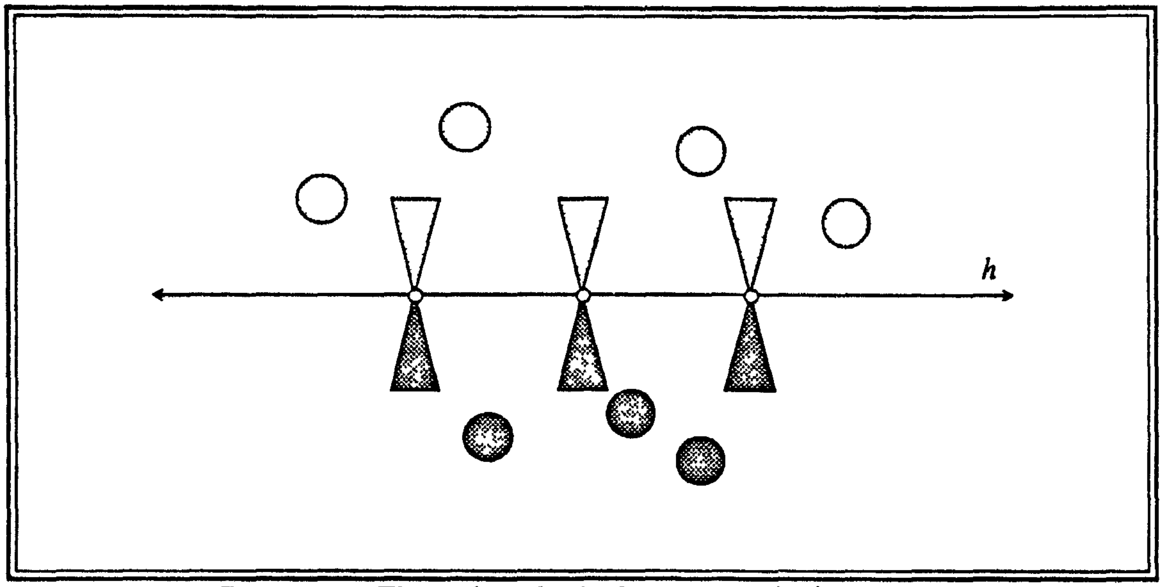


Figure 3.4: The only spherical separator is degenerate

compact (closed and bounded) sets, we can guarantee the non-degeneracy of the separating hyperspheres:

**Theorem 3.11** *Let $\mathcal{R}$ and $\mathcal{G}$ be non-empty finite families of compact subsets of $\boldsymbol{E}^d$, and let $P$ be any non-empty member of $\mathcal{R} \cup \mathcal{G}$. Then $\mathcal{R}$ and $\mathcal{G}$ are strictly spherically separable if and only if for each family $\mathcal{B}$ consisting of $d+2$ or fewer members of $\mathcal{R} \cup \mathcal{G}$, the families $(\mathcal{B} \cup \{P\}) \cap \mathcal{R}$ and $(\mathcal{B} \cup \{P\}) \cap \mathcal{G}$ are strictly spherically separable.*

**Proof** As in Theorem 3.10, we embed $\boldsymbol{E}^d$ into a hyperplane $h$ of $\boldsymbol{E}^{d+1}$ and apply a stereographic projection, arriving at a hyperplane $h'_s$ that separates $\tau(\mathcal{R})$ and $\tau(\mathcal{G})$. If $h'_s$ contains the previously-defined point $p'$, then due to the compactness of $\tau(\mathcal{R})$ and $\tau(\mathcal{G})$, we may perturb $h'_s$ into some new separator $h''$ that avoids $p'$. If $s''$ is the intersection of $h''$ and $\Sigma$, then the $(d-1)$-dimensional sphere $\tau^{-1}(s'')$ is a non-degenerate strict separator for $\mathcal{R}$ and $\mathcal{G}$. □

In Theorem 3.8, the need for a distinguished set $P$ of $\mathcal{R} \cup \mathcal{G}$ for linear separability was obviated by the introduction of a distinguished direction. In the setting of spherical separability, this distinction of direction becomes more natural. Let $s^+$ and $s^-$ be the open interior and exterior of hypersphere $s$, respectively. We shall say that the points of $s^+$ are *inside* $s$, and that the points of $s^-$ are *outside* $s$. We now state a theorem of spherical separability analogous to Theorem 3.8.

**Theorem 3.12** *Let $\mathcal{R}$ and $\mathcal{G}$ be non-empty finite families of subsets of $\mathbf{E}^d$. Then $\mathcal{R}$ and $\mathcal{G}$ are strictly separable by a hypersphere with $\mathcal{R}$ inside the hypersphere and $\mathcal{G}$ outside, if and only if for each family $\mathcal{B}$ consisting of $d+2$ or fewer members of $\mathcal{R}\cup\mathcal{G}$, the families $\mathcal{B}\cap\mathcal{R}$ and $\mathcal{B}\cap\mathcal{G}$ are strictly separable by a hypersphere with $\mathcal{B}\cap\mathcal{R}$ inside and $\mathcal{B}\cap\mathcal{G}$ outside.*

**Proof** Let $\mathcal{B}$ be a family of $d+2$ or fewer members of $\mathcal{R}\cup\mathcal{G}$ as defined above, and let $s_b$ be a hypersphere whose interior contains $\mathcal{B}\cap\mathcal{R}$ and whose exterior contains $\mathcal{B}\cap\mathcal{G}$. Also, let $s_p$ be a hypersphere whose interior contains $s_b$ for all finitely many choices of subfamily $\mathcal{B}$ of $\mathcal{R}\cup\mathcal{G}$. If $P$ is the exterior of $s_p$, then the families $(\mathcal{B}\cup\{P\})\cap\mathcal{R}$ and $(\mathcal{B}\cup\{P\})\cap(\mathcal{G}\cup\{P\})$ are spherically separable. Therefore the families $\mathcal{R}$ and $\mathcal{G}\cup\{P\}$ are (possibly degenerately) strictly spherically separable by Theorem 3.10. But every hyperplane intersects $P$, so the separator cannot be degenerate. Finally, $P$ being outside the spherical separator implies the result. $\square$

# Chapter 4

# Separation and Duality

## 4.1 Introduction

In the preceding chapter, we made use of a well-known dual relationship between points and hyperplanes to prove results concerning strict strong linear and spherical separability. In the linear case, member sets were transformed into convex sets, and strong separators (if they existed) were transformed into points in the common intersection of these convex sets. A situation where the convex sets have convenient properties is that in which each member set of both families consists of a single point. In this case, the convex sets corresponding to these points are simply open half-spaces. The orientation of each half-space is determined by the family of which its dual point was a member. Since these half-spaces intersect in a (possibly empty) convex polytope, the set of strong linear separators for point sets is implicitly characterized. In some sense, in limiting our attention to this intersection, we sacrifice a great deal of combinatorial information that is inherent in the arrangement of these half-spaces.

The aim of this chapter is to provide a characterization not only of strict strong linear and spherical separators for sets of points, but also of weak linear and spherical separators, separable components, and separable subsets in general, for both the strict and non-strict cases. This will be achieved through the transformation of the original setting in $\boldsymbol{E}^d$, with its two sets of points, its hyperplanes, half-spaces, and hyperspheres, into new settings in $E^{d+1}$ (for the linear case) and $E^{d+2}$ (for the spherical case). In these new settings, dual arrangements of half-spaces will be exhibited that capture all of the combinatorial qualities of the original. In the chapters to follow, these arrangements will be used as a framework for both proofs of combinatorial results, and for various separation algorithms.

In the next section, the dual transform of Chapter 3 will be extended for the case of linear separation. Points, hyperplanes, and open and closed half-spaces will all be given an interpretation in the transformed space. In Section 4.2, the properties of the dual arrangement with respect to separation will be examined. In Section 4.4, the spherical case will be considered. The transformations of this section are extensions of a geometric transformation due to Edelsbrunner and Seidel [ES86] that relates Voronoi diagrams in $E^d$ with hyperplane arrangements in $E^{d+1}$. Their transform is in turn an extension of the connection established by Brown [Bro79,Bro80] between Voronoi diagrams in two dimensions and convex hulls of point sets in three dimensions.

## 4.2 Dual Transforms for Linear Separation

Let $p$ be a point in $E^d$, expressed in homogeneous coordinates as $(p_0, p_1, \ldots, p_d)$, and let $h = \{x \in E^d | u \cdot x = 0\}$ be a hyperplane in $E^d$, whose points are also represented using homogeneous coordinates. Note that for $h$ to be well-defined, the vector $u = (u_0, u_1, \ldots, u_d)$ must have $u_i \neq 0$ for some $i \in \{1, 2, \ldots, n\}$. Consider the dual transform $\mathcal{D}$ from $E^d$ to $E^d$ that maps the point $p$ into the hyperplane $\mathcal{D}(p) = \{x \in E^d | p \cdot x = 0\}$, and the hyperplane $h$ into the point $\mathcal{D}(h) = u$. Unfortunately, this transform is ill-defined if $p$ is the cartesian origin of $E^d$, or if $u_0 = 0$; in the former case, the origin is mapped to the "hyperplane at infinity", and in the latter, any hyperplane containing the origin is mapped to a point "at infinity". These annoying features may be eliminated by abandoning Euclidean space in favour of projective space. However, with this approach new difficulties arise.

In the $d$-dimensional space $P^d$, a hyperplane is parameterized using homogeneous coordinates as $h = \{x \in P^d | u \cdot x = 0\}$, where $u$ is a vector in $R^{d+1}$ other than $(0, 0, \ldots, 0)$. Let us examine the constraint $\{x \in P^d | u \cdot x > 0\}$. At first glance, this constraint would seem to describe a half-space bounded by $h$. However, noting that the homogeneous tuples $x$ and $-x$ represent the same point of $\boldsymbol{P}^d$, and that $u \cdot (-x) = -u \cdot x$, the constraint is meaningless for all $x \notin h$. This ambiguity may be eliminated for the proper points of $\boldsymbol{P}^d$ with the introduction of the additional constraint $x_0 \geq 0$. Since an improper point $x$ has $x_0 = 0$, the constraints are still without meaning for these points. Indeed, the assignment of an improper point to one of the two open half-spaces bounded by $h$ cannot be done except in some arbitrary fashion. Consequently, we are forced to reject projective space as being unsuitable for our purposes. Instead, we shall rely upon a natural embedding of $\boldsymbol{E}^d$ in $\boldsymbol{E}^{d+1}$.

A tuple representing a point of $\boldsymbol{E}^d$ in homogeneous coordinates can also be made to represent a point of $\boldsymbol{E}^{d+1}$ in cartesian coordinates. When not otherwise clear from the context, the tuple $x = (x_0, x_1, \ldots, x_d)$ will be written $x_H = (x_0, x_1, \ldots, x_d)_H$ when denoting a point of $\boldsymbol{E}^d$, and written $x_C = (x_0, x_1, \ldots, x_d)_C$ when denoting a point of $\boldsymbol{E}^{d+1}$. Using this notation, we reinterpret the $d$-dimensional "homogeneous" space $\boldsymbol{E}^d$ as a subset of the $(d+1)$-dimensional "cartesian" space $\boldsymbol{E}^{d+1}$. By the definition of homogeneous coordinates, if $p_H$ represents a point of $\boldsymbol{E}^d$, then $\lambda p_H$ represents the same point, for any scalar $\lambda \neq 0$. Interpreted as cartesian coordinates, these tuples define a pair of oppositely-oriented open rays with the origin of $\boldsymbol{E}^{d+1}$ as their common endpoint.

Given that our problem concerns the linear separation of two point sets, we introduce two labels, *red* and *green*, to be applied to the points of $\boldsymbol{E}^d$. A red point of $\boldsymbol{E}^d$ will be represented by tuples of the form $p_H = (p_0, p_1, \ldots, p_d)_H$, where $p_0 > 0$. In contrast, a green point will be represented by tuples where $p_0 < 0$. In this sense, the first coordinate is made to carry information concerning the labeling of the point.

In the cartesian space, a labeled point corresponds to one of the pair of oppositely-oriented rays mentioned above: the ray lying in the half-space $\{x_C \in \boldsymbol{E}^{d+1} | x_0 > 0\}$ corresponds to the point with label *red*, and the ray contained in the complementary open half-space corresponds to the point with label *green*. Let $\vec{E}^{d+1}$ be the set of all open rays with endpoint the origin of $\boldsymbol{E}^{d+1}$. The set of rays of $\vec{E}^{d+1}$ contained in $\{x_C \in \boldsymbol{E}^{d+1} | x_0 > 0\}$ shall be denoted by $\vec{E}_R^{d+1}$; that is, $\vec{E}_R^{d+1}$ consists of those rays of $\vec{E}^{d+1}$ associated with red points of $\boldsymbol{E}^d$. Similarly, the set of rays contained in

$\{x_c \in \boldsymbol{E}^{d+1} | x_0 < 0\}$ shall be called $\vec{E}_G^{d+1}$. Note that not all rays in $\vec{E}^{d+1}$ correspond to points of $\boldsymbol{E}^d$; namely, those rays contained in $\{x_c \in \boldsymbol{E}^{d+1} | x_0 = 0\}$. The set of such rays we shall call $\vec{E}_0^{d+1}$.

If $p$ is a labeled point in $\boldsymbol{E}^d$, we will denote its unique corresponding ray in $\vec{E}^{d+1}$ by $\vec{l}(p)$. Naturally enough, the ray $-\vec{l}(p)$ is associated with that labeled point with the same location as $p$, but with the opposite labeling. If $P$ is a set of labeled points in $\boldsymbol{E}^d$, we will denote by $\vec{l}(P)$ the set of rays in $\vec{E}^{d+1}$ associated with the points of $P$.

In order to interpret the hyperplanes and half-spaces of $\boldsymbol{E}^d$ in the higher dimensional setting, the inner product operation for points of $\boldsymbol{E}^d$ must be extended to the rays of $\vec{E}^{d+1}$. For $\vec{u}$ and $\vec{v}$ in $\vec{E}^{d+1}$, let us define the *ray* inner product $\vec{u} \cdot \vec{v}$ as follows

$$\vec{u} \cdot \vec{v} = \begin{cases} 1 & \text{if } a \in \vec{u} \text{ and } b \in \vec{v} \Rightarrow a \cdot b > 0, \\ 0 & \text{if } a \in \vec{u} \text{ and } b \in \vec{v} \Rightarrow a \cdot b = 0, \\ -1 & \text{if } a \in \vec{u} \text{ and } b \in \vec{v} \Rightarrow a \cdot b < 0. \end{cases}$$

To verify that the ray product is well-defined, consider the points $a \in \vec{u}$ and $b \in \vec{v}$. Every point of $\vec{u}$ and $\vec{v}$ may be expressed as $\lambda a$ and $\xi b$, respectively, for some choice of $\lambda > 0$ and $\xi > 0$. Then

$$(\lambda a) \cdot (\xi b) = \lambda \xi (a \cdot b),$$

which is guaranteed to have the same sign as $a \cdot b$.

Using the ray inner product, we may parameterize the cartesian equivalent of hyperplanes in $\boldsymbol{E}^d$ in terms of the rays of $\vec{E}^{d+1}$. If we consider these hyperplanes as consisting of both red and green points of $\boldsymbol{E}^d$, the expression $h = \{x_H \in \boldsymbol{E}^d | u \cdot x_H = 0\}$ for a hyperplane in the homogeneous space $\boldsymbol{E}^d$, in the cartesian setting of $\boldsymbol{E}^{d+1}$, becomes $\vec{l}(h) = \{\vec{x} \in \vec{E}^{d+1} | \vec{u} \cdot \vec{x} = 0\}$, where $\vec{x}$ is not a ray of $\vec{E}_0^{d+1}$. It will be convenient at times to ignore this restriction on the rays of $\vec{l}(h)$. This allows us to think of the collection of points contained in rays of $\vec{l}(h)$ as a hyperplane in $\boldsymbol{E}^{d+1}$ passing through the origin. Certainly, there exists a unique hyperplane in $\boldsymbol{E}^{d+1}$ containing all the rays of $\vec{l}(h)$. We will denote by $\mathrm{H}^{d+1}$ the set of all "hyperplanes" in $\boldsymbol{E}^{d+1}$ of the form $\{\vec{x} \in \vec{E}^{d+1} | \vec{u} \cdot \vec{x} = 0\}$, for every choice of $\vec{u}$ in $\vec{E}^{d+1}$.

The cartesian equivalents of half-spaces in $\boldsymbol{E}^d$ may also be parameterized in terms of the rays of $\vec{E}^{d+1}$. The homogeneous expressions $h_{\geq} = \{x_H \in \boldsymbol{E}^d | x_0(u \cdot x_H) \geq 0\}$

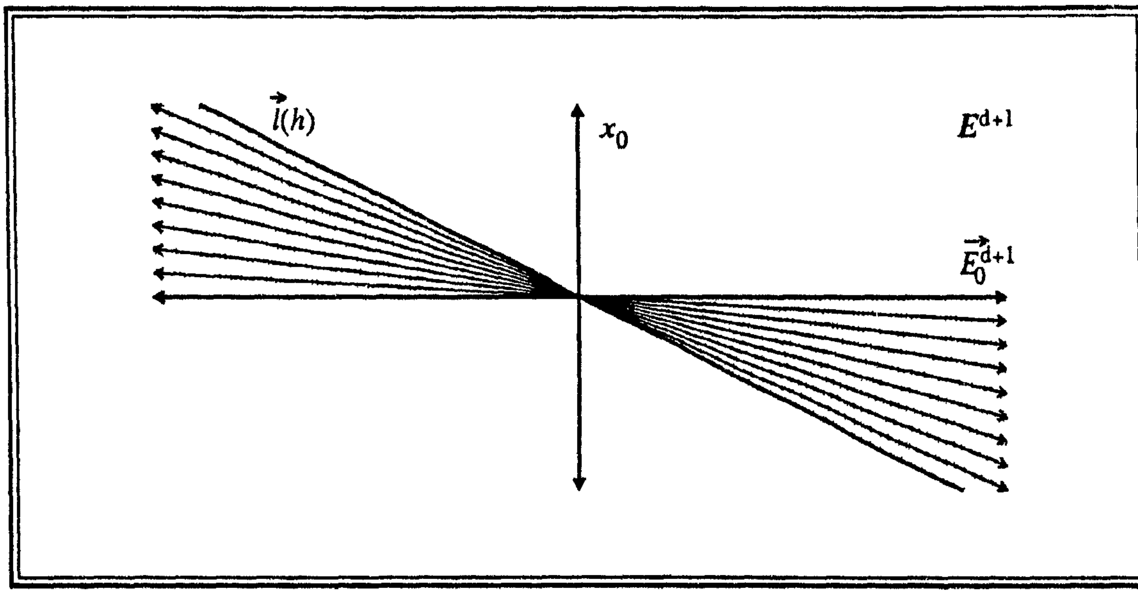


Figure 4.1: A half-space of $\boldsymbol{E}^d$ in the cartesian setting of $\boldsymbol{E}^{d+1}$

and $h_{\leq} = \{x_H \in \boldsymbol{E}^d | x_0(u \cdot x_H) \leq 0\}$ for the closed half-spaces bounded by hyperplane $h$ in $\boldsymbol{E}^d$ become

$$\begin{aligned} \vec{l}(h_{\geq}) &= \{\vec{x} \in \vec{E}_R^{d+1} | \vec{u} \cdot \vec{x} \geq 0\} \cup \{\vec{x} \in \vec{E}_G^{d+1} | \vec{u} \cdot \vec{x} \leq 0\} \text{ and} \\ \vec{l}(h_{\leq}) &= \{\vec{x} \in \vec{E}_R^{d+1} | \vec{u} \cdot \vec{x} \leq 0\} \cup \{\vec{x} \in \vec{E}_G^{d+1} | \vec{u} \cdot \vec{x} \geq 0\}, \end{aligned}$$

respectively, in the new setting of $\boldsymbol{E}^{d+1}$. Informally, these expressions describe "wedges" bounded by the hyperplanes $\vec{l}(h)$ and $\vec{E}_0^{d+1}$ (see Figure 4.1). As expected, the expressions associated with the open half-spaces of $\boldsymbol{E}^d$ are obtained from those associated with the closed half-spaces by making the inequalities strict.

In the same way that the set $\vec{h} = \{\vec{x} \in \vec{E}^{d+1} | \vec{u} \cdot \vec{x} = 0\}$ can be said to be a hyperplane of $\boldsymbol{E}^{d+1}$ parameterized using the rays of $\vec{E}^{d+1}$, expressions such as $\{\vec{x} \in \vec{E}^{d+1} | \vec{u} \cdot \vec{x} \geq 0\}$ and $\{\vec{x} \in \vec{E}^{d+1} | \vec{u} \cdot \vec{x} > 0\}$ can be said to denote closed and open half-spaces of $E^{d+1}$, respectively, whose bounding hyperplanes pass through the origin. Let $\vec{\Pi}_c^{d+1}$ be the family consisting of all the closed half-spaces of this type, and let $\vec{\Pi}_o^{d+1}$ be the family of all such open half-spaces.

The parameterization of half-spaces in terms of rays induces bijective mappings between the rays of $\vec{E}^{d+1}$ and the half-spaces of $\vec{\Pi}_o^{d+1}$ and $\vec{\Pi}_c^{d+1}$. Consider the transform $\rho_o$ that maps ray $\vec{u} \in \vec{E}^{d+1}$ to the open half-space $\rho_o(\vec{u}) = \{\vec{x} \in \vec{E}^{d+1} | \vec{u} \cdot \vec{x} >$

$0\} \in \vec{\Pi}_o^{d+1}$, and the open half-space $\vec{h}_v = \{\vec{x} \in \vec{E}^{d+1} | \vec{v} \cdot \vec{x} > 0\} \in \vec{\Pi}_o^{d+1}$ to the ray $\rho_o(\vec{h}_v) = \vec{v} \in \vec{E}^{d+1}$. It is easily verified that $\rho_o$ is indeed a bijection, and that $\rho_o(\rho_o(\vec{u})) = \vec{u}$ and $\rho_o(\rho_o(\vec{h}_v)) = \vec{h}_v$. Moreover, we observe that $\rho_o$ preserves incidence between rays and half-spaces, the proof of which is obvious:

**Observation 4.1** *Let $\vec{u}$ be a ray in $\vec{E}^{d+1}$, and let $\vec{h}_v$ be a half-space in $\vec{\Pi}_o^{d+1}$. Then $\vec{u}$ is contained in $\vec{h}_v$ if and only if ray $\rho_o(\vec{h}_v)$ is contained in half-space $\rho_o(\vec{u})$.*

In an entirely analogous manner, we may define the dual transform $\rho_c$ mapping rays in $\vec{E}^{d+1}$ to closed half-spaces in $\vec{\Pi}_c^{d+1}$. Observation 4.1 is also true for $\rho_c$, with the set $\vec{\Pi}_o^{d+1}$ replaced by $\vec{\Pi}_c^{d+1}$. For the ray $\vec{u} \in \vec{E}^{d+1}$, the hyperplane bounding the half-spaces $\rho_o(\vec{u})$ and $\rho_c(\vec{u})$ is also of interest. We define $\rho(\vec{u})$ to be the hyperplane $\{\vec{x} \in \vec{E}^{d+1} | \vec{u} \cdot \vec{x} = 0\}$ consisting of all rays of $\vec{E}^{d+1}$ orthogonal to $\vec{u}$. Since $\rho(\vec{u}) = \rho(-\vec{u})$, $\rho$ is not a bijection. However, each hyperplane passing through the origin is associated with exactly one pair of rays of the form $(\vec{u}, -\vec{u})$.

## 4.3 Linear Separation and Arrangements

Let $R$ and $G$ be sets of points in $E^d$, where the points of $R$ are labeled *red*, and the points of $G$ are labeled *green*. Let $P = \{p_1, p_2, \ldots, p_n\}$ be the union of $R$ and $G$. Each point $p_i$ of $P$, being labeled, corresponds uniquely to the ray $\vec{l}(p_i)$ of $\vec{E}^{d+1}$. Applying the transform $\rho_o$ to the rays of $\vec{l}(P)$ yields a collection of open half spaces in $\vec{\Pi}_o^{d+1}$, which we will call $\rho_o(\vec{l}(P))$. The collection of closed half spaces in $\vec{\Pi}_c^{d+1}$ obtained through the application of $\rho_c$ on $\vec{l}(P)$ will be denoted $\rho_c(\vec{l}(P))$, and the set of hyperplanes bounding these half-spaces will be called $\rho(\vec{l}(P))$.

The hyperplanes of $\rho(\vec{l}(P))$ form a *homogeneous* arrangement $\mathcal{A}(\rho(\vec{l}(P)))$ in $\vec{E}^{d+1}$, so called because each hyperplane may be expressed as a homogenous linear equation in $d+1$ variables. For the same reason, the arrangement is symmetric about the origin. Since each face of $\mathcal{A}(\rho(\vec{l}(P)))$ can be expressed as the intersection of half spaces whose bounding hyperplanes contain the origin, each ray of $\vec{E}^{d+1}$ belongs to precisely one face. As with the hyperplanes of $\Pi^{d+1}$ and the half-spaces of $\vec{\Pi}_c^{d+1}$ and $\vec{\Pi}_o^{d+1}$, the faces of $\mathcal{A}(\rho(\vec{l}(P)))$ may be parameterized in terms of the rays of $\vec{E}^{d+1}$. In this fashion, the faces can be considered not as sets of points in $\boldsymbol{E}^{d+1}$, but rather as sets of rays.

The only face of $\mathcal{A}(\rho(\vec{l}(P)))$ that could possibly contain no rays of $\vec{E}^{d+1}$ is a vertex at the origin. In fact, it is the only vertex possible in a homogeneous arrangement, all other faces being unbounded. Strictly speaking, however, the origin is contained in no ray of $\vec{E}^{d+1}$, and no hyperplane of $\Pi^{d+1}$, except where it is convenient to consider it so. For this reason, the vertex at the origin is an artificial entity, the "empty set" of rays, whose presence completes the structure of the homogeneous arrangement.

Each ray of $\vec{E}^{d+1}$ may be classified according to the region of $\mathcal{A}(\rho(\vec{l}(P)))$ in which it lies. Usually, this is accomplished by evaluating the ray according to its position with respect to each hyperplane of the arrangement - information obtainable by means of the ray inner product. Given a ray $\vec{u} \in \vec{E}^{d+1}$, its position vector is $\nu(\vec{u}) = (\nu_1(\vec{u}), \nu_2(\vec{u}), \ldots, \nu_n(\vec{u}))$, where $\nu_i(\vec{u}) = \vec{l}(p_i) \cdot \vec{u}$, for all $i = 1, \ldots, n$. Thus, if $\nu_i(\vec{u}) = 0$, $\vec{u}$ lies in the hyperplane $\rho(\vec{l}(p_i))$ and the closed half-space $\rho_c(\vec{l}(p_i))$, but not in the open half-space $\rho_o(\vec{l}(p_i))$. If $\nu_i(\vec{u}) = 1$, $\vec{u}$ is contained in both half-spaces, and if $\nu_i(\vec{u}) = -1$, it is contained in neither.

Since every face $f$ of the arrangement (other than the possible vertex at the origin) consists of rays sharing a common position vector, we will let $\nu(f) = \nu(\vec{u})$ for any $\vec{u} \in f$. To the vertex at the origin, if it exists, we assign the position vector $(0, 0, \ldots, 0)$, since it is the common intersection of all the hyperplanes of $\rho(\vec{l}(P))$.

The arrangement $\mathcal{A}(\rho(\vec{l}(P)))$, together with the half-spaces of $\rho_o(\vec{l}(P))$ and $\rho_c(\vec{l}(P))$, completely captures the combinatorial nature of the linearly separable components of $P$. Thus, we will refer to the hyperplane arrangement $\mathcal{A}(\rho(\vec{l}(P)))$, together with the position vectors as defined above, as the homogeneous *half-space* arrangement $\mathcal{A}(\rho_o(\vec{l}(P)))$ in $\boldsymbol{E}^{d+1}$.

The following lemma shows how the linearly separable components of $P$ with respect to the hyperplane $h$ of $\boldsymbol{E}^d$ may be derived from the positions of the rays $\vec{u}$ and $-\vec{u}$ in the arrangement, where $\vec{u}$ is the ray such that $\rho(\vec{u}) = \rho(-\vec{u}) = \vec{l}(h)$:

**Lemma 4.2** *Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of labeled points in $\boldsymbol{E}^d$, and let $h$ be a hyperplane. Let $\vec{u}$ and $-\vec{u}$ be the rays of $\vec{E}^{d+1}$ such that $\rho(\vec{u}) = \rho(-\vec{u}) = \vec{l}(h)$. Then*

1. *The strictly linearly separable components of $P$ with respect to $h$ are* $\{p_i \in P \mid \vec{u} \in \rho_o(\vec{l}(p_i))\}$ *and* $\{p_i \in P \mid -\vec{u} \in \rho_o(\vec{l}(p_i))\}$.
2. *The non-strictly linearly separable components of $P$ with respect to $h$ are* $\{p_i \in P \mid \vec{u} \in \rho_c(\vec{l}(p_i))\}$ *and* $\{p_i \in P \mid -\vec{u} \in \rho_c(\vec{l}(p_i))\}$.

**Proof** We will prove only the first claim, the proof of the second being similar to that of the first. Let $R$ and $G$ be the sets of red and green points of $P$, respectively. Let $h$ be parameterized as $\{x_H \in \boldsymbol{E}^d \mid u_H \cdot x_H = 0\}$, where (without loss of generality) $u_0$ is assumed to be positive. The two open half-spaces bounded by $h$ can then be expressed in the cartesian setting as

$$
\begin{aligned}
\vec{l}(h_>) &= \{\vec{x} \in \vec{E}_R^{d+1} \mid \vec{u} \cdot \vec{x} > 0\} \cup \{\vec{x} \in \vec{E}_G^{d+1} \mid \vec{u} \cdot \vec{x} < 0\} \text{ and} \\
\vec{l}(h_<) &= \{\vec{x} \in \vec{E}_R^{d+1} \mid \vec{u} \cdot \vec{x} < 0\} \cup \{\vec{x} \in \vec{E}_G^{d+1} \mid \vec{u} \cdot \vec{x} > 0\}.
\end{aligned}
$$

By definition, the strictly linearly separable components of $P$ with respect to $h$ are $C_1 = (R \cap h_>) \cup (G \cap h_<)$ and $C_2 = (R \cap h_<) \cup (G \cap h_>)$. In the cartesian setting of $\boldsymbol{E}^{d+1}$ we have

$$
\begin{aligned}
\vec{l}(C_1) &= (\vec{l}(R) \cap \{\vec{x} \in \vec{E}_R^{d+1} \mid \vec{u} \cdot \vec{x} > 0\}) \cup (\vec{l}(R) \cap \{\vec{x} \in \vec{E}_G^{d+1} \mid \vec{u} \cdot \vec{x} < 0\}) \\
&\quad \cup (\vec{l}(G) \cap \{\vec{x} \in \vec{E}_R^{d+1} \mid \vec{u} \cdot \vec{x} < 0\}) \cup (\vec{l}(G) \cap \{\vec{x} \in \vec{E}_G^{d+1} \mid \vec{u} \cdot \vec{x} > 0\}) \\
&= \{\vec{x} \in \vec{l}(R) \mid \vec{u} \cdot \vec{x} > 0\} \cup \{\vec{x} \in \vec{l}(G) \mid \vec{u} \cdot \vec{x} > 0\} \\
&= \{\vec{l}(p_i) \in \vec{l}(P) \mid \nu_i(\vec{u}) > 0\},
\end{aligned}
$$

and similarly

$$
\vec{l}(C_2) = \{\vec{l}(p_i) \in \vec{l}(P) \mid \nu_i(\vec{u}) < 0\} = \{\vec{l}(p_i) \in \vec{l}(P) \mid \nu_i(-\vec{u}) > 0\}
$$

Interpreting these conditions in the context of the half-spaces of $\rho_o(\vec{l}(P))$ yields $C_1 = \{p_i \in P \mid \vec{u} \in \rho_o(\vec{l}(p_i))\}$ and $C_2 = \{p_i \in P \mid -\vec{u} \in \rho_o(\vec{l}(p_i))\}$ □

An immediate implication of Lemma 4.2 concerns the strong and weak separation of $R$ and $G$. Let $\chi_o(\vec{x})$ and $\chi_c(\vec{x})$ be the number of half-spaces of $\rho_o(\vec{l}(P))$ and $\rho_c(\vec{l}(P))$, respectively, that contain ray $\vec{x} \in \vec{E}^{d+1}$. If $\vec{u}$ is (almost) any ray of $\vec{E}^{d+1}$ such that $\chi_o(\vec{u}) = n$, then the hyperplane $h$ such that $\vec{l}(h) = \rho(\vec{u})$ is a strong strict linear separator of $R$ and $G$ in $E^d$. If instead $\chi_o(\vec{u}) < n$, and there exists no $\vec{v} \in \vec{E}^{d+1}$ such that $\chi_o(\vec{v}) > \chi_o(\vec{u})$, then $h$ is a weak strict linear separator of $R$ and $G$. The same is true for non-strict separation, using $\chi_c(\vec{u})$ in place of $\chi_o(\vec{u})$. The only rays of $\vec{E}^{d+1}$ that are exceptions to the above observations are $\vec{r}_\infty = \{\lambda(1,0,\ldots,0)_C \in E^{d+1} | \lambda > 0\}$ and its opposite, $-\vec{r}_\infty$, as it is this pair of rays that maps to $\vec{E}_0^{d+1}$ under $\rho$.

Another implication of Lemma 4.2 concerns the linearly separable subsets of $P$, both strict and non-strict. Given any subset $Q$ of $P$, we wish to know whether the sets $Q \cap R$ and $Q \cap G$ are linearly separable. If we pose this question in the context of the arrangement $\mathcal{A}(\rho_o(\vec{l}(Q)))$, the lemma leads us to the following corollary for the strict case:

**Corollary 4.3** *Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of labeled points in $E^d$, and let $Q$ be a subset of $P$. The points of $Q$ are strictly linearly separable if and only if there exists some ray $\vec{u} \in \vec{E}^{d+1}$ contained in the intersection of all the half-spaces of $\rho_o(\vec{l}(Q))$.*

The corollary holds equally well for non-strict linear separability when considering $\rho_c$ in place of $\rho_o$.

The faces of $\mathcal{A}(\rho_o(\vec{l}(P)))$ may also be related to the linearly separable subsets and components of the labeled point set $P$ in $E^d$. Since all rays forming a given face $f$ of $\mathcal{A}(\rho_o(\vec{l}(P)))$ have identical position vectors, and are contained in the same half-spaces of $\rho_o(\vec{l}(P))$ and $\rho_c(\vec{l}(P))$, we let $\chi_o(f) = \chi_o(\vec{u})$ and $\chi_c(f) = \chi_c(\vec{u})$ for any ray $\vec{u} \in f$. In this manner, every face of the arrangement (other than the possible vertex at the origin) can be said to correspond to one strict and one non-strict linearly separable component – so long as the vertex at the origin exists. Without it, there would be some face $g$ such that if ray $\vec{u}$ is contained in $g$, its opposite ray $-\vec{u}$ is also contained in $g$. The face $g$ would then correspond to two components of each type.

**Lemma 4.4** *Let $P$ be a set of labeled points of $E^d$. There exists a vertex of $\mathcal{A}(\rho_o(\vec{l}(P)))$ at the origin of $E^{d+2}$ if and only if there exists no hyperplane of $E^d$ containing all points of $P$.*

**Proof** Let $h$ be a hyperplane in $E^d$. Let $\vec{u}$ and $-\vec{u}$ be the rays of $\vec{E}^{d+1}$ such that $\rho(\vec{u}) = \rho(-\vec{u}) = \vec{l}(h)$. Let $p$ be a point of $P$. By Observation 4.1, $\vec{l}(h)$ contains $\vec{l}(p)$

if and only if $\rho(\vec{l}(p))$ contains $\vec{u}$ and $-\vec{u}$. Therefore $h$ contains all points of $P$ if and only if $\vec{u}$ and $-\vec{u}$ are contained in all hyperplanes of $\mathcal{A}(\rho_o(\vec{l}(P)))$; that is, if there is no vertex of $\mathcal{A}(\rho_o(\vec{l}(P)))$ at the origin. □

The fact that some face of the arrangement must contain $\vec{r}_\infty$ causes no great inconvenience, as the following argument shows. first, we observe that no ray of $\vec{l}(P)$ lies in $\vec{E}_0^{d+1}$. Hence there exists no ray $\vec{l}(p) \in \vec{l}(P)$ such that hyperplane $\rho(\vec{l}(p))$ contains $\vec{r}_\infty$. Therefore $\vec{r}_\infty$ and $-\vec{r}_\infty$ can only be contained in cells of $\mathcal{A}(\rho_o(\vec{l}(P)))$, and are not the only rays of these cells. Despite containing rays which do not correspond to a separator under $\rho$, these cells still may be associated with components of $R$ and $G$.

## 4.4 Dual Transforms for Spherical Separation

In the previous sections, a transformation of instances of labeled sets in $\boldsymbol{E}^d$ into homogeneous half-space arrangements in $\boldsymbol{E}^{d+1}$ was exhibited - a transformation that preserves the combinatorial structure of linear separation. We shall now do the same for spherical separation, also by means of a transformation into the setting of homogeneous arrangements, this time in $\boldsymbol{E}^{d+2}$. Following the precedent established for linear separation, the transformation for the spherical case makes the following correspondences:

| PRIMAL | | DUAL |
|---:|---|---|
| point | $\longleftrightarrow$ | hyperplane through origin, |
| hypersphere | $\longleftrightarrow$ | open ray with endpoint the origin, |
| labeled point | $\longleftrightarrow$ | half-space with bounding hyperplane through origin. |

The approach will be based upon one used by Edelsbrunner and Seidel [ES86] to relate Voronoi diagrams in $\boldsymbol{E}^d$ with hyperplane arrangements in $\boldsymbol{E}^{d+1}$. Before introducing the spherical separation transform, we briefly describe their Voronoi transform and how it relates to spherical separation.

Consider the bijective mapping $\mathcal{V}$ of every site $v$ of the set of Voronoi sites $V$ (represented using cartesian coordinates) onto the hyperplane $h_v(x) = 2v \cdot x - v \cdot v$ in $\boldsymbol{E}^{d+1}$. This hyperplane may be visualized as the tangent hyperplane to the unit

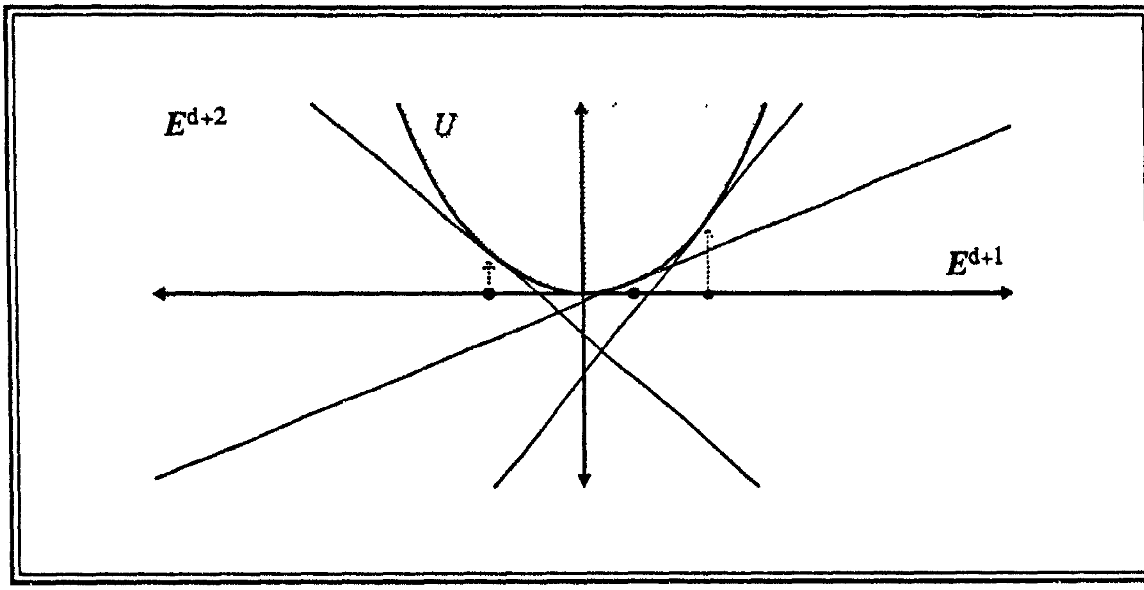


Figure 4.2: Mapping of Voronoi sites in $E^d$ to hyperplanes in $E^{d+1}$

paraboloid $U(x) = x \cdot x$ in $E^{d+1}$ (see Figure 4.2). Since the square of the distance from $x$ to $v$ is $\|x - v\|^2 = x \cdot x - 2v \cdot x + v \cdot v$, $\|x - v\| = \sqrt{U(x) - h_v(x)}$. Hence the closest Voronoi site to $x \in E^d$ is that whose associated hyperplane $h_v$ is such that $h_v(x) \geq h_w(x)$ for all $w \in V$.

The arrangement of these hyperplanes provides much more information than is required for the construction of Voronoi diagrams. If the hyperplanes are evaluated at $x \in E^d$, and are then considered in order of decreasing value, their associated Voronoi sites are ordered by increasing distance from the point $x$. Thus a point $y = (y_1, y_1, \ldots, y_{d+1})_C$ located in this arrangement could be thought to correspond to a hypersphere in the primal space $E^d$ with centre $(y_1, y_1, \ldots, y_d)_C$ and radius

$$\sqrt{\sum_{i=1}^{d} {y_i}^2 - y_{d+1}}.$$

The sites inside the hypersphere in the primal space would have as dual hyperplanes those "above" the point $y$, where "above" is defined as being in the direction of the positive $x_{d+1}$-axis. The sites outside the hypersphere would have as dual hyperplanes those "below" the point $y$.

One feature of the linear separation transform that is shared by this transformation strategy is that potential separating surfaces are mapped into locations within an arrangement of hyperplanes. However, the Voronoi transform cannot differentiate between two sets of points in $\boldsymbol{E}^d$. In the linear separation case, this differentiation was possible due to the introduction of labeled points of $\boldsymbol{E}^d$, and the conversion of these points into rays of $\vec{E}^{d+1}$. For the case of spherical separation, we shall apply the techniques of the Voronoi transform in the conversion of the labeled points of $\boldsymbol{E}^d$ into hyperplanes and half-spaces in $\boldsymbol{E}^{d+2}$.

Let $u$ be a point of $\boldsymbol{E}^d$, represented in cartesian coordinates. The image of $u$ under the Voronoi transform $\mathcal{V}$ is the hyperplane

$$\mathcal{V}(u) = \{x_C \in \boldsymbol{E}^{d+1} | -u \cdot u + \sum_{i=1}^{d} 2u_i x_i - x_{d+1} = 0\}.$$

Expressed in homogeneous coordinates, this expression becomes

$$\mathcal{V}(u) = \{x_H \in \boldsymbol{E}^{d+1} | (-u \cdot u, 2u_1, 2u_2, \ldots, 2u_d, -1) \cdot x_H = 0\}.$$

Exploiting the relationship between points in $\boldsymbol{E}^{d+1}$ represented using homogeneous coordinates, and the rays of $\vec{E}^{d+2}$, we may convert the hyperplane $\mathcal{V}(u)$ into a hyperplane of $\boldsymbol{E}^{d+2}$. With this goal in mind, we introduce notation that will simplify the representation of this hyperplane in $\boldsymbol{E}^{d+2}$.

Let $p$ be a labeled point of $\boldsymbol{E}^d$, represented using cartesian coordinates. The ray of $\vec{E}^{d+2}$ with which we will associate $p$ shall be given by

$$\vec{s}(p) = \begin{cases} \{\lambda(p \cdot p, -2p_1, -2p_2, \ldots, -2p_d, 1) \in \boldsymbol{E}^{d+2} | \lambda > 0\} & \text{if } p \text{ is } \textit{red}, \text{ and} \\ \{-\lambda(p \cdot p, -2p_1, -2p_2, \ldots, -2p_d, 1) \in \boldsymbol{E}^{d+2} | \lambda > 0\} & \text{if } p \text{ is } \textit{green}. \end{cases}$$

Although this relationship is not bijective, it is certainly injective. A ray of $\vec{E}^{d+2}$ can be the image of only one labeled point of $\boldsymbol{E}^d$ under this transformation. Also, as one would expect, the ray $-\vec{s}(p)$ is associated with that labeled point with the same location as $p$, but with the opposite labeling. If $P$ is a set of labeled points in $\boldsymbol{E}^d$, the set of rays that are the images of points of $P$ will be denoted by $\vec{s}(P)$.

Except for the the points at the origin, the red points of $\boldsymbol{E}^d$ are mapped onto rays of $\vec{E}_R^{d+2}$, and the green points are mapped onto rays of $\vec{E}_G^{d+2}$. The origin of $\boldsymbol{E}^d$

is mapped to the ray $\{\lambda(0,0,\ldots,0,1) \in E^{d+2} | \lambda > 0\}$ if labeled red, and to the ray $\{\lambda(0,0,\ldots,0,-1) \in E^{d+2} | \lambda > 0\}$ if labeled green.

Let $u$ be a labeled point of $E^d$. If we extend the definition of the ray inner product to rays of $\vec{E}^{d+2}$, we may express the set of rays of $\vec{E}^{d+2}$ corresponding to the points of hyperplane $\mathcal{V}(u)$ as $\sigma(\vec{s}(u)) = \{\vec{x} \in \vec{E}^{d+2} | \vec{s}(u) \cdot \vec{x} = 0\}$. Since $\sigma(-\vec{u}) = \sigma(\vec{u})$, the labeling of $u$ is not relevant in the determination of the rays of this hyperplane. However, the labeling does provide information that determines the orientation of the half-spaces $\sigma_o(\vec{u})$ and $\sigma_c(\vec{u})$.

The function $\sigma$, as well as the dual transforms $\sigma_o$ and $\sigma_c$, may be used to convert the labeled point set $P = \{p_1, p_2, \ldots, p_n\}$ of $E^d$ into the homogeneous half-space arrangement $\mathcal{A}(\sigma_o(\vec{s}(P)))$ in $E^{d+2}$. The position vector of the ray $\vec{u} \in \vec{E}^{d+2}$ with respect to the arrangement is $\nu(\vec{u}) = (\nu_1(\vec{u}), \nu_2(\vec{u}), \ldots, \nu_n(\vec{u}))$, where $\nu_i(\vec{u}) = \vec{s}(p_i) \cdot \vec{u}$, for all $i = 1, \ldots, n$. Thus, if $\nu_i(\vec{u}) = 0$, $\vec{u}$ lies in the hyperplane $\sigma(\vec{s}(p_i))$ and the closed half-space $\sigma_c(\vec{s}(p_i))$, but not in the open half-space $\sigma_o(\vec{s}(p_i))$. If $\nu_i(\vec{u}) = 1$, $\vec{u}$ is contained in both half-spaces, and if $\nu_i(\vec{u}) = -1$, it is contained in neither. As in the previous section, the position vector of a face $f$ of the arrangement shall be taken to be the position vector shared by the rays contained in $f$.

To show how the arrangement $\mathcal{A}(\sigma_o(\vec{s}(P)))$ captures the combinatorial properties of the spherically separable components of $P$, we must first show how the hyperspheres of $E^d$ relate to the rays of $\vec{E}^{d+2}$. The hyperspheres of $E^d$ are often parameterized in terms of the cartesian coordinates of their centres, and their radii. Let $s$ be a hypersphere with centre $c = (c_1, c_1, \ldots, c_d)_C$ and radius $r$. An equally viable parameterization for $s$ is the $(d+1)$-tuple $(s_1, s_2, \ldots, s_{d+1})$, where $s_i = c_i$ for $i = 1, 2, \ldots, d$, and $s_{d+1} = \sum_{i=1}^{d} s_i{}^2 - r^2$. This tuple may be interpreted as the cartesian coordinates of a point in $E^{d+1}$. It should be noted that not every point of $E^{d+1}$ corresponds to a hypersphere of $E^d$ in this manner; if $s_{d+1} \geq \sum_{i=1}^{d} s_i{}^2$, then $r^2 \leq 0$. Consider now the homogeneous coordinate equivalents of the points of $E^{d+1}$. If $x \in E^{d+1}$ is represented by the homogeneous tuple $x_H = (x_0, x_1, \ldots, x_{d+1})_H$, where $x_0 \neq 0$, point $x$ is

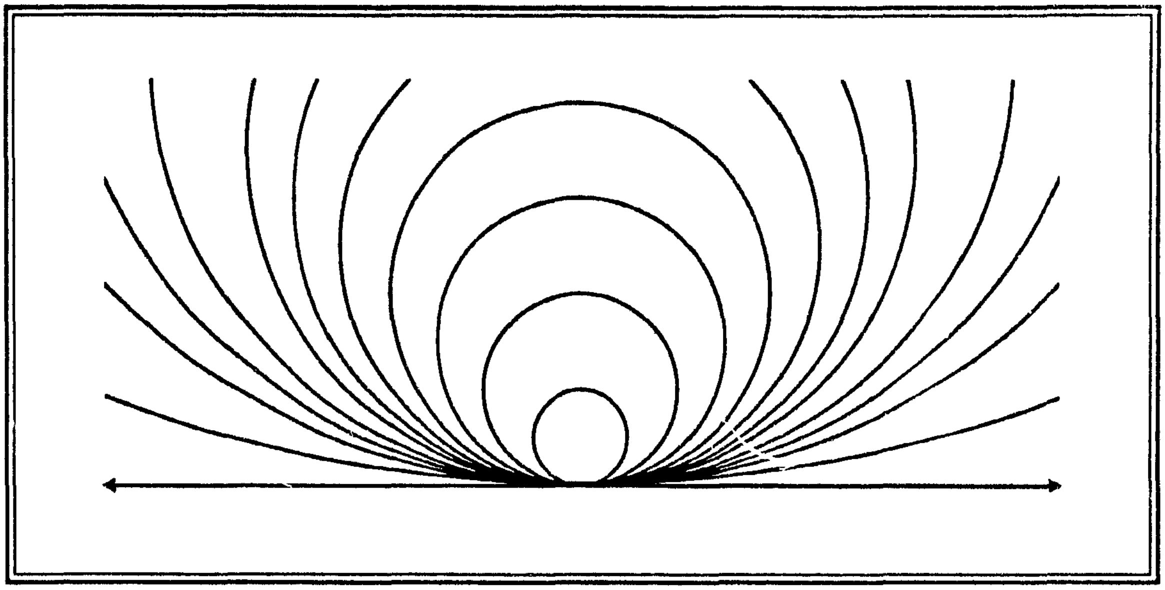

Figure 4.3: A sequence of hyperspheres tending toward a hyperplane in $\boldsymbol{E}^d$

associated with the unique hypersphere of $\boldsymbol{E}^d$ with centre and squared radius

$$ctr(x) = \frac{1}{x_0}(x_1, x_1, \ldots, x_d)_C \text{ and } rad^2(x) = ctr(x) \cdot ctr(x) - \frac{x_{d+1}}{x_0},$$

respectively.

In Chapter 3 we allowed the use of hyperplanes as degenerate spherical separators, since a hyperplane of $\boldsymbol{E}^d$ may be regarded as the limit of some infinite sequence of hyperspheres whose centres are successively farther from the origin, and whose radii grow proportionately (see Figure 4.3). A hyperplane of $\boldsymbol{E}^d$ may be parameterized as follows: let $\alpha$ be some fixed real value, and let $c = (c_1, c_1, \ldots, c_d)_C$ be a point of $\boldsymbol{E}^d$ other than the origin. Let $\hat{c}$ represent the unit vector in the direction of $c$. The unique hyperplane of $\boldsymbol{E}^d$ with normal vector $c$ and passing through the point $\alpha\hat{c}$ is given using cartesian coordinates by the expression $h = \{x \in \boldsymbol{E}^d | c \cdot x = \alpha||c||\}$. We will determine the homogeneous tuples to which hyperplane $h$ may be associated by first constructing a sequence of hyperspheres that converges to $h$, and then taking the limit of homogeneous tuples representing these hyperspheres.

Let $\psi(t)$ be the hypersphere of $\boldsymbol{E}^d$ with radius $r$ and with centre $tc$, where $t > \frac{\alpha}{||c||}$. In addition, let us restrict the radius $r$ to be $||tc|| - \alpha$, for some fixed real value $\alpha$. The condition on $t$ ensures that the radius of $\psi(t)$ is positive. The point $\alpha\hat{c}$ belongs to $\psi(t)$ for any $t > \frac{\alpha}{\|c\|}$, since

$$\|tc - \alpha\hat{c}\| = (\|tc\| - \alpha)\|\hat{c}\| = r.$$

Clearly, $h$ is tangent to $\psi(t)$ for all $t > \frac{\alpha}{\|c\|}$, since it passes through $\alpha\hat{c}$ and is orthogonal to the line passing through $\alpha\hat{c}$ and $tc$. Since the radius of $\psi(t)$ increases as $t \to \infty$, $h = \lim_{t\to\infty} \psi(t)$.

Let $x_H(t)$ be a homogeneous tuple of $E^{d+1}$ representing the hypersphere $\psi(t)$. Hence

$$\begin{aligned} x_H(t) &= \lambda\left(1, tc_1, \ldots, tc_d, (tc)\cdot(tc) - r^2\right) \\ &= \lambda\left(1, tc_1, \ldots, tc_d, (\|tc\| - r)(\|tc\| + r)\right) \\ &= \lambda\left(1, tc_1, \ldots, tc_d, 2\alpha\|tc\| - \alpha^2\right), \end{aligned}$$

for any choice of $\lambda \neq 0$. Let $t \to \infty$ and $\lambda t \to \xi$, for some real-valued choice of $\xi \neq 0$. These conditions together cause $\lambda$ to tend to zero. The resultant limit of $x_H(t)$ is

$$\xi\,(0, c_1, \ldots, c_d, 2\alpha\|c\|)$$

Noting that $h = \{x \in E^d | \, c \cdot x = \alpha\|c\|\}$, we conclude that the homogeneous tuples of the form $\xi(0, u_1, u_2, \ldots, u_{d+1})$, and only these tuples, correspond to the hyperplane $\{x \in E^d | \, 2\sum_{i=1}^{d} u_i x_i = u_{d+1}\}$ of $E^d$.

As with the linear separation transformations, the homogeneous tuples of $E^{d+1}$ representing hyperspheres of $E^d$ (both degenerate and non-degenerate) shall be interpreted as cartesian coordinates in $E^{d+2}$. Hence every such hypersphere corresponds to a pair of oppositely-oriented rays of $\vec{E}^{d+2}$: the non-degenerate hyperspheres correspond to rays of $\vec{E}_R^{d+2}$ and $\vec{E}_G^{d+2}$, and the degenerate hyperspheres are associated with rays of $\vec{E}_0^{d+2}$. Again, not all rays of $\vec{E}^{d+2}$ are associated with hyperspheres of $E^d$. The subset of $\vec{E}^{d+2}$ whose elements do represent non-degenerate hyperspheres or hyperplanes will be called $\vec{E}_\star^{d+2}$. If $\vec{u}$ is an element of $\vec{E}_\star^{d+2}$, then we let $\gamma(\vec{u})$ denote the unique (possibly degenerate) hypersphere associated with $\vec{u}$.

## 4.5 Spherical Separation and Arrangements

So far, we have detailed a transform mapping the labeled points of $P$ in $\boldsymbol{E}^d$ into a homogeneous arrangement in $\boldsymbol{E}^{d+2}$, and related rays in $\boldsymbol{E}^{d+2}$ with hyperspheres and hyperplanes in $\boldsymbol{E}^d$. However, we have not yet seen how these transformations mesh together. The following lemma shows how the combinatorial structure of the dual arrangement of a point set in $\boldsymbol{E}^d$ reflects the combinatorial structure of this set with respect to spherical separation.

**Lemma 4.5** *Let $P = \{p_1, p_2, \dots, p_n\}$ be a set of labeled points in $\boldsymbol{E}^d$, and let $s$ be a (possibly degenerate) hypersphere. Let $\vec{u}$ and $-\vec{u}$ be the rays of $\vec{E}_*^{d+2}$ such that $\gamma(\vec{u}) = \gamma(-\vec{u}) = s$. Then*

1. *The strictly spherically separable components of $P$ with respect to $s$ are $\{p_i \in P \mid \vec{u} \in \sigma_o(\vec{s}(p_i))\}$ and $\{p_i \in P \mid -\vec{u} \in \sigma_o(\vec{s}(p_i))\}$.*
2. *The non-strictly spherically separable components of $P$ with respect to $s$ are $\{p_i \in P \mid \vec{u} \in \sigma_c(\vec{s}(p_i))\}$ and $\{p_i \in P \mid -\vec{u} \in \sigma_c(\vec{s}(p_i))\}$.*

**Proof** We will prove only the first claim, the proof of the second being similar to that of the first. Let $R$ and $G$ be the sets of red and green points of $P$, respectively, and let $u = (u_0, u_1, \dots, u_{d+1})$ be a point contained in $\vec{u}$. Without loss of generality, we assume that $u_0 \geq 0$. Consider point $p_i \in P$. If ray $\vec{u}$ is contained in $\sigma_o(\vec{s}(p_i)) = \{\vec{x} \in \vec{E}^{d+2} \mid \vec{s}(p_i) \cdot \vec{x} > 0\}$, and if $p_i$ is red, then

$$\left(\sum_{i=1}^{d} p_i{}^2, -2p_1, -2p_2, \dots, -2p_d, 1\right) \cdot u \;>\; 0,$$

or equivalently,

$$u_0 \sum_{i=1}^{d} p_i{}^2 - 2\sum_{i=1}^{d} u_i p_i + u_{d+1} \;>\; 0. \tag{4.1}$$

Instead, if $p_i$ is green, then similarly

$$u_0 \sum_{i=1}^{d} p_i{}^2 - 2\sum_{i=1}^{d} u_i p_i + u_{d+1} \;<\; 0. \tag{4.2}$$

If $s$ is non-degenerate, then $u_0 > 0$, and the parameterization of $s$ using cartesian coordinates is

$$\begin{aligned} s &= \left\{ x \in \boldsymbol{E}^d \,\middle|\, \sum_{i=1}^{d} \left( x_i - \frac{u_i}{u_0} \right)^2 = \sum_{i=1}^{d} \left( \frac{u_i}{u_0} \right)^2 - \frac{u_{d+1}}{u_0} \right\} \\ &= \left\{ x \in \boldsymbol{E}^d \,\middle|\, \sum_{i=1}^{d} x_i{}^2 - 2\sum_{i=1}^{d} \left( \frac{u_i}{u_0} \right) x_i + \frac{u_{d+1}}{u_0} = 0 \right\} \\ &= \left\{ x \in \boldsymbol{E}^d \,\middle|\, u_0 \sum_{i=1}^{d} x_i{}^2 - 2\sum_{i=1}^{d} u_i x_i + u_{d+1} = 0 \right\} \end{aligned}$$

The open connected components of $\boldsymbol{E}^d \setminus s$ are the regions

$$\begin{aligned} s_> &= \left\{ x \in \boldsymbol{E}^d \,\middle|\, u_0 \sum_{i=1}^{d} x_i{}^2 - 2\sum_{i=1}^{d} u_i x_i + u_{d+1} > 0 \right\} \text{ and} \\ s_< &= \left\{ x \in \boldsymbol{E}^d \,\middle|\, u_0 \sum_{i=1}^{d} x_i{}^2 - 2\sum_{i=1}^{d} u_i x_i + u_{d+1} < 0 \right\}, \end{aligned}$$

respectively. By definition, the strictly spherically separable components of $P$ with respect to $s$ are $C_1 = (R \cap s_>) \cup (G \cap s_<)$ and $C_2 = (R \cap s_<) \cup (G \cap s_>)$. Of the points of $R$, only those contained in $s_>$ satisfy (4.1), and of the points of $G$, only those contained in $s_<$ satisfy (4.2). Therefore $C_1 = \{p_i \in P | \vec{u} \in \sigma_o(\vec{s}(p_i))\}$. The same argument applied for the ray $-\vec{u}$ yields $C_2 = \{p_i \in P | -\vec{u} \in \sigma_o(\vec{s}(p_i))\}$.

If $s$ is degenerate, and the parameterization of $s$ using cartesian coordinates is $\{x \in \boldsymbol{E}^d | \sum_{i=1}^{d} 2u_i x_i = u_{d+1}\}$, the open half-spaces bounded by $s$ are

$$\begin{aligned} s_> &= \left\{ x \in \boldsymbol{E}^d \,\middle|\, -2\sum_{i=1}^{d} u_i x_i + u_{d+1} > 0 \right\} \text{ and} \\ s_< &= \left\{ x \in \boldsymbol{E}^d \,\middle|\, -2\sum_{i=1}^{d} u_i x_i + u_{d+1} < 0 \right\}, \end{aligned}$$

respectively. Noting that $u_0 = 0$, the arguments used for the non-degenerate case also suffice here. Hence the result follows. □

One of the by-products of this lemma concerns the incidence relation between points and hyperspheres of $\boldsymbol{E}^d$. If $\sigma$ is considered in place of $\sigma_o$ in the above proof, one arrives at the following corollary:

**Corollary 4.6** *Let $p$ be a point of $\mathbf{E}^d$, and let $\vec{u}$ be a ray of $\vec{E}_\star^{d+2}$. Then point $p$ is contained in hypersphere $\gamma(\vec{u})$ if and only if ray $\vec{u}$ is contained in hyperplane $\sigma(\vec{s}(p))$ of $\mathbf{E}^{d+2}$.*

As in the linear separation case, we may define $\chi_o(\vec{x})$ and $\chi_c(\vec{x})$ to be the number of half-spaces of $\sigma_o(\vec{s}(P))$ and $\sigma_c(\vec{s}(P))$, respectively, that contain ray $\vec{x}$ of $\vec{E}_\star^{d+2}$. These quantities may also be applied to the faces of $\mathcal{A}(\sigma_o(\vec{s}(P)))$, since the rays constituting a particular face of the arrangement share a common position vector. If $\vec{u}$ is any ray of $\vec{E}_\star^{d+2}$ such that $\chi_o(\vec{u}) = n$, then the hypersphere $s$ such that $\vec{s}(s) = \sigma(\vec{u})$ is a strong strict spherical separator of $R$ and $G$ in $\mathbf{E}^d$. If instead $\chi_o(\vec{u}) < n$, and there exists no $\vec{v} \in \vec{E}_\star^{d+2}$ such that $\chi_o(\vec{v}) > \chi_o(\vec{u})$, then $s$ is a weak strict spherical separator of $R$ and $G$. The same is true for non-strict separation, using $\chi_c(\vec{u})$ in place of $\chi_o(\vec{u})$.

We now state two additional corollaries of Lemma 4.5. The next corollary is analogous to Corollary 4.3, and concerns the spherical separability of subsets of $P$. Corollary 4.8 deals with the distinction between the separable components of $P$ with respect to the hypersphere $s$, relative to which points are contained in the interior of $s$, and which are contained in the exterior.

**Corollary 4.7** *Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of labeled points in $\mathbf{E}^d$, and let $Q$ be a subset of $P$. The points of $Q$ are strictly spherically separable if and only if there exists some ray $\vec{u} \in \vec{E}_\star^{d+2}$ contained in the intersection of all the half-spaces of $\sigma_o(\vec{s}(Q))$.*

**Corollary 4.8** *Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of labeled points in $\mathbf{E}^d$, and let $\vec{u}$ be a ray of $\vec{E}_\star^{d+2}$. Let $s$ be the hypersphere $\gamma(\vec{u})$ in $\mathbf{E}^d$. If $\vec{u} \in \vec{E}_R^{d+2}$, then the strict spherically separable component $C = \{p_i \in P \mid \vec{u} \in \sigma_o(\vec{s}(p_i))\}$ of $P$ with respect to $s$ consists of the red points exterior to $s$ and the green points interior to $s$. If instead $\vec{u} \in \vec{E}_G^{d+2}$, then $C$ consists of the green points exterior to $s$ and the red points interior to $s$.*

Both Corollary 4.7 and Corollary 4.8 hold for non-strict spherical separation when considering $\sigma_c$ in place of $\sigma_o$, and the closures of the exterior and interior of $s$ in place of the (open) exterior and interior. Corollary 4.8 indicates that if one limits one's attention to that part of $\mathcal{A}(\sigma_o(\vec{s}(P)))$ contained in $\vec{E}_R^{d+2}$, only the spherically separable components with red points exterior to the separator are considered. The part of $\mathcal{A}(\sigma_o(\vec{s}(P)))$ contained in $\vec{E}_G^{d+2}$ yields compcnents with green points exterior to the spherical separator, and the part of the arrangement contained in $\vec{E}_0^{d+2}$ correponds to degenerate spherical separators – that is, to linear separators. By taking the intersection of $\mathcal{A}(\sigma_o(\vec{s}(P)))$ with $\vec{E}_0^{d+2}$, one obtains a $(d+1)$-dimensional homogeneous arrangement that captures the combinatorial qualities of $P$ with respect to linear separation, equivalent in every respect to $\mathcal{A}(\rho_o(\vec{l}(P)))$.

One concern that will arise in later chapters is whether every face of $\mathcal{A}(\sigma_o(\vec{s}(P)))$ (except for the vertex at the origin, if it exists) contains a ray of $\vec{E}_\star^{d+2}$. This is indeed the case. Recall that in $\vec{E}_0^{d+2}$, only the rays $\vec{z} = \{\lambda(0,0,\ldots,0,1) \in E^{d+2} | \lambda > 0\}$ and $-\vec{z}$ are not members of $\vec{E}_\star^{d+2}$. However, given the labeled point $p \in P$, the ray inner product $\vec{s}(p) \cdot \vec{z}$ is always either 1 or $-1$, and therefore $\vec{z}$ and $-\vec{z}$ are contained in cells of $\mathcal{A}(\sigma_o(\vec{s}(P)))$. As a result, if $\vec{z}$ or $-\vec{z}$ is contained in cell $f$ of the arrangement, there exists some other ray $\vec{u}$ of $\vec{E}_0^{d+1}$ in $f$ such that $\vec{u} \in \vec{E}_\star^{d+2}$.

In $\vec{E}_R^{d+2}$ and $\vec{E}_G^{d+2}$, recall that the rays of the form $\{\lambda x \in E^{d+2} | \lambda > 0, x_0 \neq 0\}$ that are not elements of $\vec{E}_\star^{d+2}$ are those where $\sum_{i=1}^{d} x_i{}^2 - x_0 x_{d+1} \leq 0$; that is, those that would correspond to hyperspheres having the square of their radii less than or equal to zero. The rays of $\vec{E}_R^{d+2} \cup \vec{E}_G^{d+2}$ not in $\vec{E}_\star^{d+2}$ consist of two connected components symmetric with respect to the origin, one in $\vec{E}_R^{d+2}$ and the other in $\vec{E}_G^{d+2}$. Let us restrict our attention to the former, and denote it by $\varrho$.

Let $\vec{z}$ be a ray of $\varrho$, and let $p$ be a red point of $P$. Let $x \in E^{d+2}$ be that point of $\vec{z}$ where $x_0 = 1$. Then

$$
\begin{aligned}
(p \cdot p, -2p_1, -2p_2, \ldots, -2p_d, 1) \cdot x &= p \cdot p - 2\sum_{i=1}^{d} p_i x_i + x_{d+1} \\
&\geq p \cdot p - 2\sum_{i=1}^{d} p_i x_i + \sum_{i=1}^{d} x_i{}^2 \\
&\geq \|p - x'\|^2,
\end{aligned}
$$

where $x' = (x_1, x_2, \ldots, x_d)$ in $E^d$. This implies that $\vec{z}$ is contained in $\sigma(\vec{s}(p))$ if and only if $x' = p$, which is true for only one ray in $\vec{E}_R^{d+1}$. Otherwise, it is contained in $\sigma_o(\vec{s}(p))$. Similarly, if $p$ is labeled green, $\vec{z}$ is contained in $\sigma(\vec{s}(p))$ if and only if $x' = p$, but otherwise, it is contained in the complement of $\sigma_c(\vec{s}(p))$. Thus all but a finite number of rays of $\varrho$ are contained in a common cell $f$, the remaining rays being located in faces contained in the closure of $f$.

Let $g$ a face in the closure of $f$, where $dim(g) \geq 1$ and $g \cap \vec{E}_R^{d+2} \neq \emptyset$. If $dim(g) > 1$, then $g \cap \vec{E}_R^{d+2}$ contains an infinite number of rays, and therefore some ray of $\vec{E}_\star^{d+2}$ Otherwise, $g$ consists of a single ray of $\vec{E}_R^{d+2}$, and is contained in the intersection of at least $d$ hyperplanes of $\mathcal{A}(\sigma_o(\vec{s}(P)))$. Let this ray be an element of $\varrho$, say $\vec{z}$. Then $x' = p$ for at least $d$ choices of $p$ in $P$, which is an impossibility. Therefore $g$ contains some ray of $\vec{E}_\star^{d+2}$.

This leaves only the face $f$ to be considered. Let us assume that $f \cap \vec{E}_R^{d+2}$ is entirely contained in $\varrho$. Since the set $\varrho$ is closed relative to $\vec{E}_R^{d+2}$, the closure of $f$ in $\vec{E}_R^{d+2}$ must also be entirely contained in $\varrho$, which is a contradiction.

We summarize these arguments in the following lemma:

**Lemma 4.9** *Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of labeled points in $\boldsymbol{E}^d$, and let $f$ be a face of $\mathcal{A}(\sigma_o(\vec{s}(P)))$ of dimensionality at least* 1. *Then $f$ contains at least one ray of $\vec{E}_\star^{d+2}$.*

From this lemma we may conclude that the faces of $\mathcal{A}(\sigma_o(\vec{s}(P)))$ have the same significance as those of $\mathcal{A}(\rho_o(\vec{l}(P)))$, in that each face of both arrangements of dimensionality greater than zero corresponds to one strict and one non-strict component of the appropriate type – again, only if the arrangement has a vertex located at the origin.

**Lemma 4.10** *Let $P$ be a set of labeled points of $\boldsymbol{E}^d$. There exists a vertex of $\mathcal{A}(\sigma_o(\vec{s}(P)))$ at the origin of $\boldsymbol{E}^{d+2}$ if and only if there exists no hypersphere of $\boldsymbol{E}^d$, degenerate or non-degenerate, that contains all points of $P$*

**Proof** Let $s$ be a hypersphere in $\boldsymbol{E}^d$. Let $\vec{u}$ and $-\vec{u}$ be the rays of $\vec{E}^{d+1}$ such that $\gamma(\vec{u}) = \gamma(-\vec{u}) = s$. Let $p$ be a point of $P$. By Corollary 4.6, hyperplane $\sigma(\vec{u}) = \sigma(-\vec{u})$ contains ray $\vec{s}(p)$ of $\vec{E}_\star^{d+2}$ if and only if hyperplane $\sigma(\vec{s}(p))$ contains rays $\vec{u}$ and $-\vec{u}$. Therefore $s$ contains all points of $P$ if and only if $\vec{u}$ and $-\vec{u}$ are contained in all hyperplanes of $\sigma(\vec{s}(P))$; that is, if there is no vertex of $\mathcal{A}(\sigma_o(\vec{s}(P)))$ at the origin. □

# Chapter 5

# Separable Subsets and Components

## 5.1 Introduction

The setting of half-space arrangements will prove to be very convenient for both the investigation of combinatorial properties of sets of points with respect to linear and spherical separation, and the algorithms determining the various linear and spherical separators of point sets and other object classes. In this chapter, we address two combinatorial topics; the first concerns separable components, and the second concerns separable subsets.

In the next section, we will provide upper bounds on the maximum number of separable components of fixed size $k$ of a set of labeled points. More precisely, given a set $P$ of $n$ distinct labeled points of $E^d$, we will bound the number of separable components of $P$ of size less than or equal to $k$, in each of the linear and spherical cases, and the strict and non-strict cases.

The bound for the case of linear separation will also turn out to apply to the number of *k-sets* in $d$ dimensions. Let $h$ be a hyperplane in $E^d$, and let $h_>$ and $h_<$ be the two open half-spaces bounded by $h$. If $k$ is the cardinality of $P \cap h_>$, then the point set $P \cap h_>$ is called a *k-set* of $P$, and the point set $P \cap h_<$ is called an $(n-k)$-*set* of $P$. If the points of $P$ all share the same label, then the $k$-sets of $P$ are identical to

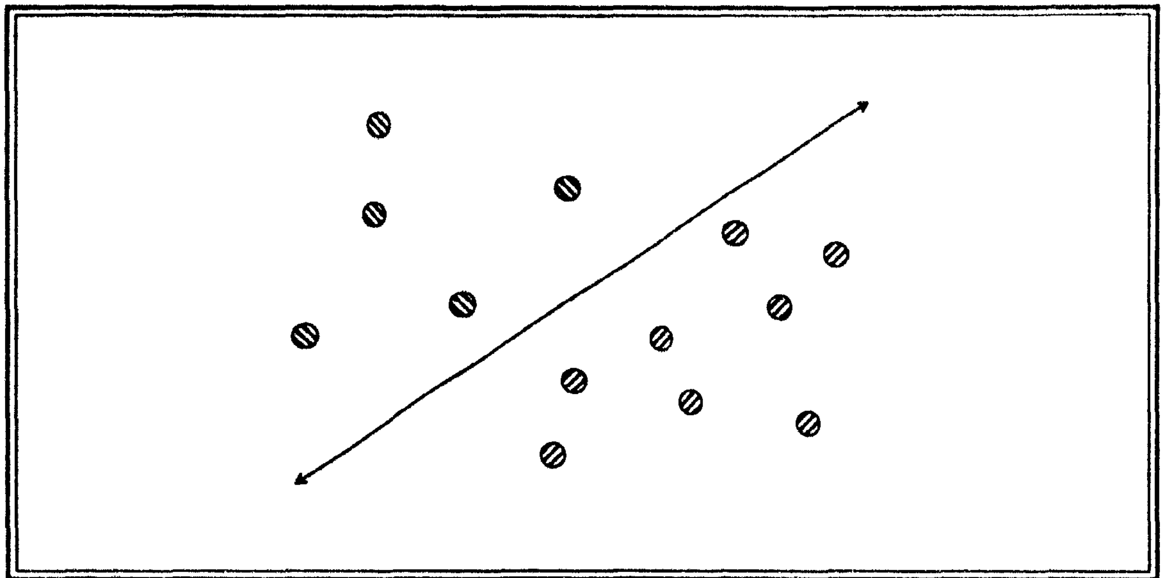

Figure 5.1: A 5-set and an 8-set

the strict linearly separable components of $P$ of size $k$ (see Figure 5.1).

The theory of $k$-sets has many applications, generally involving the analysis of space and time complexities of algorithms. Some notable examples include higher-order Voronoi diagram construction [Lee82,Ede87], half-space range queries [CP86, Cla88], and approximation of sets of points by hyperplanes [YKII88]. The first asymptotic bounds on the number of $k$-sets of $n$ points in the plane were developed by Lovász [Lov71] and Erdős, Lovász, Simmons, and Strauss [ELSS73]. These bounds of $O(n\sqrt{k})$ and $\Omega(n\log(k+1))$ are still the best known to date in two dimensions. In [Ede87], Edelsbrunner credits Raimund Seidel with an extension of this lower bound to higher dimensions for the case $k = n/2$, obtaining the bound $\Omega(n^{d-1}\log n)$. In three dimensions, Chazelle and Preparata [CP86] derived an upper bound of $O(nk^5)$, which was subsequently improved for large values of $k$ to $O(n^2k)$ by Cole, Sharir, and Yap [CSY87]. Very recently, Bárány, Füredi, and Lovász [BFL89] showed that for $k = n/2$, the bound may be reduced to $O(n^{2.998})$. In higher dimensions, Clarkson [Cla88] obtained an upper bound of $O(n^{\lfloor d/2\rfloor}k^{\lceil d/2\rceil})$ using random sampling methods.

In Chapter 3, the strong strict linear and spherical separability of a set of labeled points $P$ was related to the existence of separable subsets of $P$ having a certain cardinality dependent upon $d$ alone. In Section 5.3, expressions shall be given for the number of linearly and spherically separable subsets, both strict and non-strict, and of both fixed and arbitrary cardinalities. These expressions will be the summations of functions of $\chi_o(f)$ and $\chi_c(f)$ over all faces $f$ of the appropriate dual arrangement.

The implication of these formulae is that separable subsets may be counted without generating them explicitly. If $P$ is a set of $n$ labeled points in $E^d$, the number of separable subsets of $P$ may approach $2^n$. However, this number may be determined in time proportional to the number of faces of the dual half-space arrangement of $P$. Indeed, as we shall see in the next chapter, it is well known that the number of these faces is polynomial in $n$ with order dependent upon $d$.

## 5.2 Upper Bounds for Separable Components

Let $P$ be a set of distinct labeled points of $E^d$, such that no hyperplane contains every point of $P$. Consider the homogeneous hyperplane arrangement $\mathcal{A}(\rho_o(\vec{l}(P)))$ in $E^{d+1}$. Since the arrangement has a vertex located at the origin. each face of the arrangement has associated with it precisely one strict and one non-strict linearly separable component of $P$. The following lemma shows that only the cells of the homogeneous arrangement need be considered when looking for strict linearly separable components.

**Lemma 5.1** *Let $P$ be a set of distinct labeled points of $E^d$, such that no hyperplane contains every point of $P$. Let $f$ be a face of $\mathcal{A}(\rho_o(\vec{l}(P)))$ associated with the strict linearly separable component $C$ of $P$. Then there exists some cell $g$ of $\mathcal{A}(\rho_o(\vec{l}(P)))$, also associated with $C$, whose closure contains $f$.*

**Proof** Let $H$ be the set of hyperplanes of $\rho(\vec{l}(P))$ containing face $f$, and let $H_>$ be the set of half-spaces of $\rho_o(\vec{l}(P))$ bounded by the hyperplanes of $H$. Let $H_<$ be the set of open half-spaces complementary to those of $H_>$. Consider the common intersection of the half-spaces of the set $H_< \cap (\rho(\vec{l}(P)) \setminus H_>)$. This common intersection is a cell $g$ of $\mathcal{A}(\rho_o(\vec{l}(P)))$ contained in the same half-spaces of $\rho_o(\vec{l}(P))$ as $f$. Clearly, $f$ is contained in the closure of $g$. □

The proof of this lemma applies equally well to the case of strict spherical separation and the setting of $\mathcal{A}(\sigma_o(\vec{s}(P)))$. If no hypersphere or hyperplane of $E^d$ contains every point of $P$, then every face of the homogeneous arrangement $\mathcal{A}(\sigma_o(\vec{s}(P)))$ in $E^{d+2}$ corresponds to one strict and one non-strict spherically separable component of $P$. Accordingly, we state the following corollary:

**Corollary 5.2** *Let $P$ be a set of distinct labeled points of $E^d$, such that no hypersphere or hyperplane contains every point of $P$. Let $f$ be a face of $\mathcal{A}(\sigma_o(\vec{s}(P)))$ associated with the strict linearly separable component $C$ of $P$. Then there exists some cell $g$ of $\mathcal{A}(\sigma_o(\vec{s}(P)))$, also associated with $C$, whose closure contains $f$.*

Quite clearly, no two cells of $\mathcal{A}(\rho_o(\vec{l}(P)))$ correspond to the same strict linearly separable component, and no two cells of $\mathcal{A}(\rho_o(\vec{l}(P)))$ correspond to the same strict spherically separable component. For both the linear and spherical cases, one may conclude that each strictly separable component corresponds to a unique cell of the appropriate arrangement.

Another corollary of Lemma 5.1 deals with the non-strict components of labeled point sets:

**Corollary 5.3** *Let $P$ be a set of distinct labeled points of $E^d$, such that no hyperplane contains every point of $P$. Let $f$ be a face of $\mathcal{A}(\rho_o(\vec{l}(P)))$ associated with the non-strict linearly separable component $C$ of $P$. Then there exists some edge $g$ of $\mathcal{A}(\rho_o(\vec{l}(P)))$, also associated with $C$, contained in the closure of $f$.*

Naturally, Corollary 5.3 holds equally well for non-strict spherically separable components and the arrangement $\mathcal{A}(\sigma_o(\vec{s}(P)))$.

The number of strictly separable components of $P$ of size $k$ or less, whether linear or spherical, may be enumerated by counting the number of cells $f$ of the appropriate dual arrangement of $P$ having $\chi_o(f) \leq k$. By obtaining an upper bound on the number of such cells over all homogeneous arrangements in $E^m$, we now derive upper bounds on the number of strict components of cardinality $k$ or less, over all distributions of $n$ points in $d$ dimensions into the labeled sets $R$ and $G$. Once these bounds have been derived, we will see how they apply to the case of non-strict components.

Let $H$ be a set of $n$ hyperplanes in $E^m$ whose common intersection consists of the single point $\mathbf{0}$. Of course, this implies that $n \geq m$. For every hyperplane $h \in H$, let us define $h_>$ and $h_<$ to be the two open half-spaces of $E^m$ bounded by $h$. The set of half-spaces $\{h_> | \, h \in H\}$ shall be denoted by $H_>$, and the set of half-spaces $\{h_< | \, h \in H\}$ shall be denoted by $H_<$. Given a face $f$ of the homogeneous half-space arrangement $\mathcal{A}(H_>)$, let $\chi_o(f)$ be the number of half-spaces of $H_>$ containing face $f$, and let $\chi_c(f)$ be the number of half-spaces of $H_>$ whose closures contain $f$. Equivalently, $\chi_c(f)$ is the number of half-spaces of $H_<$ avoiding $f$. With these definitions, $\mathcal{A}(H_>)$ may be considered as a half-space arrangement in $E^m$.

Let $\mathcal{A}_n^m$ be the set of all such homogeneous half-space arrangements in $E^m$ having $n$ distinct hyperplanes, and containing the vertex located at the origin. Given some integer $k$ between 0 and $n$, inclusive, let us define $\mathcal{C}(m,k,n)$ to be the maximum number of cells $f$ of $\mathcal{A}(H_>)$ where $\chi_o(f) = k$, over all arrangements $\mathcal{A}(H_>)$ in $\mathcal{A}_n^m$. Also, given integers $k_1$ and $k_2$ such that $k_1 \leq k_2$, we let $\mathcal{C}(m, k_1 : k_2, n)$ be the maximum number of cells $f$ where $k_1 \leq \chi_o(f) \leq k_2$, over all arrangements in $\mathcal{A}_n^m$. We will adopt the convention that $\mathcal{C}(m,k,n) = 0$ for all integer values of $k$ less than 0 and greater than $n$. Also, we will say that $\mathcal{C}(m, k_1 : k_2, n) = \mathcal{C}(m, k_1 : n, n)$ if $k_2 > n$, and that $\mathcal{C}(m, k_1 : k_2, n) = \mathcal{C}(m, 0 : k_2, n)$ if $k_1 < 0$.

It should be noted that $\mathcal{C}(m,0,n) = \mathcal{C}(m,n,n) = 1$. Also, we have $\mathcal{C}(m,k,n) = \mathcal{C}(m, n-k, n)$ and $\mathcal{C}(m, k_1 : k_2, n) = \mathcal{C}(m, n-k_2 : n-k_1, n)$, as the following argument shows: Let $H$, $H_>$, and $H_<$ be defined as above, such that $\mathcal{A}(H_>)$ is an arrangement in $E^m$. If $f$ is a cell of $\mathcal{A}(H_>)$ such that $\chi_o(f) = k$, then the cell $-f$ radially opposite from $f$ about the origin has $\chi_o(-f) = n-k$. Hence the maximum number of cells $f$ of $\mathcal{A}(H_>)$ where $\chi_o(f) = k$, over all arrangements $\mathcal{A}(H_>)$ in $\mathcal{A}_n^m$, is the same as the maximum number of cells $g$ where $\chi_o(g) = n-k$.

A hyperplane $h$ of $H$, when intersected with the remaining $n-1$ hyperplanes of $H$, yields a set of $(m-1)$-dimensional flats $H' = (H \setminus \{h\}) \cap h$ in $h$. The half-spaces of $H_>$ and $H_<$, when intersected with $h$, are given by $H'_> = (H_> \setminus \{h_>\}) \cap h$ and $H'_< = (H_< \setminus \{h_<\}) \cap h$, respectively. Since every hyperplane of $H$ passes through the origin of $E^m$, these intersections of hyperplanes and half-spaces with $h$ are all non-empty. Also, the common intersection of the flats of $H'$ must be the vertex of

$\mathcal{A}(H_>)$ at the origin. For these reasons, the flats of $H'$, together with the sets $H'_>$ and $H'_<$, form an $(m-1)$-dimensional homogeneous half-space arrangement $\mathcal{A}(H'_>)$ in $h$.

Each cell $f'$ of $\mathcal{A}(H'_>)$ is an $(m-1)$-face of $\mathcal{A}(H_>)$, and as such is a facet contained in the closure of exactly two cells $f_1$ and $f_2$ of $\mathcal{A}(H_>)$. We will let $\mathcal{C}_h(k_1 : k_2)$ denote the number of cells $f$ of $\mathcal{A}(H_>)$, such that $k_1 \leq \chi_o(f) \leq k_2$ and $f$ has a bounding facet contained in $h$. With these definitions and observations, we are now able to prove the following lemma:

**Lemma 5.4** *Let $\mathcal{A}(H_>)$ be a homogeneous half-space arrangement of $\mathcal{A}_n^m$ as defined above. If $h$ is a hyperplane of $H$, then*

$$\mathcal{C}_h(0 : k) \leq \mathcal{C}(m-1, 0 : k, n-1),$$

*for $0 \leq k \leq n$ and $m \geq 2$.*

**Proof** Let $f'$ be a cell of $\mathcal{A}(H'_>)$, and let $f_1$ and $f_2$ be the two cells of $\mathcal{A}(H_>)$ having $f'$ as a common facet. Without loss of generality, we assume that $f_1$ is contained in $h_>$ and that $f_2$ is contained in $h_<$. Because $f_1$, $f_2$, and $f'$ are all contained in the same half-spaces of $H_> \setminus \{h_>\}$, we have $\chi_o(f_1) = \chi_o(f') + 1$ and $\chi_o(f_2) = \chi_o(f')$. Hence if $f$ is a cell of $\mathcal{A}(H_>)$ with facet $g$ contained in $h$, then $\chi_o(f) = j$ implies that $\chi_o(g)$ equals either $j$ or $j-1$. We therefore have

$$\mathcal{C}_h(0 : k) \leq \mathcal{C}(m-1, 0 : k, n-1)$$

as desired. □

By summing the quantity $\mathcal{C}_h(0 : k)$ over all hyperplanes $h$ in $\mathcal{A}(H_>)$, we arrive at the following lemma:

**Lemma 5.5** $\mathcal{C}(m, 0 : k, n) \leq \dfrac{n}{m}\mathcal{C}(m-1, 0 : k, n-1)$, *for $0 \leq k \leq n$ and $n \geq m \geq 2$.*

**Proof** Let $\mathcal{A}(H_>)$ be a homogeneous half-space arrangement of $\mathcal{A}_n^m$ as defined above. Let $n$ be the number of hyperplanes in $H$. Consider the sum of $\mathcal{C}_h(0 : k)$ over all hyperplanes $h$ in $H$. By Lemma 5.4,

$$\begin{aligned}\sum_{h \in H} \mathcal{C}_h(0 : k) &\leq \sum_{h \in H} \mathcal{C}(m-1, 0 : k, n-1) \\ &\leq n\,\mathcal{C}(m-1, 0 : k, n-1).\end{aligned}$$

But each cell $f$ having $\chi_o(f) \leq k$ is counted exactly as many times as it has bounding hyperplanes. Since cell $f$ has a vertex at the origin in its closure, it must have at least $m$ facets. Thus

$$\begin{aligned} \mathcal{C}(m, 0:k, n) &\leq \frac{1}{m}\sum_{h \in H} \mathcal{C}_h(0:k) \\ &\leq \frac{n}{m}\mathcal{C}(m-1, 0:k, n-1) \end{aligned}$$

as required. □

Lemma 5.5 gives us a recurrence relation that we will exploit in deriving our bound of Theorem 5.8. The boundary conditions for the recurrence arise out the examination of the two-dimensional situation. However, we first need the following result concerning the overlap of rays on the real line. Let $Q = \{q_1, q_2, \ldots, q_n\}$ be a sequence of distinct points on the real line, in increasing order, and let $Q^* = \{q_1^*, q_2^*, \ldots, q_n^*\}$ be a sequence of rays such that ray $q_i^*$ has endpoint $q_i$. Also, let $Q^+$ and $Q^-$ be the subsequences of positively-directed and negatively-directed rays of $Q^*$, respectively.

Let $I = \{I_0, I_1, \ldots, I_n\}$ be the sequence of open intervals where $I_0 = (-\infty, q_1)$, $I_n = (q_n, \infty)$, and $I_i = (q_i, q_{i+1})$ for all $i = 1, 2, \ldots, n-1$. For every interval $I_i \in I$, let us define $r^+(I_i)$ and $r^-(I_i)$ as the number of rays of $Q^+$ and $Q^-$, respectively, that contain $I_i$. Note that if $i < j$, then $r^+(I_i) \geq r^+(I_j)$ and $r^-(I_i) \leq r^-(I_j)$. Given some integer $k$ between 0 and $n$, inclusive, we wish to find the maximum number of intervals $I_i$ such that $r^+(I_i) + r^-(I_i) \leq k$, over all such sequences of rays $Q^*$. Denoting this number by $r(k, n)$, and using these definitions, we prove the following by induction on $k$:

**Lemma 5.6** $r(k,n) \leq 2k+1$, *for* $0 \leq k \leq n$.

**Proof** The lemma holds trivially for $k = 0$. Assume that the claim is true for all $k = 0, 1, \ldots, k'-1$. We will show that it must be true for $k = k'$.

Let $Q^*$ be a set of $n$ rays, as defined above, that realizes the maximum $r(k', n)$. Let $I_i$ be an interval of $I$ such that $r^+(I_i) + r^-(I_i) = k'$. If no such interval exists, then $r(k', n) \leq r(k'-1, n)$, and the lemma holds.

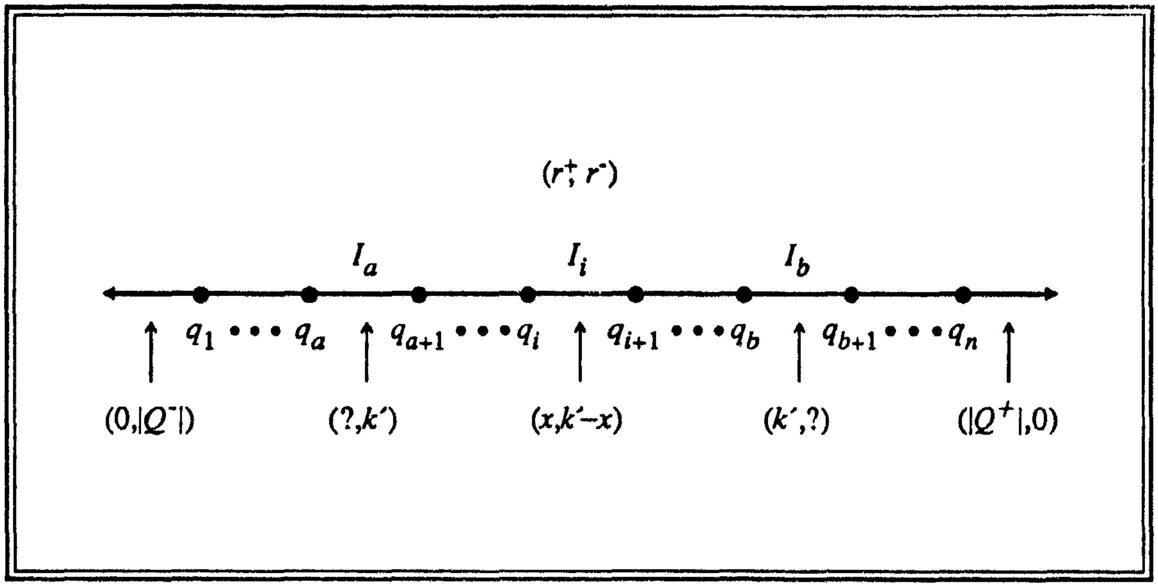


Figure 5.2: Construction for the proof of Lemma 5.6

Otherwise, let $I_a$ be the unique interval such that $r^-(I_a) = k'$, if it exists. If not, let $I_a = I_0$. Similarly, let $I_b$ be the interval such that $r^+(I_b) = k'$, if it exists. If not, let $I_b = I_n$. Note that $a \leq i$ and $b \geq i$. By construction, there can be no cells $I_j$ with $j < a$ or $j > b$ such that $r^+(I_j) + r^-(I_j) \leq k'$ (see Figure 5.2).

There can be no more than $k' - r^-(I_i)$ rays of $Q^-$ with endpoints between $I_a$ and $I_i$, and similarly no more than $k' - r^+(I_i)$ rays of $Q^+$ with endpoints between $I_i$ and $I_b$. Since $r^+(I_i) + r^-(I_i) = k'$, there are at most $r^+(I_i)$ endpoints of rays of $Q^+$ between $I_a$ and $I_i$, and at most $r^-(I_i)$ endpoints of rays of $Q^-$ between $I_i$ and $I_b$. Therefore the total number of points of $Q$ between $I_a$ and $I_b$ is no more than $2k'$, implying that $r(k', n) \leq 2k' + 1$. □

The bound of the previous lemma, while certainly correct, is not very meaningful for $k \geq n/2$, since there are only $n + 1$ intervals on the line. Despite this seeming deficiency, we use Lemma 5.6 to justify the boundary conditions for the recurrence of Lemma 5.5.

**Lemma 5.7** *$C(2,0 : k,n) \leq 4k$, for $1 \leq k \leq n$.*

**Proof** Let $H = \{h_1, h_2, \ldots, h_n\}$ be a set of $n$ lines passing through the origin of $E^2$, and let $H^* = \{h_1^*, h_2^*, \ldots, h_n^*\}$ be a set of open half-planes where $h_j^*$ is bounded by $h_j$

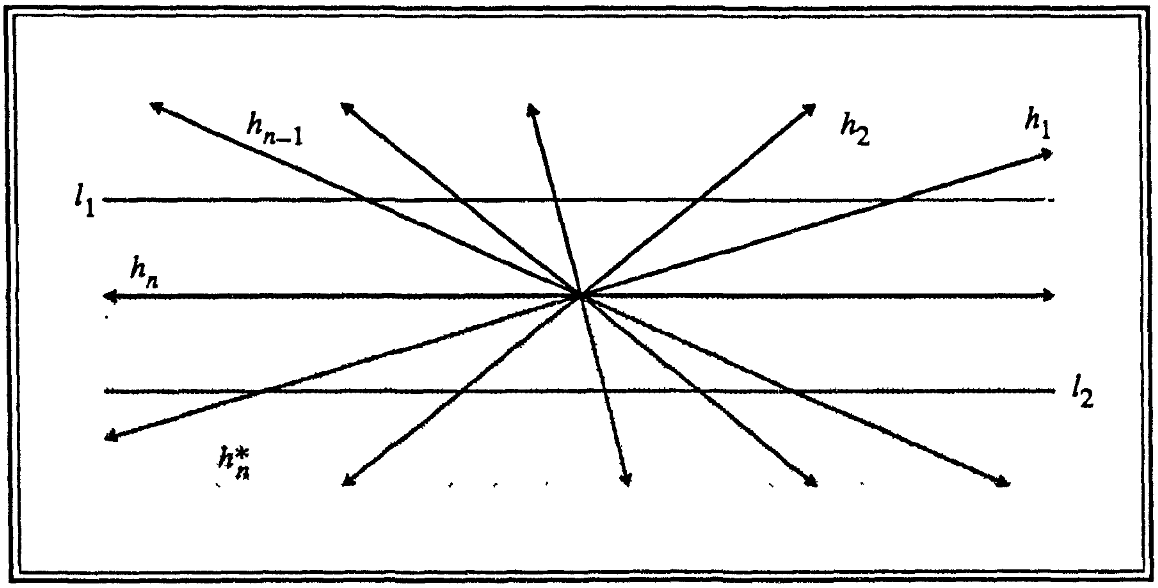


Figure 5.3: Construction for the proof of Lemma 5.7

for all $j = 1, 2, \ldots, n$. Without loss of generality, let us assume that $h_n$ coincides with the "horizontal" axis, and that $h_n^*$ is bounded by $h_n$ "from above". Then the lines of $H$, together with the half-planes of $H^*$, form a half-plane arrangement $\mathcal{A}(H^*)$ in $\mathcal{A}_n^2$.

Consider two lines $l_1$ and $l_2$ parallel to and distinct from $h_n$, such that $l_2$ is contained in $h_n^*$, but $l_1$ is not (see Figure 5.3). The intersection of $l_1$ and the half-planes of $H^* \setminus \{h_n^*\}$ form a collection of rays in the line $l_1$; we have a similar collection of rays in $l_2$. If $f_1$ is a cell of $\mathcal{A}(H^*)$ intersecting $l_1$, then the interval $f_1 \cap l_1$ in $l_1$ is contained in $\chi_o(f_1)$ rays of the collection. If $f_2$ is a cell intersecting $l_2$, then the interval $f_2 \cap l_2$ is contained in $\chi_o(f_2) - 1$ rays of the collection in $l_2$. Hence $\mathcal{C}(2, 0 : k, n) \leq r(k, n-1) + r(k-1, n-1) = 4k$. □

We may now state and prove the main theorem of this section.

**Theorem 5.8** $\mathcal{C}(m, 0 : k, n) \leq \dfrac{8k}{m(m-1)} \dbinom{n}{m-2}$, *for* $1 \leq k \leq n$ *and* $n \geq m \geq 2$.

**Proof** By induction on $m$. If $m = 2$, then

$$\mathcal{C}(2, 0 : k, n) \leq 4k = \frac{8k}{m(m-1)} \binom{n}{m-2}$$

by Lemma 5.7. Otherwise, if $m > 2$, assume that the theorem holds for dimensions less than $m$. By Lemma 5.5,

$$\begin{aligned}
\mathcal{C}(m,0:k,n) &\leq \frac{n}{m}\mathcal{C}(m-1,0:k,n-1) \\
&\leq \frac{n}{m}\left[\frac{8k}{(m-1)(m-2)}\binom{n-1}{m-3}\right] \\
&\leq \frac{8k}{m(m-1)}\binom{n}{m-2}
\end{aligned}$$

as required. □

The bound of Theorem 5.8 holds equally well for homogeneous half-space arrangements in $E^m$ that are not in $\mathcal{A}_n^m$; that is, for those whose hyperplanes do not intersect in a vertex of the arrangement. If $\mathcal{A}(H_>)$ is such an arrangement, with $H$ the set of hyperplanes of the arrangement, let $z$ be the common intersection of the hyperplanes of $H$, where $dim(z) > 0$. Any hyperplane $h$ of $H$ may be perturbed infinitesimally into the hyperplane $h'$ containing the origin, such that the intersection $z'$ of the hyperplanes of $(H \setminus \{h\}) \cup \{h'\}$ has $dim(z') = dim(z) - 1$. If the perturbation is sufficiently small, none of the cells of the arrangement are destroyed (although some new cells are created), and these cells are still contained in the same half-spaces (with $h'_>$ replacing the half-space $h_>$). These perturbations may be repeated until an arrangement of $\mathcal{A}_n^m$ is produced; this arrangement has at least as many cells $f$ with $\chi_o(f) \leq k$ as does $\mathcal{A}(H_>)$, and therefore the bound of Theorem 5.8 applies to $\mathcal{A}(H_>)$.

Given a labeled point set $P$, Theorem 5.8 may be directly applied to bound the number of its strict linearly or spherically separable components of a given size or smaller. We also use this result to bound the number of non-strict separable components, by noticing that if $C$ is a strict separable component of $P$, either linearly or spherically, then the set $P \setminus C$ is a non-strict separable component. Thus the bound on the number of strict separable components of size $n - k$ or more also bounds the number of non-strict separable components of size $k$ or less. Recalling that $\mathcal{C}(m, 0 : k, n) = \mathcal{C}(m, n - k : n, n)$, we now state the following corollaries:

**Corollary 5.9** *Let $P$ be a set of $n$ labeled distinct points of $E^d$, where $n \geq d+1$. Let $k$ be an integer between 1 and $n-1$, inclusive. Then the expression*

$$\frac{8k}{d(d+1)}\binom{n}{d-1}$$

*is an upper bound for*

1. *the number of strict linearly separable components of $P$ of size $\leq k$,*
2. *the number of strict linearly separable components of $P$ of size $\geq n-k$,*
3. *the number of non-strict linearly separable components of $P$ of size $\leq k$,*
4. *the number of non-strict linearly separable components of $P$ of size $\geq n-k$.*

**Corollary 5.10** *Let $P$ be a set of $n$ labeled distinct points of $E^d$, where $n \geq d+2$. Let $k$ be an integer between 1 and $n-1$, inclusive. Then the expression*

$$\frac{8k}{(d+1)(d+2)}\binom{n}{d}$$

*is an upper bound for*

1. *the number of strict spherically separable components of $P$ of size $\leq k$,*
2. *the number of strict spherically separable components of $P$ of size $\geq n-k$,*
3. *the number of non-strict spherically separable components of $P$ of size $\leq k$,*
4. *the number of non-strict spherically separable components of $P$ of size $\geq n-k$*

The asymptotic behaviour of these bounds, for fixed dimension $d$, is $O(kn^{d-1})$ in the linear case and $O(kn^d)$ in the spherical case. The best known result for $k$-sets in dimensions higher than three has been recently developed by Clarkson [Cla88], who gives asymptotic bounds of $\Theta(n^{\lfloor d/2 \rfloor}k^{\lceil d/2 \rceil})$ for the maximum number of $j$-sets, summed over all $j \leq k$, and taken over all sets of $n$ points in $E^d$. Here $d$ is taken to be fixed, and $n/k \to \infty$. Clearly, the lower bound $\Omega(n^{\lfloor d/2 \rfloor}k^{\lceil d/2 \rceil})$ applies to the number of strict linearly separable components of $P$ of size $\leq k$ or $\geq n-k$.

## 5.3 Counting the Number of Separable Subsets

We now shift our attention to the separable subsets of a set of labeled points. Let $P$ be a set of distinct labeled points of $E^d$. Corollary 4.3 relates the linear separability of a subset $Q$ of $P$ to the existence of a ray in the common intersection of the half-spaces of $\rho_o(\vec{l}(Q))$ or $\rho_c(\vec{l}(Q))$, depending upon whether the separability is strict or non-strict. Similarly, Corollary 4.7 relates the spherical separability of $Q$ to the existence of a ray in the common intersection of the half-spaces of $\sigma_o(\vec{s}(Q))$ or $\sigma_c(\vec{s}(Q))$. In this way, the strict and non-strict separable subsets of labeled point sets relate to the *intersecting subsets* of sets of open half-spaces and sets of closed half-spaces, respectively. In this section, we shall first develop expressions for the number of intersecting subsets in homogeneous half-space arrangements, and the number of intersecting subsets of fixed size. We shall then exploit these relationships by reinterpreting these results in the original setting of labeled point sets.

Let $\mathcal{A}(H_>)$ be a homogeneous half-space arrangement in $\mathcal{A}_n^m$, where $H_>$ consists of $n$ open half-spaces whose bounding hyperplanes pass through the origin. Let $H$ be the set of these bounding hyperplanes, and let $H_o$ and $H_c$ be the sets consisting of the open and closed half-spaces bounded by hyperplanes of $H$, respectively. Let $Q_o$ be a non-empty subset of $H_o$, and let $Q_c$ be a non-empty subset of $H_c$. We define the *open cone* $\Lambda(Q_o)$ of $Q_o$ to be the set of faces of $\mathcal{A}(H_>)$ contained in the region of intersection of the half-spaces of $Q_o$, and the *closed cone* $\Lambda(Q_c)$ of $Q_c$ to be that part of $\mathcal{A}(H_>)$ contained in the region of intersection of the half-spaces of $Q_c$ (see Figure 5.4). If a cone contains no ray of $\vec{E}^m$, then it will be called *empty*; otherwise, it will be said to be *non-empty*. Let $Q$ be the set of bounding hyperplanes of $Q_o$ and $Q_c$. The common intersection of the hyperplanes of $Q$ shall be called the *apex* of $\Lambda(Q_o)$ and $\Lambda(Q_c)$.

Given an arrangement $\mathcal{A}(H_>)$ in $\mathcal{A}_n^m$, our goal is to develop expressions for the number of subsets of $H_>$ and $H_\geq$ determining non-empty open and closed cones, respectively. We also wish to know the number of subsets of $H_>$ and $H_\geq$ of fixed cardinality $k$ that determine non-empty open and closed cones. In the first step in the derivation of these expressions, we will make use of Euler's relation for convex

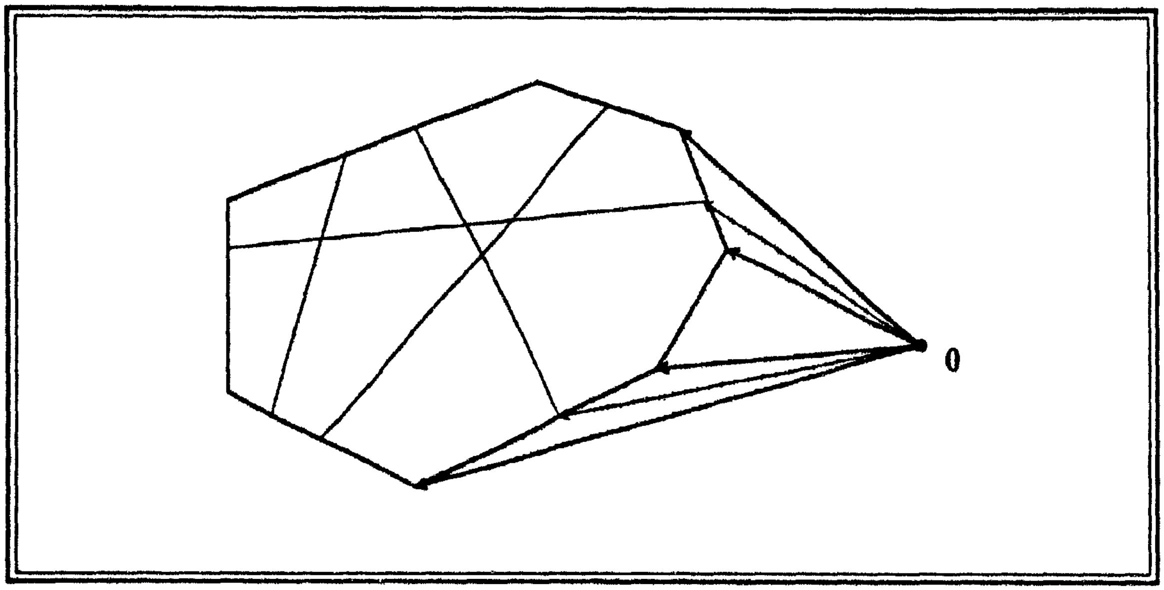


Figure 5.4: A closed cone in a homogeneous half-space arrangement

polytopes in $\boldsymbol{E}^d$ - an important result usually attributed to Ludwig Euler [Eul1,Eul2], but first proven in its higher dimensional form by Schläfli [Sch01]:

**Theorem 5.11 (Euler's Relation)** *Let $\pi$ be a convex polytope in $\boldsymbol{E}^d$. Let $\phi_i(\pi)$ be the set of faces of $\pi$ of dimension $i$. Then*

$$\sum_{i=0}^{d}(-1)^i \sum_{f\in\phi_i(\pi)} 1 = 1.$$

Here, as well in the rest of this chapter, we will adopt the convention that any summation over an empty range evaluates as zero. In particular, if $\phi_i(\pi) = \emptyset$, then

$$\sum_{f\in\phi_i(\pi)} 1 = 0.$$

In effect, Euler's relation states that if the number of faces of odd dimension is subtracted from the number of faces of even dimension, the difference is always 1. Another way of looking at Euler's relation is that the sum of $(-1)^{dim(f)}$ over all faces $f$ of $\pi$ is always 1. We choose to express Euler's relation in the manner of Theorem 5.11 because the range of dimensions of the faces of $\pi$ is given explicitly.

Consider now the intersection of a polytope $\pi$ in $E^d$ with some hyperplane $h$ Let $h_>$ and $h_<$ be the two open half-spaces bounded by $h$. If $f$ is a face of $\pi$, there are three possible ways in which $f$ may interact with $h$:

1. $f$ may be contained in $h$,
2. $f$ may avoid $h$ entirely, and
3. $f$ may be split by $h$ into the faces $f_>$, $f_=$, and $f_<$, contained in $h_>$, $h$, and $h_<$, respectively.

The main question here is how Euler's relation may be extend to account for the "splitting" of the polytope by $h$. In the first two cases mentioned, the face $f$ does not change. In the third case, $f$ is replaced by three new faces, two of which ($f_>$ and $f_<$) have the same dimension as $f$; the other ($f_=$) having dimension one less. Thus we have

$$\begin{aligned}
&(-1)^{dim(f_>)} + (-1)^{dim(f_=)} + (-1)^{dim(f_<)} \\
&\quad = (-1)^{dim(f)} + (-1)^{dim(f)-1} + (-1)^{dim(f)} \\
&\quad = (-1)^{dim(f)}.
\end{aligned}$$

If the summation of Euler's relation is applied to the new faces of the "split" polytope, the result is the same. Indeed, if new hyperplanes are successively introduced, the same argument shows that Euler's relation still holds. A (closed) polytope $\pi$, together with a set of hyperplanes $H$, shall be said to form a *sliced* polytope $\pi(H)$. With this definition, we state the following variant of Euler's relation:

**Lemma 5.12** *Let $\pi(H)$ be a closed sliced polytope in $E^d$, and let $\phi_i(\pi(H))$ be the set of faces of $\pi(H)$ of dimension $i$. Then*

$$\sum_{i=0}^{d}(-1)^i \sum_{f\in\phi_i(\pi(H))} 1 = 1.$$

The technique of polytope slicing is not new; in fact, the proof of Euler's relation due to Nef [Nef81,Nef84] relies on it.

Later in this section, we will need a result similar to Lemma 5.12 for *open* sliced polytopes. As a starting point, consider this contrived but not entirely pointless statement of Euler's relation for open polytopes.

**Observation 5.13** *Let $\pi$ be the non-empty interior of a convex polytope in $E^d$. Let $\phi_i(\pi)$ be the set of faces of $\pi$ of dimension $i$. Then*

$$\sum_{i=0}^{d}(-1)^{i+d}\sum_{f\in\phi_i(\pi)}1=1.$$

The only face of $\pi$ is the $d$-dimensional face $\pi$ itself, and so of course the observation is true. However, the same splitting argument used in the case of closed polytopes is equally effective in this setting. In fact, it is effective even when $\pi$ is taken to be an unbounded polyhedral set. Observation 5.13 thus gives way to the more useful Lemma 5.14:

**Lemma 5.14** *Let $\pi(H)$ be a non-empty open sliced polyhedral set in $E^d$, and let $\phi_i(\pi(H))$ be the set of faces of $\pi(H)$ of dimension $i$. Then*

$$\sum_{i=0}^{d}(-1)^{i+d}\sum_{f\in\phi_i(\pi(H))}1=1.$$

Lemma 5.14 and Lemma 5.12 allow us to extend Euler's relation to open and closed cones, respectively. For the remainder of the chapter, we will use the following notation to refer to the faces of an open or closed cone. If $\Lambda$ is a cone located in the arrangement $\mathcal{A}(H_>)$ of $\mathcal{A}_n^m$, we shall define $\phi_i(\Lambda)$ to be the set of faces of $\Lambda$ of dimension $i$. With this notation, we state Euler's relation for closed cones in homogeneous half-space arrangements:

**Lemma 5.15** *Let $\mathcal{A}(H_>)$ be a homogeneous half-space arrangement of $\mathcal{A}_n^m$, where $m\geq 2$, and let $Q_c$ be a non-empty subset of the set of closed half-spaces $H_c$. If the closed cone $\Lambda(Q_c)$ is non-empty, then the faces of $\Lambda(Q_c)$ satisfy*

$$\sum_{i=1}^{m}(-1)^{i-1}\sum_{f\in\phi_i(\Lambda(Q_c))}1=1.$$

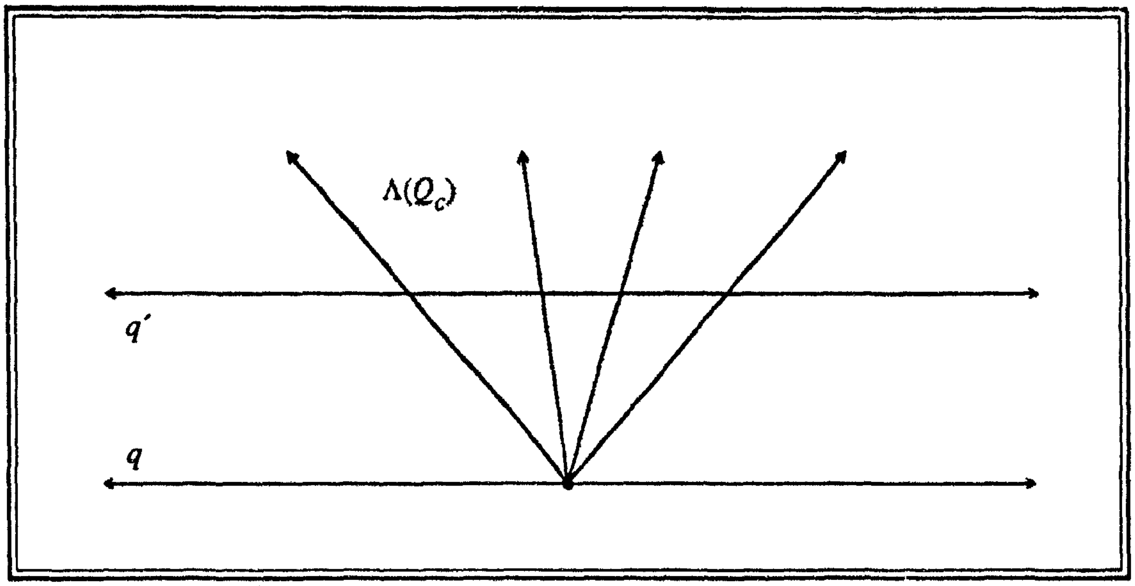


Figure 5.5: Construction for proof of Lemma 5.15

**Proof** By induction on the dimensionality of the apex of $\Lambda(Q_c)$. First, let us assume that the apex of $\Lambda(Q_c)$ has dimension zero; that is, the apex is the vertex of $\mathcal{A}(H_>)$ at the origin. Then there must be a hyperplane $q$ passing through the apex of $\Lambda(Q_c)$ and avoiding the remaining faces of the non-empty cone $\Lambda(Q_c)$ (see Figure 5.5). Also, there must exist a translate $q'$ of $q$ intersecting every ray contained in the faces of $\Lambda(Q_c)$. The intersection of $q'$ with the closed cone $\Lambda(Q_c)$ yields a closed sliced polytope of dimension $m-1$. Every face $f \in \Lambda(Q_c)$ of dimension greater than zero intersects $q'$ in the face $f'$ of the sliced polytope. Noting that $dim(f) = dim(f') + 1$, Lemma 5.12 may be applied to obtain the result for this case.

Now let us assume that the lemma holds for cones whose apices are of dimension less than $j$, where $j \geq 1$. We shall show that the lemma is true for cones with apices of dimension $j$. Let the apex of $\Lambda(Q_c)$ be a $j$-flat passing through the origin. Since $\mathcal{A}(H_>)$ is an arrangement in $\mathcal{A}_n^m$, a vertex of $\mathcal{A}(H_>)$ is situated at the origin. Hence, if $H$ is the set of bounding hyperplanes of the half-spaces of $H_>$, there must exist some $h \in H$ that does not contain the apex of $\Lambda(Q_c)$. Let $h_\geq$ and $h_\leq$ be the closed half-spaces bounded by $h$. The sets $Q_1 = Q_c \cup \{h_\geq\}$, $Q_2 = Q_c \cup \{h_\leq\}$, and $Q_3 = Q_c \cup \{h_\geq, h_\leq\}$ all determine closed cones of $\mathcal{A}(H_>)$ having apices of dimension

$j-1$.

The faces of $\Lambda(Q_c)$ avoiding $h$ lie in $Q_1$ or $Q_2$, but not both. The faces contained in $h$ lie in $Q_1$, $Q_2$, and $Q_3$. Hence

$$\begin{aligned}\sum_{i=1}^{m}(-1)^{i-1}\sum_{f\in\phi_i(\Lambda(Q_c))}1 &= \sum_{i=1}^{m}(-1)^{i-1}\sum_{f\in\phi_i(\Lambda(Q_1))}1\\ &\quad+\sum_{i=1}^{m}(-1)^{i-1}\sum_{f\in\phi_i(\Lambda(Q_2))}1-\sum_{i=1}^{m}(-1)^{i-1}\sum_{f\in\phi_i(\Lambda(Q_3))}1\\ &= 1+1-1=1,\end{aligned}$$

by the induction hypothesis. □

An inductive proof was used here because no hyperplane may intersect all faces $f$ of a closed cone $\Lambda$ such that $dim(f)>0$, if the apex of $\Lambda$ is of dimension greater than zero. The only hyperplane that could possibly intersect all 1-faces of the apex is one passing through the origin – but in addition to other shortcomings, this hyperplane would avoid all faces of $\Lambda$ not in the apex. For the case of an open cone, there always exists a hyperplane intersecting all faces, and thus we avoid having to resort to induction.

**Lemma 5.16** *Let $\mathcal{A}(H_>)$ be a homogeneous half-space arrangement of $\mathcal{A}_n^m$, where $m\geq 2$, and let $Q_o$ be a non-empty subset of the set of open half-spaces $H_o$. If the open cone $\Lambda(Q_o)$ is non-empty, then the faces of $\Lambda(Q_o)$ satisfy*

$$\sum_{i=1}^{m}(-1)^{i+m}\sum_{f\in\phi_i(\Lambda(Q_o))}1=1.$$

**Proof** Let $q$ be a hyperplane passing through the apex of $\Lambda(Q_o)$ and avoiding the remaining faces of the non-empty cone $\Lambda(Q_o)$. Because cone $\Lambda(Q_o)$ is open and convex, such a hyperplane may always be exhibited. There must exist a translate $q'$ of $q$ intersecting every ray contained in the faces of $\Lambda(Q_o)$. The intersection of $q'$ with the open cone $\Lambda(Q_o)$ yields not necessarily an open sliced polytope, but an open sliced polyhedral set of dimension $m-1$ in $q'$. Every face $f\in\Lambda(Q_c)$ intersects $q'$ in the face $f'$ of the sliced polyhedral set. Noting that $dim(f)=dim(f')+1$, we have

$$(-1)^{dim(f)+m}=(-1)^{dim(f')+m+1}=(-1)^{dim(f')+m-1}.$$

By applying Lemma 5.14 to the faces of the sliced polyhedral set, the result follows. □

Lemma 5.15 and Lemma 5.16 allow us to prove the main theorem of this section. For this proof, we will follow the convention that the combination $\binom{a}{b}$ equals 0, if $a$ and $b$ are integers such that $b > a$. If $H^*$ is a set of half-spaces, either all closed or all open, let $\Gamma(H^*)$ be the family of subsets of $H^*$ whose associated cones are non-empty, and let $|\Gamma(H^*)|$ be the cardinality of $\Gamma(H^*)$. Also, for integer $k$ greater than 1, let $\Gamma_k(H^*)$ be the family of subsets of $H^*$ of fixed cardinality $k$ whose associated cones are non-empty, and let $|\Gamma_k(H^*)|$ be the cardinality of $\Gamma_k(H^*)$. Lastly, if $\mathcal{A}(H_>)$ is a homogeneous half-space arrangement, we will denote by $\mathcal{A}_i(H_>)$ the set of faces of $\mathcal{A}(H_>)$ having dimensionality $i$.

**Theorem 5.17** *Let $H_>$ be a collection of half-spaces in $\mathbf{E}^m$ forming the homogeneous arrangement $\mathcal{A}(H_>)$ of $\mathcal{A}^m_n$, for $m \geq 2$. Let $H_\geq$ be the set of half-spaces generated by taking the closures of the half-spaces of $H_>$. Then*

$$\begin{aligned}
\text{a)}\quad |\Gamma_k(H_\geq)| &= \sum_{i=1}^{m}(-1)^{i-1}\sum_{f\in\mathcal{A}_i(H_>)}\binom{\chi_c(f)}{k},\\
\text{b)}\quad |\Gamma_k(H_>)| &= \sum_{i=1}^{m}(-1)^{i+m}\sum_{f\in\mathcal{A}_i(H_>)}\binom{\chi_o(f)}{k},\\
\text{c)}\quad |\Gamma(H_\geq)| &= \sum_{i=1}^{m}(-1)^{i-1}\sum_{f\in\mathcal{A}_i(H_>)}(2^{\chi_c(f)}-1), \text{ and}\\
\text{d)}\quad |\Gamma(H_>)| &= \sum_{i=1}^{m}(-1)^{i+m}\sum_{f\in\mathcal{A}_i(H_>)}(2^{\chi_o(f)}-1).
\end{aligned}$$

**Proof** We shall prove only the first and last claims; the proofs of the other two follow from the same arguments.

Given some face $f$ in $\mathcal{A}_i(H_>)$, the number of half-spaces of $H_\geq$ containing $f$ is $\chi_c(f)$, and the number of subsets of $H_\geq$ of cardinality $k$ containing $f$ is

$$\sum_{Q_\geq\in\Gamma_k(H_\geq)|\, f\in\phi_i(\Lambda(Q_\geq))} 1 = \binom{\chi_c(f)}{k}.$$

This summation may be thought of as a contribution of 1 from every non-empty closed cone of $\Gamma_k(H_{\geq})$ containing $f$. Hence we have

$$\begin{aligned}
\sum_{i=1}^{m}(-1)^{i-1}\sum_{f\in\mathcal{A}_i(H_>)}\binom{\chi_c(f)}{k} &= \sum_{i=1}^{m}(-1)^{i-1}\left(\sum_{f\in\mathcal{A}_i(H_>)}\left(\sum_{Q_{\geq}\in\Gamma_k(H_{\geq})\mid f\in\phi_i(\Lambda(Q_{\geq}))}1\right)\right)\\
&= \sum_{i=1}^{m}(-1)^{i-1}\left(\sum_{Q_{\geq}\in\Gamma_k(H_{\geq})}\left(\sum_{f\in\phi_i(\Lambda(Q_{\geq}))}1\right)\right)\\
&= \sum_{Q_{\geq}\in\Gamma_k(H_{\geq})}\left(\sum_{i=1}^{m}(-1)^{i-1}\left(\sum_{f\in\phi_i(\Lambda(Q_{\geq}))}1\right)\right)\\
&= \sum_{Q_{\geq}\in\Gamma_k(H_{\geq})}1
\end{aligned}$$

by Lemma 5.15. Thus the first claim holds.

For the last claim, the argument is very similar to that for the first. Given some face $f$ in $\mathcal{A}_i(H_>)$, the number of half-spaces of $H_>$ containing $f$ is $\chi_o(f)$, and the number of non-empty subsets of $H_>$ of cardinality $k$ containing $f$ is

$$\sum_{Q_>\in\Gamma(H_>)\mid f\in\phi_i(\Lambda(Q_>))}1 = 2^{\chi_o(f)}-1.$$

Hence we have

$$\begin{aligned}
\sum_{i=1}^{m}(-1)^{i+m}\sum_{f\in\mathcal{A}_i(H_>)}(2^{\chi_o(f)}-1) &= \sum_{i=1}^{m}(-1)^{i+m}\left(\sum_{f\in\mathcal{A}_i(H_>)}\left(\sum_{Q_>\in\Gamma(H_>)\mid f\in\phi_i(\Lambda(Q_>))}1\right)\right)\\
&= \sum_{i=1}^{m}(-1)^{i+m}\left(\sum_{Q_>\in\Gamma(H_>)}\left(\sum_{f\in\phi_i(\Lambda(Q_>))}1\right)\right)\\
&= \sum_{Q_>\in\Gamma(H_>)}\left(\sum_{i=1}^{m}(-1)^{i+m}\left(\sum_{f\in\phi_i(\Lambda(Q_>))}1\right)\right)\\
&= \sum_{Q_>\in\Gamma(H_>)}1
\end{aligned}$$

by Lemma 5.16. Thus the last claim also holds. □

Interpreted in the original settings of linear and spherical separation of point sets in $E^d$, these expressions count the number of separable subsets of a given labeled point set, whether strict or non-strict, linear or spherical, of fixed cardinality or of all cardinalities. The next two corollaries summarize these results.

**Corollary 5.18** *Let $P$ be a set of $n$ distinct labeled points of $\mathbf{E}^d$, such that no hyper plane contains every point of $P$. Let $k$ be a positive integer.*

1. *The number of non-empty non-strict linearly separable subsets of $P$ of fixed cardinality $k$ is given by*

$$\sum_{i=1}^{d+1}(-1)^{i-1}\sum_{f\in\mathcal{A}_i(\rho_o(\vec{l}(P)))}\binom{\chi_c(f)}{k}.$$

2. *The number of non-empty strict linearly separable subsets of $P$ of fixed cardinality $k$ is given by*

$$\sum_{i=1}^{d+1}(-1)^{i+d+1}\sum_{f\in\mathcal{A}_i(\rho_o(\vec{l}(P)))}\binom{\chi_o(f)}{k}.$$

3. *The number of non-empty non-strict linearly separable subsets of $P$ is given by*

$$\sum_{i=1}^{d+1}(-1)^{i-1}\sum_{f\in\mathcal{A}_i(\rho_o(\vec{l}(P)))}(2^{\chi_c(f)}-1).$$

4. *The number of non-empty strict linearly separable subsets of $P$ is given by*

$$\sum_{i=1}^{d+1}(-1)^{i+d+1}\sum_{f\in\mathcal{A}_i(\rho_o(\vec{l}(P)))}(2^{\chi_o(f)}-1).$$

**Corollary 5.19** *Let $P$ be a set of $n$ distinct labeled points of $\mathbf{E}^d$, such that no hyperplane or hypersphere contains every point of $P$. Let $k$ be a positive integer.*

1. *The number of non-empty non-strict spherically separable subsets of $P$ of fixed cardinality $k$ is given by*

$$\sum_{i=1}^{d+2}(-1)^{i-1}\sum_{f\in\mathcal{A}_i(\sigma_o(\vec{s}(P)))}\binom{\chi_c(f)}{k}.$$

2. *The number of non-empty strict spherically separable subsets of $P$ of fixed cardinality $k$ is given by*

$$\sum_{i=1}^{d+2}(-1)^{i+d}\sum_{f\in\mathcal{A}_i(\sigma_o(\vec{s}(P)))}\binom{\chi_o(f)}{k}.$$

3. *The number of non-empty non-strict spherically separable subsets of $P$ is given by*

$$\sum_{i=1}^{d+2}(-1)^{i-1}\sum_{f\in\mathcal{A}_i(\sigma_o(\vec{s}(P)))}(2^{\chi_c(f)}-1).$$

4. *The number of non-empty strict spherically separable subsets of $P$ is given by*

$$\sum_{i=1}^{d+2}(-1)^{i+d}\sum_{f\in\mathcal{A}_i(\sigma_o(\vec{s}(P)))}(2^{\chi_o(f)}-1).$$

If the points of $P$ are contained in a hyperplane $h$, then the expressions of Corollary 5.18 are not valid. The proof of Theorem 5.17 relies heavily upon the existence of a vertex at the origin in the dual homogeneous half-space arrangement. However, all subsets of $P$ are non-strictly separable by the hyperplane $h$ itself. To count strictly separable subsets, it suffices to consider the problem in the $(d-1)$-dimensional setting of $h$, by converting the coordinates of $P$ in $E^d$ into suitable coordinates in $E^{d-1}$. Alternatively, one could handle this in the dual space itself, by cutting every face of $\mathcal{A}(\rho_o(\vec{l}(P)))$ with a hyperplane $h'$, and applying the counting methods recursively. Of course, this second strategy lends itself equally well to the spherical case.

# Chapter 6

# Weak Separation Algorithms

## 6.1 Introduction

Up to this point in the thesis, we have been concerned with the combinatorial aspects of separation. In this chapter, and in the chapter to follow, we will investigate the algorithmic aspects of separation. Most of the algorithms of this chapter will be based upon the transformations of Chapter 4 which map sets of labeled points in $E^d$ into homogeneous half-space arrangements in $E^{d+1}$ and $E^{d+2}$.

Several algorithms already exist for finding strong linear and spherical separators of labeled point sets. It has been known for some time that the problem of finding a strong linear separator may be expressed as a linear programming problem. With the techniques due to Megiddo [Meg84], and later refined by Dyer [Dye86] and Clarkson [Cla86], such problems may be solved in time and space linear in the number of points, assuming that the dimension of the problem is fixed. The problem of finding strong separators with certain desirable qualities will be discussed in Chapter 7.

For strong spherical separation in $E^2$, O'Rourke, Kosaraju, and Megiddo [OKM86] have shown that the problem of finding a smallest separating circle in two dimensions may be performed in linear time, also using the techniques of [Meg84]. They also show that a largest such circle may be found in optimal $O(n \log n)$ worst-case time, where $n$ is the number of labeled points to be separated. Using nearest-point and farthest-point Voronoi diagrams, Bhattacharya [Bha88] presented an algorithm to determine the set of all circular separators in $O(n \log n)$ worst-case time.

In this chapter, we will focus on the problem of finding weak separators of various types. Initially, we will examine the hyperplane construction algorithm due to Edelsbrunner, O'Rourke, and Seidel [EOS86], including the data structures appropriate for the storage of such an arrangement. Secondly, in Section 6.3, modifications to this algorithm will be outlined that allow the degenerate character of homogeneous arrangements to be exploited: it will be shown that homogeneous arrangements in $\boldsymbol{E}^{m+1}$ may be constructed using time and storage of the same order as in the construction of non-homogeneous arrangements in $\boldsymbol{E}^m$. In Section 6.4, an algorithm is presented that enables the determination of weak linear and spherical separators in arbitrary dimensions. Finally, in Section 6.5, the topological sweep approach of Edelsbrunner and Guibas [EG86] will be used to reduce the storage required for some of these separation problems.

Throughout this chapter, we will assume that the primitive comparative and arithmetic operations (addition, subtraction, multiplication, and division) may be performed in unit time. Also, we assume that the storage required by a real number or integer is unit space. Thus a point in $\boldsymbol{E}^d$ requires $O(d)$ storage space. However, in the discussions of the asymptotic complexities of these algorithms, we will consider the dimension of the problem to be fixed.

## 6.2 Constructing Homogeneous Hyperplane Arrangements

To represent an arrangement of hyperplanes in storage, a data structure known as an *incidence graph* is used. This representation technique was first developed by Grünbaum [Grü67] for convex polytopes.

Let $\mathcal{A}(H)$ be an arrangement of hyperplanes in $\boldsymbol{E}^m$, not necessarily homogeneous. For convenience, we define the two *improper faces* of $\mathcal{A}(H)$ as being the $(-1)$-face $\emptyset$ and the $(m+1)$-face $\mathcal{A}(H)$. We say that the $(-1)$-face $\emptyset$ is incident upon every vertex of $\mathcal{A}(H)$, and that the $(m+1)$-face $\mathcal{A}(H)$ is incident upon every cell of the

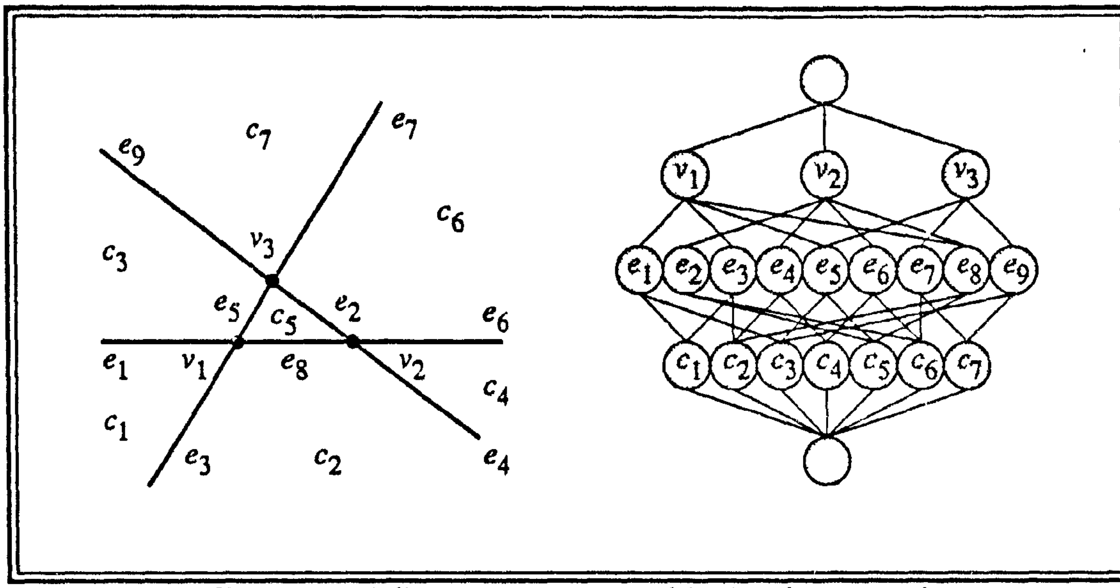


Figure 6.1: A line arrangement and its incidence graph

arrangement. The usual 0- to $m$-dimensional faces will be called *proper*. The incidence graph of $\mathcal{A}(H)$ shall be denoted by $\mathcal{I}(H)$, and is defined as follows: for each proper and improper face of $\mathcal{A}(H)$, there exists a node of $\mathcal{I}(H)$. If faces $f_1$ and $f_2$ are incident upon each other, then their nodes in $\mathcal{I}(H)$ are adjacent. An example of a line arrangement and its corresponding incidence graph is shown in Figure 6.1. In discussing the incidence graph, we will often refer to a given node by the face it represents.

In the implementation of the incidence graph, each node is represented by a record that contains a description of the face to which it corresponds, additional space for such accounting purposes as marking of faces and so forth, and two lists of pointers to other node records. One of these lists is devoted to the subfaces of the current face, and the other is devoted to its superfaces. Each of the pointers may also have additional space associated with them, for the labeling of incidences between faces, or for other purposes. The description of a face usually consists of some parameterization of the affine hull of the face, and the coordinates of some point belonging to the face. There are many ways of choosing such a point; for most applications, the actual choice itself is irrelevant. The exact allocation of additional space depends heavily upon the

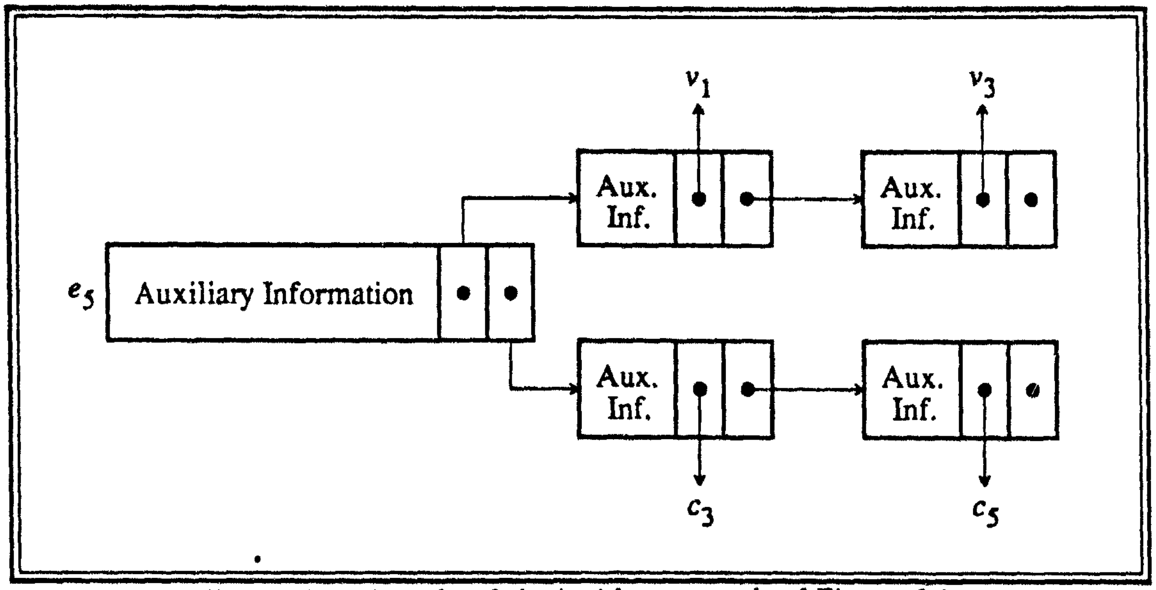


Figure 6.2: A node of the incidence graph of Figure 6.1

algorithms using this data structure. The two improper nodes allow access to the structure: the vertices of $\mathcal{A}(H)$ may be accessed using the list of superfaces in the node corresponding to improper face $\emptyset$, and the cells of $\mathcal{A}(H)$ may be accessed using the list of subfaces in the node corresponding to the improper face $\mathcal{A}(H)$. Figure 6.2 contains a description of the internal layout of a node of the incidence graph of Figure 6.1.

The size of the incidence graph $\mathcal{I}(H)$ is strictly proportional to the number of faces and incidences between faces of the arrangement $\mathcal{A}(H)$. We define $\boldsymbol{f}_k(H)$ to be the number of $k$-faces of $\mathcal{A}(H)$, for $0 \leq k \leq m$, and define $\boldsymbol{i}_k(H)$ to be the number of incidences between $k$-faces and $(k+1)$-faces of $\mathcal{A}(H)$, for $0 \leq k \leq m-1$. We also define $\boldsymbol{f}_{k,m}(n)$ and $\boldsymbol{i}_{k,m}(n)$ to be the maxima of $\boldsymbol{f}_k(H)$ and $\boldsymbol{i}_k(H)$ respectively, taken over all sets of $n$ hyperplanes $H$ in $\boldsymbol{E}^m$. Using the well-known results due to Buck [Buc43] for $\boldsymbol{f}_{k,m}(n)$ and $\boldsymbol{i}_{k,m}(n)$, we can place a bound on the size of $\mathcal{I}(H)$ in terms of the cardinality and dimensionality of $H$.

**Theorem 6.1 (Buck)** *Given $n \geq 1$ and $m \geq 1$, then*

$$\boldsymbol{f}_{k,m}(n) \leq \sum_{i=0}^{k} \binom{m-i}{k-i}\binom{n}{m-i}, \text{ for } 0 \leq k \leq m, \text{ and}$$

$$\boldsymbol{i}_{k,m}(n) \quad \leq \quad 2(m-k)\boldsymbol{f}_{k,m}(n), \text{ for } 0 \leq k \leq m-1.$$

*In addition, if $H$ is a set of $n$ hyperplanes in $\boldsymbol{E}^m$, then $\boldsymbol{f}_k(H) = \boldsymbol{f}_{k,m}(n)$ and $\boldsymbol{i}_k(H) = \boldsymbol{i}_{k,m}(n)$ if and only if $\mathcal{A}(H)$ is simple.*

Theorem 6.1 readily implies that the worst-case amount of space required to store the incidence graph of a set of $n$ hyperplanes in $m$ dimensions is in $\Theta(n^m)$.

Given a set of hyperplanes in $\boldsymbol{E}^m$, the incidence graph $\mathcal{I}(H)$ may be constructed in $O(n^d)$ time using the incremental algorithm of Edelsbrunner, O'Rourke, and Seidel [EOS86]. However, due to the degenerate structure of homogeneous hyperplane arrangements, with some modifications, their algorithm may be used to construct homogeneous arrangements in $\boldsymbol{E}^m$ using only $O(n^{m-1})$ time and space. Before justifying this claim, we must first understand some of the workings of their algorithm. In the description that follows, we will not differentiate between the faces of the arrangement $\mathcal{A}(H)$, and the nodes of the graph $\mathcal{I}(H)$.

Initially, a subset $H'$ of $H$ of size $m$ is obtained whose hyperplanes intersect in a common vertex of $\mathcal{A}(H)$. If no such subset exists, then the normal vectors of the hyperplanes of $H$ are contained in a single $(m-1)$-flat, and thus the arrangement may instead be constructed in this flat. Otherwise, the arrangement $\mathcal{A}(H')$ is constructed using some *ad hoc* method.

The main step of the algorithm consists of introducing the hyperplanes of $H \setminus H'$, one by one, into the growing arrangement. The order in which these hyperplanes are added is irrelevant. Let us assume that hyperplane $h$ is being added to the arrangement $\mathcal{A}(H')$. First, an edge $e_0$ of $\mathcal{A}(H')$ is found whose closure $cl(e_0)$ intersects $h$. Next, starting from $e_0$, all faces of $\mathcal{A}(H')$ whose closure intersects $h$ are marked. Finally, each marked face is updated. Those faces intersected by $h$ are split into new faces if necessary. When the last hyperplane of $h$ has been added, the algorithm terminates.

The time required to insert one hyperplane $h$ into an existing arrangement $\mathcal{A}(H')$ is bounded by the number of faces of $\mathcal{A}(H' \cup \{h\})$ contained in the closures of cells bounded by $h$. This subset of the faces of $\mathcal{A}(H' \cup \{h\})$ is known as the *zone* of the arrangement defined by $h$ (see Figure 6.3). In the analysis of their algorithm,

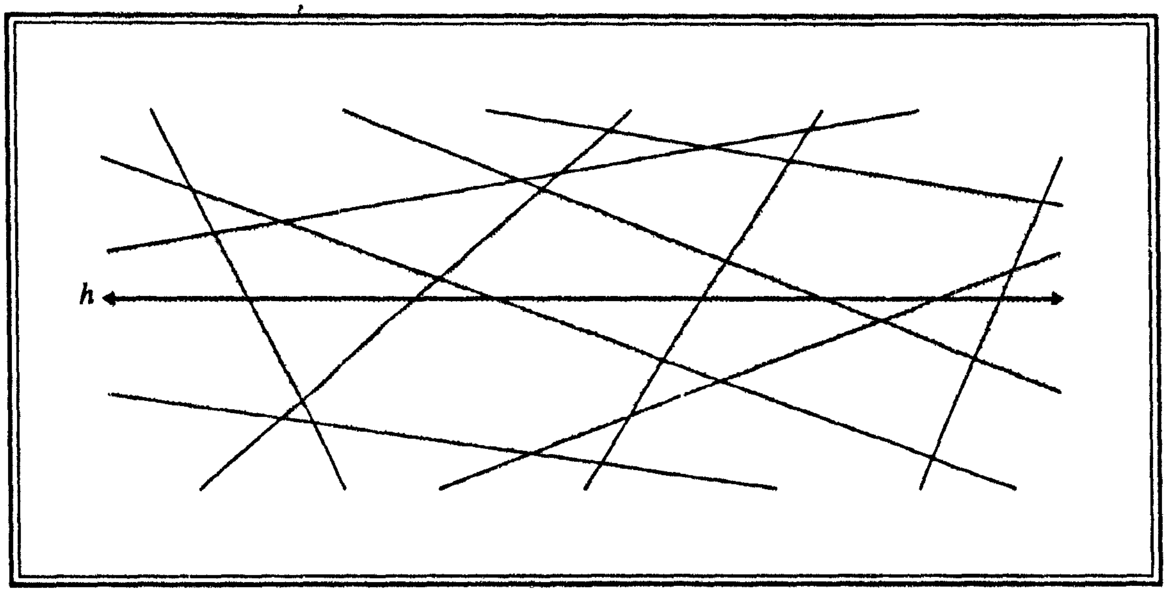


Figure 6.3: A zone in an arrangement of lines

Edelsbrunner, O'Rourke, and Seidel show that the worst-case complexity of a zone in an arrangement of hyperplanes in $\boldsymbol{E}^m$ is in $\Theta(n^{m-1})$. Thus the time required to perform all the incrementations of the arrangement is $O(n^m)$. The worst-case time complexity of their algorithm, being dominated by the incremental step, is then the same as the worst-case space complexity: $O(n^m)$.

The space complexity of a homogeneous arrangement in $\mathcal{A}_n^m$ certainly does not attain the worst case for arrangements in $\boldsymbol{E}^m$. Since every 1-face of a homogeneous arrangement is a ray of $\vec{E}^m$, there must exist some hyperplane $h'$ passing through the origin that intersects no 1-face. Since the closures of the 1-faces contain the origin, given any pair of hyperplanes $h_a$ and $h_b$ parallel to $h'$ and on opposite sides of $h'$, every 1-face must intersect either $h_a$ or $h_b$ in a single point. Thus $h_a$ and $h_b$ together intersect every face of the arrangement other than the vertex at the origin. Since the intersections of the homogeneous arrangement with $h_a$ and $h_b$ form $(m-1)$-dimensional arrangements in $h_a$ and $h_b$, the worst-case size of the incidence graph of a homogeneous arrangement is $O(n^{m-1})$.

The highly degenerate structure of a homogeneous arrangement suggests that the construction time may be reduced by limiting the number of faces of the homogeneous arrangement that are marked unnecessarily. The rest of this section shall be devoted to this topic. Due to the complexity of the marking and update technique of the standard construction algorithm, in the following discussion we will concern ourselves only with the problem of which faces need to be marked, and how to visit these faces. The nature of the mark values assigned to each face will not change, nor will the method of updating marked faces.

In the standard algorithm, the faces that are marked initially are the vertices, edges, and 2-faces of $\mathcal{A}(H')$ whose closures intersect the new hyperplane $h$, using a breadth-first search strategy starting from an initial edge $e_0$ whose closure $cl(e_0)$ intersects $h$. The feasibility of this step is affirmed by observing that the intersection of $h$ with the union of the vertices, edges and 2-faces of $\mathcal{A}(H')$ is connected. The incidence graph of $\mathcal{A}(H')$ is then used iteratively to visit the $(i+1)$-faces which are superfaces of the marked $i$-faces, for $i$ increasing from 2 to $m-1$. In this manner, all faces that could possibly require updating when inserting $h$ into the arrangement have been marked.

Now let us consider the case where $\mathcal{A}(H')$ is a homogeneous arrangement containing the vertex at the origin. If $f$ is a face of $\mathcal{A}(H')$ whose intersection with $h$ is a single point, then $f$ must be an edge containing the origin. Since $\mathcal{A}(H')$ contains the vertex at the origin, such an edge $f$ cannot exist. This implies that a face of $\mathcal{A}(H')$ whose closure intersects $h$ in a single point (the origin) must itself not intersect $h$. This argument leads us to the following observation.

**Observation 6.2** *Let $H'$ be a set of hyperplanes such that $\mathcal{A}(H')$ is a homogeneous hyperplane arrangement in $E^m$ containing the vertex at the origin. Let $h$ be a new hyperplane to be added to $\mathcal{A}(H')$, such that $h$ passes through the origin. Then for any face $f \in \mathcal{A}(H')$, if $cl(f) \cap h = \{0\}$, the vertex at the origin, then face $f$ and its incidences are unchanged after the insertion of $h$.*

Therefore, for homogeneous arrangements, we need only mark and update those cells whose closures intersect $h$ in a face of $\mathcal{A}(H' \cup \{h\})$ of dimension at least 1. The vertex at the origin is handled in a very straightforward manner: since it is a subface of all edges of a homogeneous arrangement, it may be updated as each new edge is created. To start the marking process, we find an initial 2-face $f_0$ where $cl(f_0) \cap h$ is of dimensionality greater than 0. Observing that a 2-face in a homogeneous arrangement of $\mathcal{A}_n^m$ has exactly two 1-dimensional subfaces, and that $h$ intersects every 2-flat in the arrangement, we may pick any 2-flat and sweep radially about the origin until $h$ is encountered:

*Find starting 2-face $f_0$*

(1) Let $e$ be an arbitrary edge of $\mathcal{A}(H')$ and let $f$ be a 2-face incident with $e$.

(2) While $cl(f) \cap h = \{\mathbf{0}\}$ do the following: let $e'$ be the edge incident with $f$ other than $e$, and let $f'$ be the 2-face incident upon $e'$ and different from $f$, such that $\mathit{aff}(f') = \mathit{aff}(f)$. Set $e \leftarrow e'$ and $f \leftarrow f'$.

(3) Set $f_0 \leftarrow f$.

Once the starting 2-face $f_0$ is obtained, we must be sure that all faces of $\mathcal{A}(H')$ whose closures intersect $h \setminus \{\mathbf{0}\}$ may be reached from $f_0$ without accessing the vertex at the origin. Let $h^*$ be a hyperplane avoiding the origin and orthogonal to $h$. The intersection of $h^*$ with the homogeneous arrangement $\mathcal{A}(H')$ yields a non-homogeneous $(m-1)$-dimensional arrangement in $h^*$. Since the intersection of $h$ with the union of the vertices, edges, and 2-faces of this arrangement in $h^*$ is connected, and since the choice of $h^*$ is arbitrary, we make the following observation:

**Observation 6.3** *Let $H'$ be a set of hyperplanes such that $\mathcal{A}(H')$ is a homogeneous hyperplane arrangement in $\boldsymbol{E}^m$ containing the vertex at the origin, for $m \geq 3$. Let $h$ be a hyperplane not in $H'$, such that $h$ passes through the origin. Then the intersection of $h$ with the union of 1-faces, 2-faces and 3-faces in $\mathcal{A}(H')$ is connected.*

From this observation, it follows that every 1-face, 2-face, and 3-face whose closure contains a ray in $h$ is reachable from $f_0$. By visiting the superfaces of these faces, and iterating upon the superfaces as in the standard algorithm, every face of $\mathcal{A}(H')$ whose closure contains a ray in $h$ may be visited and marked.

## 6.3 Constructing Half-Space Arrangements

Although we have seen all the modifications that will allow the construction of homogeneous hyperplane arrangements, nothing described as yet allows for the construction of half-space arrangements. Before summarizing the modifications of the incremental construction algorithm, we discuss how information stored with the arcs of the incidence graph will prove useful for the algorithms of the next section.

Let $H_>$ be a set of open half-spaces in $\vec{\Pi}_o^{d+1}$, such that the set of hyperplanes bounding half-spaces in $H_>$ is $H$. Furthermore, let us assume that no two half-spaces of $H_>$ share a common bounding hyperplane. Let $f$ and $g$ be proper faces of $\mathcal{A}(H)$, such that $g$ is a subface of $f$. Let $H^*$ the set of hyperplanes containing $g$ but avoiding $f$, and let $H_>^*$ be the set of half-spaces of $H_>$ bounded by the hyperplanes of $H$. Note that $f$ and $g$ are contained in the same hyperplanes and half-spaces of $H \setminus H^*$ and $H_> \setminus H_>^*$, respectively. The set $H^*$ may be thought of as those hyperplanes of $H$ that "distinguish" $f$ from $g$.

In the incidence graph $\mathcal{I}(H)$, we may label the arcs from $f$ to $g$ and $g$ to $f$ in accordance with the sets $H^*$ and $H_>^*$. The set $H_>^*$ may be partitioned into two sets, $H_+^*$ and $H_-^*$, where the former consists of the those half-spaces of $H^*$ containing $f$, and the latter consists of those avoiding $f$. With both arcs $f \to g$ and $g \to f$, we associate the *differential values* $\chi_+(f,g) = \chi_+(g,f) = |H_+^*|$ and $\chi_-(f,g) = \chi_-(g,f) = |H_-^*|$. With these values, if we know the quantities $\chi_o(g)$ and $\chi_c(g)$, then

$$\chi_o(f) = \chi_o(g) + \chi_+(g,f) \quad \text{and} \quad \chi_c(f) = \chi_c(g) - \chi_-(g,f). \tag{6.1}$$

Alternatively, if we know the quantities $\chi_o(f)$ and $\chi_c(f)$, then

$$\chi_o(g) = \chi_o(f) - \chi_+(f,g) \quad \text{and} \quad \chi_c(g) = \chi_c(f) + \chi_-(f,g). \tag{6.2}$$

In the incremental construction algorithm, when a hyperplane $h$ is introduced into an arrangement $\mathcal{A}(H)$, not necessarily homogeneous, the sets of faces of $\mathcal{A}(H \cup \{h\})$ contained in $h$ become available, as well as the superfaces of these faces. If the hyperplanes of $H$ are associated with open half-spaces of $H_>$, and if differential values are being maintained, the new differential values resulting from the introduction of $h$

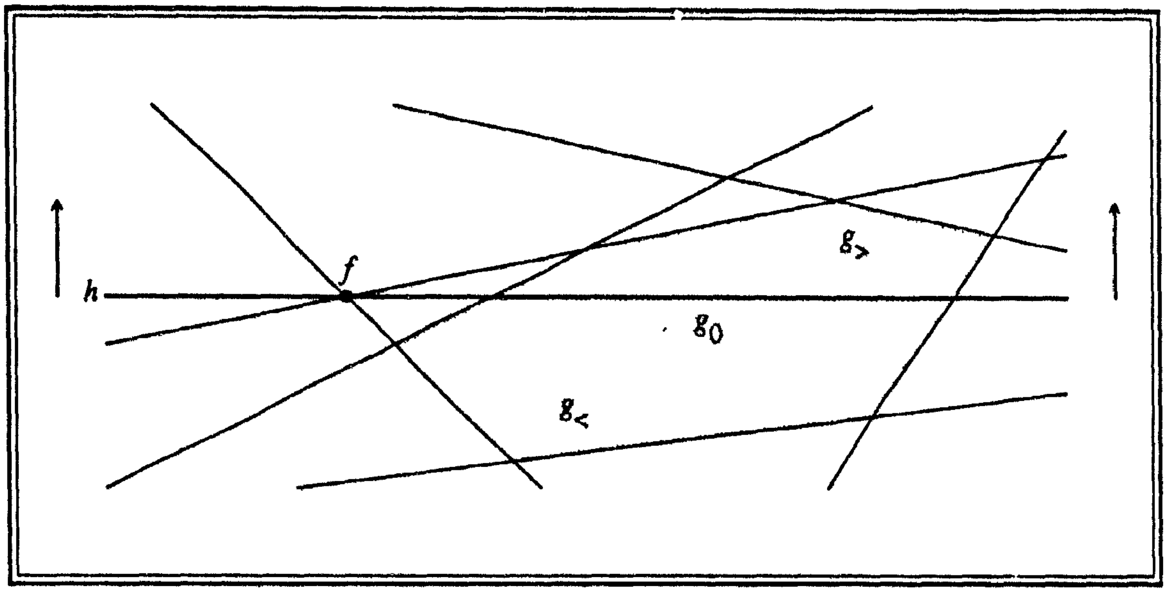


Figure 6.4: Updating differential values when inserting hyperplane $h$

are easily calculated from the old values when the marked faces of $\mathcal{A}(H)$ are updated. While we will not give the formal details of the update process, we shall illustrate how the differential values are maintained on an example.

In Figure 6.4, the orientation of the open half-space $h_>$ associated with $h$ is given by the arrows. The old face $g$ (shaded) is split by $h$ into new the faces $g_>$, $g_0$ and $g_<$ as shown. Since these faces have been newly created, the differential values $\chi_+(g_>, g_0) = \chi_+(g_0, g_>)$ and $\chi_-(g_>, g_0) = \chi_-(g_0, g_>)$ are set to 1 and 0, respectively. The values $\chi_+(g_<, g_0) = \chi_+(g_0, g_<)$ and $\chi_-(g_<, g_0) = \chi_-(g_0, g_<)$ are set to 0 and 1, respectively. The differential values of $g_>$, $g_0$, and $g_<$ with respect to other faces are inherited from $g$, unchanged. In the case of face $f$, a face of $\mathcal{A}(H)$ entirely contained in $h$, the differential values with respect to its superfaces are incremented appropriately. The values with respect to its subfaces are inherited as in the case of $g$, since these subfaces are also contained in $h$.

We shall now summarize the modifications of the incremental arrangement construction algorithm for homogeneous arrangements with a vertex at the origin. The input is assumed to be a set of half-spaces $H_>$ in $\boldsymbol{E}^m$, for $m \geq 3$, whose bounding hyperplanes contain the origin. Furthermore, if $H$ is the set of these bounding hyperplanes, no two half-spaces of $H_>$ are bounded by the same hyperplane of $H$. Finally, we assume that there exists some subset $H'$ of $H$ whose hyperplanes intersect in the single point $\mathbf{0}$. Otherwise, if there exists no such subset $H'$, the arrangement is constructed in a lower dimensional setting, as in the standard method.

*MODIFIED-CONSTRUCTION*

(1) Find subset $H'$, and construct the homogeneous arrangement $\mathcal{A}(H')$ containing the vertex at the origin, as in the standard method. Initialize the differential values of the incidences of $\mathcal{A}(H')$.

(2) If $H \setminus H'$ is empty, terminate. Otherwise, let $h$ be a hyperplane of $H \setminus H'$. Label the vertex at the origin as being entirely contained in $h$.

(3) Find a 2-face $f_0$ of $\mathcal{A}(H')$ such that $cl(f_0) \cap h \neq \{\mathbf{0}\}$ as detailed earlier.

(4) Visit and mark the faces $f$ of $\mathcal{A}(H')$ where $cl(f) \cap h \neq \{\mathbf{0}\}$ as in the standard method, but with every reference to $i$-faces replaced by a reference to $(i+1)$-faces.

(5) Update the marked faces as in the standard method, as well as their differential values. Add $h$ to $H'$, and go to step 2.

We now investigate the time complexity of the modified algorithm. The time required to add a new hyperplane $h$ to an arrangement $\mathcal{A}(H')$, including the time spent updating the differential values, is of the order of the number of faces of $\mathcal{A}(H' \cup \{h\})$ contained in the closures of cells $f$ such that $cl(f) \cap h$ is of dimension at least one. This subset of the faces of the homogeneous arrangement $\mathcal{A}(H' \cup \{h\})$ we shall call the *homogeneous zone* of $\mathcal{A}(H' \cup \{h\})$ defined by $h$.

As was established earlier, there exist parallel hyperplanes $h_a$ and $h_b$ that between them intersect every face of $\mathcal{A}(H' \cup \{h\})$ (other than the vertex at the origin). Thus each face of the homogeneous zone defined by $h$ contains a face of the non-homogeneous zone of $h_a \cap h$ in the $(m-1)$-dimensional arrangement formed by the intersection of $\mathcal{A}(H' \cup \{h\})$ and $h_a$, or of the zone of $h_b \cap h$ in the arrangement formed by $\mathcal{A}(H' \cup \{h\})$ in $h_b$. Since the size of a zone in an $(m-1)$-dimensional arrangement is $O(n^{m-2})$, the size of a homogeneous zone in $\mathcal{A}(H' \cup \{h\})$ must also be of order $O(n^{m-2})$, and therefore the time required to insert $h$ into $\mathcal{A}(h')$ is $O(n^{m-2})$. Summing this time cost over all the incrementations, we arrive at the following:

**Lemma 6.4** *Let $H_>$ be a set of half-spaces in $E^m$, for $m \geq 3$, such that no two half-spaces of $H_>$ share a common bounding hyperplane, and such that the bounding hyperplanes contain the origin. Then the homogeneous arrangement $\mathcal{A}(H)$ may be constructed in $O(n^{m-1})$ time using $O(n^{m-1})$ space.*

Because the vertex at the origin of the homogeneous arrangement is contained in every hyperplane of the arrangement, it is a subface of every edge. Hence there is no need to maintain the improper face $\emptyset$ to allow access to the faces of the arrangement of low dimension. Accordingly, we may omit the improper face $\emptyset$ from the incidence graph, and consider the vertex at the origin to be an *improper* face.

## 6.4 Finding Weak Separators

We shall now outline an algorithm to determine the maximum values of $\chi_o(f)$ and $\chi_c(f)$ over all faces $f$ of a homogeneous half-space arrangement, and to locate faces attaining these maximum values. If the half-space arrangement is the dual arrangement of a set of labeled points, then the faces produced correspond to maximal components of these points. The algorithm first constructs the homogeneous arrangement using the modification of the incremental algorithm due to Edelsbrunner, O'Rourke, and Seidel, and then performs a breadth-first search within the arrangement to visit all faces.

The input to Algorithm MAXCOMP consists of a set of $n$ distinct open halfspaces $H_>$, whose bounding hyperplanes contain the origin. Let $H$ be the set of such bounding hyperplanes. We assume that no two half-spaces of $H_>$ are bounded by the same hyperplane of $H$. We also assume that the hyperplanes of $H$ have as their common intersection the single point 0. Otherwise, the algorithm is applied in an appropriate lower dimension.

The output of Algorithm MAXCOMP will be the incidence graph $\mathcal{I}(H)$, whose arcs will be augmented by the differential values, as well as two pointers, *maxstrict* and *maxnonstrict*, to proper faces $f$ and $g$ of $\mathcal{I}(H)$ attaining the maximum values of $\chi_o(f)$ and $\chi_c(g)$, respectively.

Each record of $\mathcal{I}(H)$ has space reserved for labeling purposes. Each facet of the arrangement, being contained in precisely one hyperplane $h(f)$ of $H$, shall have storage space for a parameterization of the unique half-space $h_>(f)$ of $H_>$ bounded by $h(f)$. With each face $f$ of the arrangement, we will also reserve space for two integers: the quantities $\chi_o(f)$ and $\chi_c(f)$. With each arc $f \to g$ in $\mathcal{I}(H)$, we shall reserve space for the differential values $\chi_+(f,g)$ and $\chi_-(f,g)$. In addition, we maintain a queue $\mathcal{Q}$ of pointers to nodes of $\mathcal{I}(H)$.

*MAXCOMP*

(1) From the hyperplanes of $H$, and the half-spaces of $H_>$, construct the homogeneous hyperplane arrangement $\mathcal{A}(H_>)$, including the differential values, using the modified incremental arrangement construction algorithm of the previous sections.

(2) Choose any cell $f_0$ of $\mathcal{A}(H)$, and mark it as being *visited*. Determine $\chi_o(f_0)$ by testing the representative point of $f$ for inclusion in each of the half-spaces of $H_>$. Set $\chi_c(f) \leftarrow \chi_o(f_0)$. Set *maxstrict* $\leftarrow \uparrow f_0$, and set *maxnonstrict* $\leftarrow \uparrow f_0$. Initialize queue $\mathcal{Q}$ to contain $\uparrow f_0$.

(3) If $\mathcal{Q}$ is empty, terminate. Otherwise, dequeue the pointer to face $f$ from $\mathcal{Q}$.

(4) For every unvisited proper subface and superface $g$ of $f$ do:

(4a) Using the equations (6.1) and (6.2), calculate the quantities $\chi_o(g)$ and $\chi_c(g)$ from the differential values $\chi_+(f,g)$ and $\chi_-(f,g)$, and store them in $g$.

(4b) Let $g_o$ and $g_c$ be the faces pointed to by *maxstrict* and *maxnonstrict*, respectively. If $\chi_o(g) > \chi_o(g_o)$, then set *maxstrict* $\leftarrow \uparrow g$. If $\chi_c(g) > \chi_c(g_c)$, then set *maxnonstrict* $\leftarrow \uparrow g$.

(4c) Mark $g$ as visited, and enqueue $\uparrow g$ onto $\mathcal{Q}$.

(5) Go to Step 3.

Note that the breadth-first search strategy employed here is certainly not the only one possible: visiting every face of the arrangement could just as easily have been done using depth-first search techniques. Also, upon construction, every facet of the arrangement may easily be associated with a pointer to the unique hyperplane of $H$ containing it.

The space required by Algorithm MAXCOMP is proportional to the size of the incidence graph $\mathcal{I}(H)$; that is, $O(n^{m-1})$. Certainly, the size of the queue $\mathcal{Q}$ may not exceed the size of $\mathcal{I}(H)$. The time required is proportional to the time taken to build the arrangement, plus a constant amount for every incidence between nodes of $\mathcal{I}(H)$. By marking the faces as visited when encountered, we are guaranteed that each incidence between faces $f$ and $g$ in the arrangement may be examined at most twice: once from $f$ to $g$, and once from $g$ to $f$. Since the time required to construct the arrangement is $O(n^{m-1})$, and the number of incidences between faces of a homogeneous arrangement is $O(n^{m-1})$, the total time required by the algorithm is also $O(n^{m-1})$.

If the input to the algorithm is the set of open half-spaces $H_{>} = \rho_o(\vec{l}(P))$, for some set of distinct labeled points $P$ in $\boldsymbol{E}^d$, then Lemma 4.2 implies that the faces of $\mathcal{A}(H_{>})$ pointed to by *maxstrict* and *maxnonstrict* correspond to maximal strict and non-strict linearly separable components of $P$, respectively. Let these faces be $g_o$ and $g_c$, respectively. More precisely, a maximal strict component of $P$ is the subset $C_o$ of $P$ such that $\rho_o(\vec{l}(C_o))$ is the set of all half-spaces of $H_{>}$ containing $g_o$. If $H_{\geq}$ is the set of closures of half-spaces of $H_{>}$, then a maximal non-strict component of $P$ is the subset $C_c$ of $P$ such that $\rho_c(\vec{l}(C_c))$ is the set of all half-spaces of $H_{\geq}$ containing $g_c$. These components may be explicitly obtained in $O(n)$ time by testing representative points of $g_o$ and $g_c$ for inclusion in the half-spaces of $H_{>}$ and $H_{\geq}$.

The representative points of $g_o$ and $g_c$, with one possible exception, correspond to hyperplanes in $\boldsymbol{E}^d$ that are weak strict and non-strict linear separators of the red and green points of $P$. The exception is the set of points on the pair of rays $\vec{r}_\infty$ and $-\vec{r}_\infty$, discussed in Section 4.3. Since these rays are contained in cells of $\mathcal{A}(H_>)$, and since the cells are otherwise composed of rays corresponding to valid separators in $\boldsymbol{E}^d$, rays of $g_o$ and $g_c$ associated with weak strict and non-strict linear separators, respectively, may easily be obtained.

If the input to the algorithm is instead the set of open half-spaces $H_> = \sigma_o(\vec{s}(P))$, then Lemma 4.5 implies that the faces $g_o$ and $g_c$ correspond to weak strict and non-strict spherically separable components of $P$, respectively. The components may be obtained in $O(n)$ time as in the linear case. Although some rays of $g_o$ and $g_c$ correspond to valid weak strict and non-strict spherical separators of $P$, there are two cases where some rays are not. Some care is therefore required in the choice of a representative ray from these faces.

Let $f$ be a face of the $m$-dimensional arrangement $\mathcal{A}(H_>)$ containing the ray $\vec{r} \notin \vec{E}^m_*$. Recall that $\vec{E}^m_*$ is the set of all rays in $\vec{E}^m$ corresponding to degenerate and non-degenerate hyperspheres of $\boldsymbol{E}^{m-2}$. If $\vec{r}$ corresponds to a hypersphere with squared radius "$\leq 0$", then the rays of $f \cap \vec{E}^m_*$ correspond the non-degenerate hyperspheres of $\boldsymbol{E}^{m-2}$ not containing or intersecting any of the points of $P$. Hence any empty hypersphere guaranteed not to intersect the points of $P$ will do as a representative of $f$. If $\vec{r}$ corresponds to a degenerate hypersphere, we have seen in Section 4.5 that any other ray of $f \cap \vec{E}^m_0$ will adequately represent of $f$.

We summarize these results in the following theorem:

**Theorem 6.5** *Let $P$ be a set of $n$ distinct labeled points in $\boldsymbol{E}^d$. A weak strict or non-strict linear separator of $P$ may be found in $O(n^d)$ time and space, and a weak strict or non-strict spherical separator may be found in $O(n^{d+1})$ time and space.*

If more than one weak separator is desired, lists of pointers to candidate faces may be maintained during the visiting of the faces in Algorithm MAXCOMP. The list of faces may be examined afterwards for suitable separators. The time and space required to maintain these lists is dominated by the complexity of the homogeneous arrangement. If non-degenerate weak spherical separators are sought, then those faces whose constituent rays correspond solely to degenerate hyperspheres may be easily identified and disregarded once visited.

Finally, the number of linearly or spherically separable subsets of $P$, of fixed cardinality or of all cardinalities, either strict or non-strict, may be counted by means of a slight modification of Algorithm MAXCOMP, where the sums of the formulae of Corollary 5.18 and Corollary 5.19 are maintained during the visiting of the faces of the homogeneous arrangement $\mathcal{A}(\rho_o(\vec{l}(P)))$ or $\mathcal{A}(\sigma_o(\vec{s}(P)))$, as appropriate. The time and space complexities of the algorithm justify the following theorems:

**Theorem 6.6** *Let $P$ be a set of $n$ distinct labeled points of $\mathbf{E}^d$, such that no hyperplane contains every point of $P$. Let $k$ be a positive integer. Then*

1. *the number of non-empty non-strict linearly separable subsets of $P$ of fixed cardinality $k$,*
2. *the number of non-empty strict linearly separable subsets of $P$ of fixed cardinality $k$,*
3. *the total number of non-empty non-strict linearly separable subsets of $P$,*
4. *and the total number of non-empty strict linearly separable subsets of $P$*

*may all be calculated in $O(n^d)$ time using $O(n^d)$ space.*

**Theorem 6.7** *Let $P$ be a set of $n$ distinct labeled points of $\mathbf{E}^d$, such that no hyperplane or hypersphere contains every point of $P$. Let $k$ be a positive integer. Then*

1. *the number of non-empty non-strict spherically separable subsets of $P$ of fixed cardinality $k$,*
2. *the number of non-empty strict spherically separable subsets of $P$ of fixed cardinality $k$,*
3. *the total number of non-empty non-strict spherically separable subsets of $P$,*
4. *and the total number of non-empty strict spherically separable subsets of $P$*

*may all be calculated in $O(n^{d+1})$ time using $O(n^{d+1})$ space.*

## 6.5 Space Reduction Using the Topological Sweep

One of the drawbacks of the constructive approach of the previous section is that all the faces of the dual arrangement are constructed before any of them are scanned. Despite the intuitive appeal of exploiting the full structure of the arrangement, we may save a great deal of space by having no more than a limited number of faces on hand at any given time. This section is devoted to the application of a "topological" sweep line algorithm of Edelsbrunner and Guibas [EG86] towards the solution of some variants of the weak separation problems. The paradigm in its basic form is a two-dimensional one: it may be used to construct a planar arrangement, but as yet there is no known extension for arrangement construction in higher dimensions. This is not to say that the topological sweep has no benefits towards solving problems in dimensions greater than two: in the same paper, the authors describe how the two-dimensional topological sweep may be used to enumerate (but not construct) the faces of arrangements in higher dimensions.

We will first provide a brief overview of Edelsbrunner and Guibas' general topological sweep line method in two dimensions. For further details of their method, the interested reader is referred to [EG86].

Let $l$ be a line in the plane, parameterized using cartesian coordinates. Line $l$ will be said to be *vertical* if $l$ is a translate of the cartesian $x_2$-axis. If $l$ is non-vertical, a point $x \in E^2$ will be said to be *above* $l$ if the ray parameterized by $x + \lambda(0,1)$ does not intersect $l$, where $\lambda$ is restricted to be non-negative. Point $x$ is *below* $l$ if the same holds true with $\lambda$ instead restricted to be non-positive.

Let $L$ be a set of $n$ non-vertical lines in the plane, and let $\mathcal{A}(L)$ be the arrangement of these lines. The unique region $f_T$ above all lines of $L$ will be called the *top* region of $\mathcal{A}(L)$, and the unique region $f_B$ below all lines of $L$ will be called the *bottom* region.

A topological sweep line may be viewed as a sequence of cells and edges of $\mathcal{A}(L)$, called a *cut*, such that each cut edge is contained by a different line of $L$, and every line of $L$ contains an edge in the cut. The edges of the cut are ordered from top to bottom in that the first edge is on the lower chain of $f_T$, the last edge is on the upper chain of $f_B$, and two adjacent edges $e_i$ and $e_{i+1}$ are on the upper and lower chains of

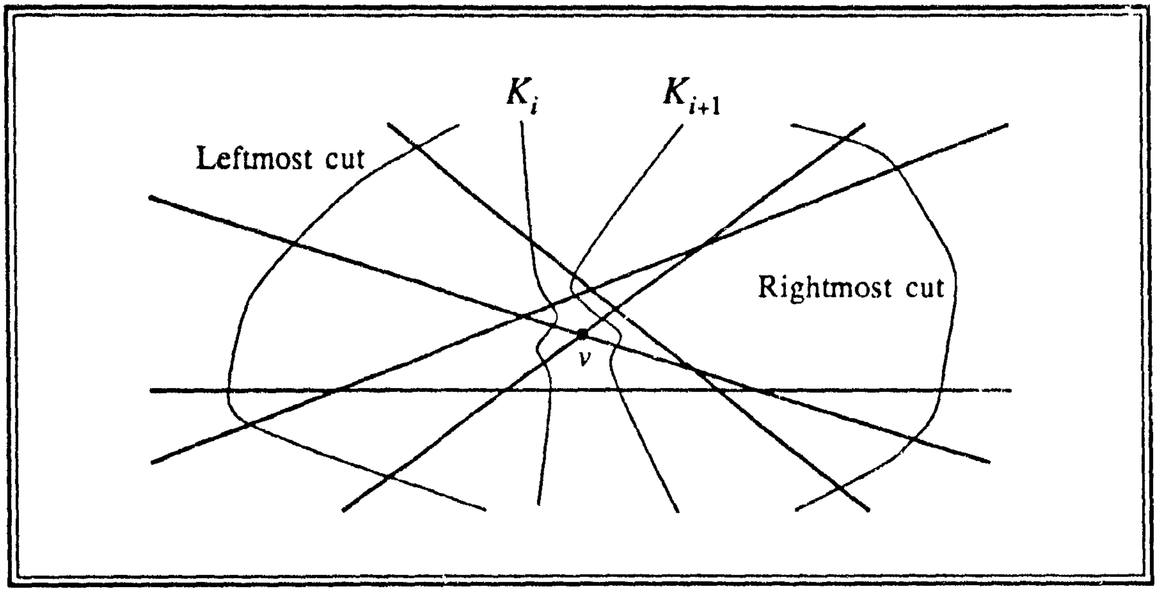


Figure 6.5: Examples of cuts in an arrangement of lines

some cell $j$, respectively. The sweep itself consists of a sequence of cuts, ordered from "left" to "right". The first or "leftmost" cut is the sequence of left unbounded edges of $\mathcal{A}(L)$, and the last or "rightmost" cut consists of the right unbounded edges. Two adjacent cuts $K_i$ and $K_{i+1}$ differ in that exactly one vertex $v$ that is to the right of $K_i$ is to the left of $K_{i+1}$ (see Figure 6.5). The advancement of the sweep line from $K_i$ to $K_{i+1}$ past $v$ is called an *elementary step.*

Edelsbrunner and Guibas use data structures they call ***horizon trees*** to store information concerning the regions intersected by the sweep line. They show that the storage required to maintain these trees is $O(n)$, enabling the entire sweep to be performed using only linear storage. Initially, these trees contain the upper and lower chains of all the regions of $\mathcal{A}(L)$ that are unbounded to the left. As the sweep line advances past vertex $v$, the regions that have $v$ as their unique leftmost bounding vertex (unique due to the absence of vertical lines in $\mathcal{A}(L)$) have their upper and lower chains immediately available from the horizon trees. This allows certain attributes of new faces in the cut to be calculated based on the attributes of the faces in the previous cut. Every face of $\mathcal{A}(L)$ is examined, since those regions without leftmost vertices are examined before the first elementary step, and each region having a leftmost vertex must be examined during some elementary step of the sweep. The worst-case time required by their algorithm is $O(n^2)$.

Consider now the homogeneous half-space arrangement $\mathcal{A}(H_>)$ in $\boldsymbol{E}^3$, where $H$ is the set of bounding planes of the half-spaces of $H_>$. Every face $f$ of $\mathcal{A}(H_>)$ of dimension greater than zero may be visited, and the quantities $\chi_o(f)$ and $\chi_c(f)$ produced, using the following application of the topological line sweep algorithm:

*3-TOPOSWEEP*

(1) Produce two parallel planes $h_a$ and $h_b$ which together intersect all faces of $\mathcal{A}(H_>)$ except the vertex at the origin, as follows:

(1a) Set $high1 \leftarrow high2 \leftarrow 0$ and $degen1 \leftarrow degen2 \leftarrow false$.

(1b) For every pair of planes $h_i$, $h_j$ in $H$, for $i \neq j$, do:

(1b1) Compute the rays $\vec{u}_{ij}$ and $-\vec{u}_{ij}$ of $\vec{E}^3$ in the intersection of $h_i$ and $h_j$.

(1b2) Let $u = (u_1, u_2, u_3)$ be any point of $\vec{u}_{ij}$. Then do:

$temp1 \leftarrow \frac{u_3^2}{u \cdot u}$;
$temp2 \leftarrow \frac{u_2^2}{(u_1^2+u_2^2)}$;
**if** $temp1 = 0$ **then**
  $degen1 \leftarrow true$;
  **if** $temp2 = 0$ **then**
    $degen2 \leftarrow true$
  **elseif** $temp2 > high2$ **then**
    $high2 \leftarrow temp2$
  **endif**
**elseif** $temp1 > high1$ **then**
  $high1 \leftarrow temp1$
**endif**

(1c) If $degen1 = false$ then let $v = (0, 0, 1)$; otherwise,

(1c1) If $degen1 = false$ then let $v = (0, v_2, 1)$ be a point such that $v_2 > 0$, and $v_2^2 < \frac{1}{1-high1}$; otherwise, (1c2) Let $v = (v_1, v_2, 1)$ be a point such that $v_1, v_2 > 0$, $v_2^2 < \frac{1}{1-high2}$, and $v \cdot v < \frac{1}{1-high1}$.

(1d) Set $h_a \leftarrow \{x \in E^3 | \, v \cdot x = 1\}$, and Set $h_a \leftarrow \{x \in E^3 | \, v \cdot x = -1\}$.

(2) In each of $h_a$ and $h_b$ do the following:

(2a) Intersect the planes of $H$ and the half-spaces of $H_>$ with $h_a$, realigning the coordinate axes such that no resulting line in $h_a$ is vertical. Let $L$ and $L_>$ be the respective intersections of $H$ and $H_>$ with $h_a$.

(2b) For all cells and edges $f$ contained in the leftmost cut of $\mathcal{A}(L_>)$, explicitly compute $\chi_o(f)$ and $\chi_c(f)$.

(2c) Visit the faces of $\mathcal{A}(L_>)$ using the topological sweep method. At each elementary step past vertex $v$, compute $\chi_o(v)$ and $\chi_c(v)$, as well as $\chi_o(f')$ and $\chi_c(f')$ for each new face $f'$ in the cut.

(3) Repeat Step 2 with $h_b$ replacing $h_a$.

Step 1 of the algorithm is an explicit construction of two planes intersecting every face of $\mathcal{A}(H)$ of dimension greater than zero. The parallel planes $h_a$ and $h_b$ are constructed in such a way that the only faces of $\mathcal{A}(H)$ that could fail to be intersected is the vertex at the origin, and any edge contained in the translate of $h_a$ and $h_b$ containing the origin. The orientation of $h_a$ and $h_b$ is chosen in order to guarantee the impossibility of the latter.

Once Step 1c is reached, the boolean variable *degen1* holds the value *false* if and only if there exists no edge of $\mathcal{A}(H_>)$ contained in the plane $h_0 = \{x \in E^3 | \, x_3 = 0\}$, and *degen2* holds the value false if and only if there exists no edge contained in the line $l_0 = \{x \in E^3 | \, x_2 = x_3 = 0\}$. If *degen1* is *false*, then $h_a$ and $h_b$ may be safely chosen to be translates of $h_0$. Otherwise, we exploit the well-known relationship between the inner product of two vectors and the angle between them (see [Bor69] for more details).

At Step 1c, the variable *high1* stores the value

$$\cos^2 \theta_1 = \left( \frac{(0,0,1) \cdot u}{\|u\|} \right)^2 = \frac{u_3^2}{u \cdot u},$$

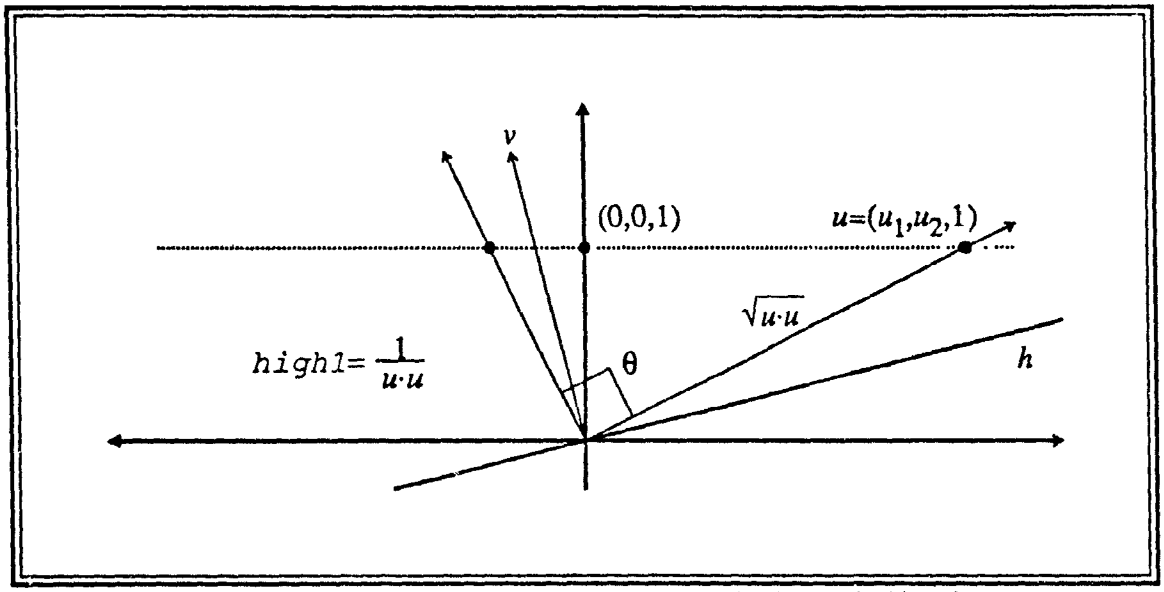


Figure 6.6: Finding a plane avoiding all edges of $\mathcal{A}(H_>)$

where $\theta_1$ is the maximum of the angles between the ray $\vec{r}_\infty = \{\lambda(0,0,1) \in \vec{E}^3 | \lambda > 0\}$ and the edges of $\mathcal{A}(H_>)$ located in $\vec{E}^3_R = \{\vec{x} \in \vec{E}^3 | \vec{r}_\infty \cdot \vec{x} > 0\}$ (see Figure 6.6). The variable *high2* stores the value

$$\cos^2\theta_2 = \left(\frac{(0,1)\cdot(u_1,u_2)}{\|(u_1,u_2)\|}\right)^2 = \frac{u_2^2}{u_1^2+u_2^2},$$

where $\theta_2$ is the maximum of the angles between the ray $\vec{r}_\infty = \{\lambda(0,1,0) \in \vec{E}^3 | \lambda > 0\}$ and the projections of the edges of $\mathcal{A}(H_>)$ located in $\vec{E}^3_R$ onto the plane $h_0$. The reader is invited to verify that the point $v$ is chosen such that the edges of $\mathcal{A}(H_>)$ in $\vec{E}^3_R$ are guaranteed not to lie in $h_v = \{x \in E^3 | v \cdot x = 0\}$, and those in $h_0$ (if any) are guaranteed not to lie in $h_v \cap h_0$.

The time required to perform Step 1 is of the same order as the maximum number of edges of $\mathcal{A}(H_>)$ – that is, $O(n^2)$. Since no attempt is made to store all pair-wise intersections at the same time, the storage required for this step is $O(n)$.

In Steps 2 and 3, the topological line sweep is performed in the planes $h_a$ and $h_b$. The computation of $\chi_o$ and $\chi_c$ may be performed in $O(n^2)$ total time for the faces of the initial cut. At each elementary step, the time required to calculate the values of $\chi_o$ and $\chi_c$ for the new faces of the cut is proportional to the number of these new faces. Hence this computation does not change the asymptotic time (or space) complexity of the algorithm.

**Theorem 6.8** *Let $H_>$ be a set of open half-spaces of $\mathbf{E}^3$ whose bounding planes contain the origin, such that no two half-spaces share the same bounding plane. Then $\chi_o$ and $\chi_c$ may be calculated for every face of the homogeneous half-space arrangement $\mathcal{A}(H_>)$ in $O(n^2)$ time using $O(n)$ space.*

Three minor refinements of Algorithm 3-TOPOSWEEP are worthy of mention First, we observe that by the symmetry of the arrangement $\mathcal{A}(H_>)$, Steps 2 and 3 may be compressed into one pass over either $h_a$ or $h_b$. Secondly, the computation of $\chi_o$ and $\chi_c$ for the faces of the initial cut may be calculated in linear time instead of quadratic time, taking advantage of the horizon tree data structure. In any event, this has no effect on the overall asymptotic time complexity. Thirdly, a face $f$ maximizing $\chi_o(f)$ may be produced, as well as those half-spaces of $H_>$ containing $f$, without changing the overall complexity. A face $f$ is abandoned in favour of a face $g$ only if $\chi_o(f) > \chi_o(g)$; since $\chi_o$ may attain only at most $n+1$ different values, this change may only occur at most $n$ times. Each such replacement may be performed in $O(n)$ time, and thus the total cost in time to maintain this list of half-spaces is $O(n^2)$. Naturally, the same methods may be applied in the case of $\chi_c$.

In their paper, Edelsbrunner and Guibas observed that the faces of an arrangement of hyperplanes (not necessarily homogeneous) may be visited by sweeping along two-dimensional slices of the arrangement. We shall now apply this general technique to the weak separation problems of the previous section, using Algorithm 3-TOPOSWEEP as a "primitive operation".

Let $H_>$ be a set of $n$ half-spaces in $\boldsymbol{E}^m$, whose bounding hyperplanes contain the origin, such that no two hyperplanes of $H_>$ share a common bounding hyperplane. Let $H$ be the set of these bounding hyperplanes. We shall place the following additional restrictions on the hyperplanes of $H$: every $j$ hyperplanes of $H$ must have an $(m-j)$ flat as their common intersection, for all $j = 1, 2, \ldots, m$. Furthermore, no $j+1$ hyperplanes may intersect in a common $(m-j)$-flat, for all $j = 1, 2, \ldots, m-1$. The homogeneous half-space arrangement $\mathcal{A}(H_>)$ shall then be said to be *simple*.

One characteristic of simple homogeneous arrangements is that (other than at the origin) there are no degeneracies. Every hyperplane avoiding the origin, when intersected with a simple homogeneous arrangement in $\boldsymbol{E}^m$, produces a simple non-homogeneous $(m-1)$-dimensional arrangement. If the homogeneous arrangement is the image of a set of labeled points under the transforms of Chapter 4, then the strict and non-strict components are easily seen to be the same. For this reason, in the algorithm to follow, we will not distinguish between $\chi_o$ and $\chi_c$, and will concentrate only upon the cells of $E^m$. Also, the simplicity of $\mathcal{A}(H_>)$ implies that every subset $H'$ consisting of $d-3$ hyperplanes of $H$ must intersect in a common 3-flat $\varphi$. The intersection of the remaining hyperplanes and their corresponding half-spaces in $\varphi$ forms a (simple) three-dimensional homogeneous half-space arrangement in $\varphi$.

The input to the following algorithm is a set of half-spaces $H_>$ in $\boldsymbol{E}^m$, for $m \geq 3$, and their bounding hyperplanes $H$, as described above. The output is a ray of $\vec{E}^m$ and a list *component* of the half-spaces of $H_>$ containing this ray.

*m-TOPOSWEEP*

(1) Initialize *component* $\leftarrow \emptyset$.

(2) For every subset $H'_>$ consisting of $m-3$ half-spaces of $H_>$ do:

(2a) Let $H'$ be the set of bounding hyperplanes of $H'_>$, and let $\varphi$ be the 3-flat formed by the common intersection of the hyperplanes of $H'$. Let $H^*$ and $H^*_>$ be the intersections of the members of $H \setminus H'$ and $H_> \setminus H'_>$ with $\varphi$, respectively.

(2b) Apply Algorithm 3-TOPOSWEEP to determine a face $f$ of $\mathcal{A}(H^*_>)$ contained in the greatest number of half-spaces of $H^*_>$. Let $\bar{H}_>$ be the set of these half-spaces.

(2c) If $|component| < |\bar{H}_> \cup H'_>|$, then set *component* $\leftarrow \bar{H}_> \cup H'_>$.

(3) Find a ray contained in the common intersection of the half-spaces of *components*, using any convenient method (such as linear programming).

Each iteration of Step 2 may be performed in $O(n^2)$ time using $O(n)$ space, due to the reference to Algorithm 3-TOPOSWEEP in Step 2b. Since there are $\binom{n}{m-3}$ iterations, Step 2 requires $O(n^{m-1})$ time overall. If $m$ is taken to be fixed, the $O(n)$ linear programming algorithm of Megiddo [Meg84] may be used to perform Step 3 Clearly, the complexity of Step 2 dominates the algorithm.

The restriction that the homogeneous arrangement $\mathcal{A}(H_>)$ be simple becomes especially important at Step 2c. In a simple arrangement, the counterpart in $\mathcal{A}(H_>)$ of every 3-face in $\mathcal{A}(H^*_>)$ must be in the closures of precisely $2^{m-3}$ cells of $\mathcal{A}(H_>)$. Hence, we are guaranteed that the common intersection of the open half-spaces of $\bar{H}_>$ is non-empty. Otherwise, we would be forced to conduct a search of the arrangement $\mathcal{A}(\bar{H}_>)$ to find the largest-cardinality subset of $\bar{H}_>$ having a non-empty intersection. Even given the degeneracy of this arrangement (the half-spaces of $\bar{H}_>$ all contain $\varphi$), the potential size of this arrangement is $O(n^{m-3})$ – much too expensive to perform in each of the $\binom{n}{m-3}$ iterations.

The results of Chapter 4 – notably Lemmas 4.2 and 4.5 – together with Algorithm $m$-TOPOSWEEP imply the following:

**Theorem 6.9** *Let $P$ be a set of $n$ distinct labeled points in $\mathbf{E}^d$. If the points of $P$ are in general position, then a weak strict or non-strict linear separator of $P$ may be found in $O(n^d)$ time and $O(n)$ space. Furthermore, if no $d+2$ points are contained in a common hypersphere, then a weak strict or non-strict spherical separator may be found in $O(n^{d+1})$ time and $O(n)$ space.*

The restrictions on the simplicity of the homogeneous arrangements, and by extension on the points of $P$, may be eliminated if non-strict separators are sought. This is true since only the edges of the homogeneous arrangement need be examined, and each edge of the arrangement is contained in some of the 3-dimensional slices examined by Algorithm $m$-TOPOSWEEP. This algorithm may easily be modified to search for these edges, and for this reason, we shall not repeat the details. Since the simplicity of the arrangement is no longer required, we have the following corollary of Theorem 6.9:

**Corollary 6.10** *Let $P$ be a set of $n$ distinct labeled points in $\mathbf{E}^d$. A weak non strict linear separator of $P$ may be found in $O(n^d)$ time and $O(n)$ space, and a weak non-strict spherical separator may be found in $O(n^{d+1})$ time and $O(n)$ space.*

# Chapter 7

# Wide Linear Separation Algorithms

## 7.1 Introduction

In the previous chapter, we examined ways in which separators of labeled point sets of various types could be obtained. Unfortunately, these methods do not concern themselves with the "quality" of the separators produced. Two separators may each determine maximal separable components, but one may be greatly superior to the other when employed as a discriminant function Some of the well-known methods for obtaining linear separators, such as straightforward linear programming, too often yield extreme separators whose effectiveness as a discriminant function is diminished. One might prefer instead a separator that does not approach the points it separates

For the case of linear separation of point sets, one measure of the quality of a strong or weak separator may be the minimum orthogonal Euclidean distance between the hyperplane and the points of the maximal component it determines. Using this criterion, a weak separator $h_a$ of labeled point set $P$ would be judged to be "better" than another weak separator $h_b$ if $\delta(P_a, h_a) > \delta(P_b, h_b)$, where $P_a$ and $P_b$ are the maximal components of $P$ with respect to $h_a$ and $h_b$, respectively. If $h_a$ is such that $\delta(P_a, h_a) \geq \delta(P_b, h_b)$ for all other weak separators $h_b$, then $h_a$ can be said to be a *widest* weak separator. Naturally, these concepts extend to strong separation as well

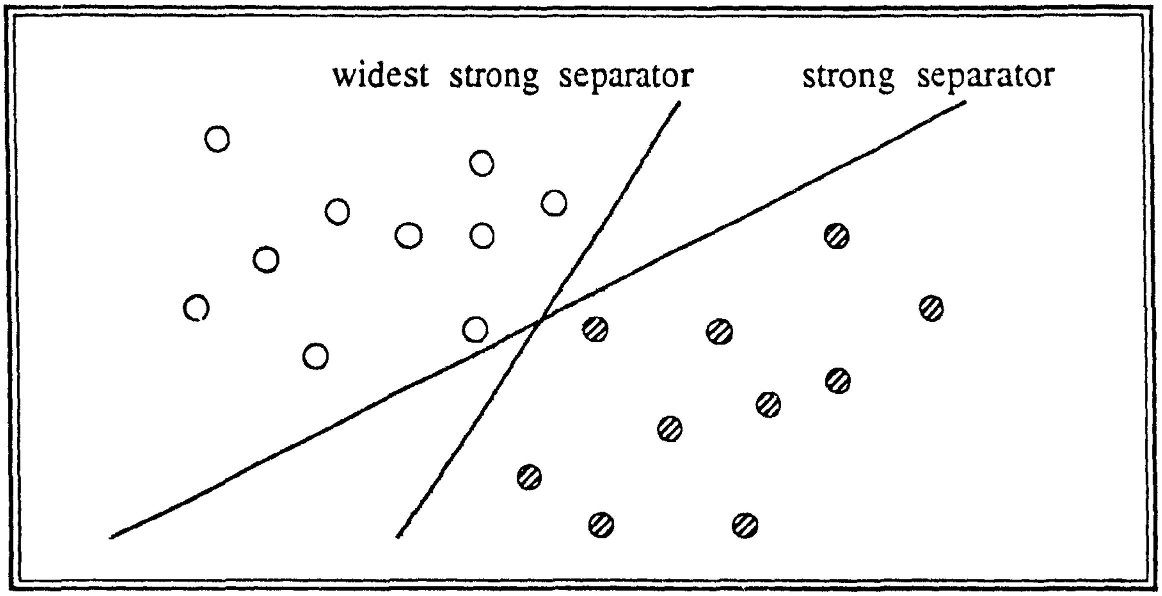


Figure 7.1: A widest strong separator – unweighted case

(see Figure 7.1).

A topic closely related to wide strong linear separation of point sets is the computation of the minimum distance between disjoint convex polytopes, or more precisely, two points determining this minimum distance. It is not difficult to show that the perpendicular bisector of the line segment joining these two points is a widest strong linear separator of the polytopes. In two dimensions, Edelsbrunner [Ede82] showed that, with preprocessing, this line segment may be obtained in $O(\log n)$ time. Schwartz [Sch81] and Chin and Wang [CW82] have also studied this problem. In three dimensions, Dobkin and Kirkpatrick [DK85] have obtained an $O(n)$ time solution. Although these methods may all be adapted to find wide strong linear separators of sets of points, they require that the convex hull of the point sets be given. The next section shall be concerned with a higher-dimensional $O(n)$ time solution to a more general form of the wide separation problem.

In some applications, some points of $P$ may be more "important" than others. Consider the case where every point $p_i$ of $P$ is not only given a label, but also a positive real-valued weight $\omega_i$. The *weighted* orthogonal Euclidean distance between $p_i$ and a hyperplane $h$ is then simply the product $\omega_i\,\delta(p_i, h)$. A second measure of the quality of a weak linear separator is then the minimum weighted orthogonal distance between the separator and the points of its maximal component. A weak separator having the greatest such minimum distance is then called a *widest weighted* weak separator. The unweighted case is simply a special instance of the weighted case where $\omega_i = 1$ for all points $p_i \in P$ Accordingly, for the most part, we shall restrict our discussion in this chapter to the weighted case In this context, we will often refer to widest weighted linear separators as simply "widest linear separators".

In some situations, a set of labeled points has no widest strong or weak linear separators. For instance, if the points of $P$ all share the same label, then any hyperplane avoiding the convex hull of $P$ is a strong separator of $P$. This separator may be moved out to infinity in such a way that the minimum (weighted) orthogonal distance to the points of $P$ diverges to infinity. In fact, whenever $P$ has a maximal strict linearly separable component consisting of points sharing a common label, then there is no widest weak separator of $P$. It is easily seen that this is the only situation where a widest weak separator does not exist.

By the criteria given above, a non-strict linear separator that contains some point of $P$ is a very poor separator indeed. In the previous chapter, we have seen how weak non-strict separators may be obtained; strong non-strict separators may be obtained using linear programming techniques. For this reason, we will only consider the strict aspect of the wide strong and weak linear separation problems.

In this chapter, we will see how wide strong and weak strict linear separators of labeled point sets may be determined. In Section 7.1, we shall examine how the problem of finding wide linear separators of labeled sets of points relates to the problem of finding linear separators of labeled sets of hyperspheres.

## 7.2 Wide Strong Linear Separation

Let $P = \{p_1, p_2, \dots, p_n\}$ be a set of $n$ distinct labeled points in $\boldsymbol{E}^d$, and let $\omega_i$ weight associated with $p_i$, for all $i = 1, 2, \dots, n$. Let the $(d+1)$-tuple $h = (h_1, h_2, \dots, h_{d+1}) \in \boldsymbol{E}^{d+1}$ represent the hyperplane in $\boldsymbol{E}^d$ described by $\{x \in \boldsymbol{E}^d | \tilde{h} \cdot x + h_{d+1} = 0\}$, using cartesian coordinates, where $\tilde{h}$ is the non-zero normal vector $(h_1, h_2, \dots, h_d)$.

Given any $(d+1)$-tuple $z$, we will say that $\tilde{z}$ is the $d$-tuple formed by taking the first $d$ coordinates of $z$. If two $(d+1)$-tuples $\alpha$ and $\beta$ represent the same hyperplane, we shall say that $\alpha$ and $\beta$ are *equivalent*, and denote this by $\alpha \equiv \beta$. The following straightforward observation illustrates the degree of freedom in the choice of $(d+1)$-tuple to represent a given hyperplane:

**Observation 7.1** *Let $\alpha$ represent a hyperplane in $E^d$.*

1. *The $(d+1)$-tuple $\beta$ is equivalent to $\alpha$ if and only if there exists some $t \neq 0$ such that $\beta = t\alpha$, and*
2. *Given any $k > 0$, there exists $(d+1)$-tuple $\beta$ such that $\beta \equiv \alpha$ and $\|\tilde{\beta}\| = k$.*

The orthogonal Euclidean distance between a point $x \in E^d$ and a hyperplane $h$ is given by

$$\delta(x, h) = \frac{|x \cdot \tilde{h} + h_{d+1}|}{\|\tilde{h}\|}$$

(see [Bor69]). If point $x$ has weight $\omega > 0$ assigned to it, then the weighted orthogonal Euclidean distance between $x$ and $h$ is given by

$$\omega\, \delta(x, h) = \omega \frac{|x \cdot \tilde{h} + h_{d+1}|}{\|\tilde{h}\|}$$

The expression

$$\omega \frac{x \cdot \tilde{h} + h_{d+1}}{\|\tilde{h}\|}$$

has the same magnitude as $\omega\, \delta(x, h)$, but the sign depends upon which open halfspace of $h$ contains $x$.

Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of $n$ distinct labeled points in $E^d$, and let $\omega_i > 0$ be a weight associated with the point $p_i \in P$. Let the set of red points of $P$ be called $R$, and the green points be called $G$. Let $\mu(P) = \{\mu(p_1), \mu(p_2), \ldots, \mu(p_n)\}$ be the set of $(d+1)$-tuples defined by $\mu(p_i) = \{\mu_1(p_i), \mu_2(p_i), \ldots, \mu_{d+1}(p_i)\}$, where

$$\mu_j(p_i) = \begin{cases} \omega_i p_{i,j} & \text{if } p_i \in R, \\ -\omega_i p_{i,j} & \text{if } p_i \in G, \end{cases}$$

for $j = 1, 2, \ldots, d$, and $\mu_{d+1}(p_i) = \omega_i$. Clearly then, if the inequality

$$\frac{\mu(p_i) \cdot h}{\|\tilde{h}\|} > 0$$

is true for all $i = 1, 2, \ldots, n$, then $h$ is a strong strict linear separator for $R$ and $G$. Conversely, if $R$ and $G$ are strictly linearly separable, then their strong separator has some parameterization $h$ satisfying the inequality for all $i$.

The problem of finding a widest strong linear separator for $P$ is then reduced to the problem of finding a $(d+1)$-tuple $h$ satisfying

$$\text{maximize} \qquad \min_i \frac{\mu(p_i) \cdot h}{\|\tilde{h}\|}. \tag{7.1}$$

If the optimal value of this problem is negative, then the points of $P$ cannot be strongly separated by a hyperplane. Otherwise, the optimal value is the minimum weighted orthogonal distance from $h$ to the points of $P$.

We now establish a strong correspondence between (7.1) and the following convex quadratic minimization problem with $n$ constraints and $d+1$ variables:

$$\begin{array}{llcl} \text{minimize} & \|\tilde{h}\|^2 & & \\ \text{subject to} & \mu(p_i) \cdot h & \geq & 1 \end{array} \tag{7.2}$$

In [Meg84,Meg83], Megiddo has shown that convex quadratic minimization problems in $m$ variables and $n$ constraints may be solved in $O(n)$ time and space, assuming that $m$ is fixed. However, the time bound for his method has a "constant" of proportionality doubly-exponential in $m$. More recently, Clarkson [Cla86] and Dyer [Dye86] independently improved his algorithm such that the new constant of proportionality is exponential in $m^2$. Thus Problem (7.2) may be solved in $O(n)$ time and space, assuming that $d$ is fixed.

The next three lemmas describe the relationship between the formulations of (7.1) and (7.2). For their proofs, we define $\Delta(h)$ as the value of (7.1) for $h$; that is,

$$\Delta(h) = \min_i \frac{\mu(p_i) \cdot h}{\|\tilde{h}\|}.$$

Also, we observe that the origin satisfies none of the constraints of (7.2), and therefore cannot be contained in the feasibility region.

**Lemma 7.2** *Problem* (7.2) *has a feasible solution if and only if there exists some feasible solution* $\alpha$ *of* (7.1) *such that* $\Delta(\alpha) > 0$.

**Proof** $\boxed{\Longrightarrow}$ Let $\alpha$ be a solution of (7.2); that is, $\min_i \mu(p_i) \cdot \alpha \geq 1$. Then $\Delta(\alpha) = \frac{1}{\|\tilde{\alpha}\|} > 0$.

$\boxed{\Longleftarrow}$ Let $\alpha$ be a solution of (7.1) such that $\Delta(\alpha) > 0$. Observation 7.1 implies that, for any $k > 0$, there exists $(d+1)$-tuple $\beta \equiv \alpha$ such that $\|\tilde{\beta}\| = k$. Since $\Delta(\beta) = \Delta(\alpha)$, choosing $k = \frac{1}{\Delta(\alpha)}$ gives $\min_i \mu(p_i) \cdot \beta = k\Delta(\alpha) = 1$. Since $\beta$ satisfies the constraints of (7.2), Problem (7.2) is feasible. □

**Lemma 7.3** *Problem* (7.2) *has an optimal value of zero if and only if Problem* (7.1) *is unbounded.*

**Proof** $\boxed{\Longrightarrow}$ Let $\alpha$ be an optimal solution of (7.2). Then $\|\tilde{\alpha}\| = 0$.

If the feasibility region of (7.2) is contained in the line $\{x \in E^{d+1} | \|\tilde{x}\| = 0\}$, then the origin must satisfy some constraint of (7.2), which is a contradiction. Hence there exists some $\beta$ feasible for (7.2) that is not contained in this line. By convexity of the feasibility region, every $(d+1)$-tuple of the form $\gamma(t) = t\beta + (1-t)\alpha$ is feasible for (7.2), for all $t \in (0, 1]$. Since $\|\tilde{\alpha}\| = 0$, we have $\|\tilde{\gamma}(t)\| = t\|\tilde{\beta}\|$. Furthermore, $\min_i \mu(p_i) \cdot \gamma(t) \geq 1$ implies that $\Delta(\gamma(t)) \geq \frac{1}{t}$. Therefore (7.1) is unbounded.

$\boxed{\Longleftarrow}$ If (7.1) is unbounded, there must exist an infinite sequence of solutions $(\alpha_j)_{j=1}^{\infty}$ such that $\Delta(\alpha_j)$ diverges monotonically to infinity as $j \to \infty$. By Observation 7.1, each $\alpha_j$ is equivalent to some $\beta_j$ such that

$$\|\tilde{\beta}_j\| = \frac{1}{\Delta(\alpha_j)}.$$

Since $\Delta(\alpha_j) = \Delta(\beta_j)$, we have $\min_i \mu(p_i) \cdot \beta_j = 1$, and thus $\beta_j$ is a feasible solution of (7.2). Moreover, each $\beta_j$ is contained in the hyperplane $\mu(p_i) \cdot h = 1$ for some $i \in \{1, 2, \ldots, n\}$. Also, since $\|\tilde{\beta}_j\| \leq \frac{1}{\Delta(\alpha_1)}$, each of the $\beta_j$ are confined to the closed and bounded region

$$\left(\bigcup_{i=1}^{n} \{h \in E^{d+1} | \mu(p_i) \cdot h = 1\}\right) \cap \{h \in E^{d+1} | \|\tilde{h}\| \leq 1/\Delta(\alpha_1)\}.$$

Therefore the limit of $(\beta_j)_{j=1}^{\infty}$ as $j \to \infty$ is a feasible solution. Since $\lim_{j \to \infty} \|\tilde{\beta}_j\| = 0$, Problem (7.2) has optimal value zero. □

**Lemma 7.4** *Let $\alpha$ be an optimal solution for* (7.2), *with optimal value greater than zero. Then $\alpha$ is also an optimal solution for* (7.1).

**Proof** Let $K_\alpha = \frac{1}{\|\tilde{\alpha}\|}$. Since $\alpha$ is optimal for (7.2), we have

$$\min_i \mu(p_i) \cdot \alpha \geq 1, \quad \text{and} \quad \|\tilde{\alpha}\| > 0.$$

Hence $K_\alpha \leq \Delta(\alpha)$. Assume that $\alpha$ is not optimal for (7.1). Then there must exist some solution $\beta$ for (7.1) such that $\Delta(\beta) > \Delta(\alpha)$. Without loss of generality, by Observation 7.1, we can choose $\beta$ such that $\|\tilde{\beta}\| = \|\tilde{\alpha}\|$. Then

$$1 = K_\alpha \|\tilde{\alpha}\| \leq \Delta(\alpha)\|\tilde{\alpha}\| < \Delta(\beta)\|\tilde{\beta}\| = \min_i \mu(p_i) \cdot \beta;$$

that is, $1 < \min_i \mu(p_i) \cdot \beta$, thus $\beta$ is also optimal for (7.2).

Let $\beta_{\min} = \min_i \mu(p_i) \cdot \beta$ and let $\gamma = \frac{1}{\beta_{\min}}\beta$. Let $\gamma_{\min} = \min_i \mu(p_i) \cdot \gamma$. Then

$$\gamma_{\min} = \min_i \frac{\mu(p_i) \cdot \beta}{\beta_{\min}} = 1.$$

Therefore $\gamma$ is a feasible solution for (7.2). But

$$\|\tilde{\gamma}\| = \|\frac{1}{\beta_{\min}}\tilde{\beta}\| = \frac{\|\tilde{\beta}\|}{\beta_{\min}} < \|\tilde{\beta}\| = \|\tilde{\alpha}\|,$$

since $\beta_{\min} > 1$. This contradicts the optimality of $\alpha$ for (7.2). Therefore $\alpha$ is optimal for (7.1). □

These three lemmas together imply that the solution technique for convex quadratic minimization problems due to Megiddo may be applied to find widest strong strict linear separators of a set of labeled points $P$. Lemma 7.2 implies that the points of $P$ are strictly separable if and only if the minimization problem has a feasible solution. Lemma 7.3 implies that there is no upper limit on the "width" of strong separators of $P$ if and only if the minimization problem has an optimal value of zero. Recall that this situation may occur only if all points of $P$ share the same labeling.

**Theorem 7.5** *Let $P$ be a set of $n$ distinct labeled points of $\mathbf{E}^d$, for some fixed dimension $d$. Let each of the points of $P$ be associated with some positive weight. Then a widest strong strict linear separator of $P$ may be found, or its non-existence determined, in $O(n)$ time.*

## 7.3 Wide Weak Linear Separation

Consider now the case where the set of labeled, weighted points $P$ has no strong strict linear separators. The convex quadratic minimization problem (7.2) on $\mu(P)$ would then be infeasible. However, if $Q$ is a subset of $P$ for which the minimization problem is feasible on $\mu(Q)$, then the points of $Q$ are separable, and vice-versa.

One way of finding a widest weak strict linear separator of $P$ is to construct the homogeneous dual arrangement $\mathcal{A}(\rho_o(\vec{l}(P)))$ of $P$ in $E^{d+1}$, using the transformations of Section 4.2 and Algorithm MAXCOMP. As was discussed earlier, the output of the algorithm optionally includes a list of pointers to the faces of the arrangement associated with maximal strict linearly separable components of $P$. For each cell in the list, we may produce its corresponding maximal component in $O(n)$ time, and then compute the widest strong linear separator of this component, also in $O(n)$ time using the methods of the previous section. The widest weak strict linear separator of $P$ would then be simply the widest separator found over all the maximal components of $P$.

Let $R$ and $G$ be the points of $P$ labeled red and green, respectively. By Corollary 5.9, we know that an asymptotic worst-case bound on the number of maximal linearly separable components is $O(kn^{d-1}+1)$, where $k$ is the interpenetration of $R$ and $G$. The time required to find a widest separator using this approach is then $O(n^d)+O(n)O(kn^{d-1}+1) = O((k+1)n^d)$, if the problem dimensionality $d$ is treated as being fixed. As the value $k$ grows, this bound becomes more and more unsatisfactory.

In the case where the points of $P$ are unweighted, the upper bound given above may be reduced. To do this, we shall examine the facets of cells of $\mathcal{A}(\rho_o(\vec{l}(P)))$ associated with maximal components of $P$.

Let $f$ be a cell of $\mathcal{A}(\rho_o(\vec{l}(P)))$ determining a maximal component, and let $Q$ be the subset of $P$ such that the hyperplanes of $\rho(\vec{l}(Q))$ each contain a facet of $f$, and every such facet is contained in some hyperplane of $\rho(\vec{l}(Q))$. The cell $f$ must be contained in each of the open half-spaces of $\rho_o(\vec{l}(Q))$, by the following argument: let $q$ be a point of $Q$ such that the half-space $\rho_o(\vec{l}(q))$ does not contain $f$. Let $g$ be the cell sharing with $f$ the facet of $f$ contained in $\rho(\vec{l}(q))$. Then $\chi_o(g) = \chi_o(f)+1$, contradicting the assumption that $f$ determines a maximal component of $P$.

One conclusion that may be drawn from this is that the set of separators of the maximal component associated with $f$ are the same as the set of strong linear separators of $Q$. For the purposes of finding any weak separators of $P$ whose duals lie in $f$, only the points of $Q$ are significant; all other points of $P$ are redundant.

With this observation in mind, we now outline a constructive algorithm to find a wide (unweighted) weak strict linear separator of a set of $n$ labeled points $P$ in $\boldsymbol{E}^d$, for some fixed $d$. The input accepted by the algorithm is a set of $n$ labeled points $P$ in $\boldsymbol{E}^d$, and the output is a widest weak strict linear separator *widest*.

*WIDEWEAK*

(1) Let $H = \rho(\vec{l}(P))$, and let $H_> = \rho_o(\vec{l}(P))$. Set *widest* $\leftarrow$ *nil* and *width* $\leftarrow$ 0.

(2) Construct the homogeneous half-space arrangement $\mathcal{A}(H_>)$ using Algorithm MAXCOMP, producing a list $\mathcal{L}$ of pointers to cells of $\mathcal{A}(H_>)$ corresponding to maximal components of $P$, and marking every facet according to the unique hyperplane of $H$ containing it.

(3) For every cell $f$ referenced by a pointer in $\mathcal{L}$, do:

   (3a) Let $Q$ be the subset of $P$ such that $\rho(\vec{l}(Q))$ is the subset of $H$ whose hyperplanes contain facets of $Q$.

   (3b) Using the techniques of Section 7.2, set *tempwidest* to be the widest strong linear separator of $Q$, and set *tempwidth* to be the minimum (unweighted) orthogonal distance from this separator to the points of $Q$.

   (3c) If *tempwidth* > *width*, then set *widest* $\leftarrow$ *tempwidest* and *width* $\leftarrow$ *tempwidth*.

If $d$ is considered fixed, the time cost of executing Steps 3a, 3b, and 3c for a given cell $f$ is $O(|Q|)$; that is, of the order of the number of incidences between cell $f$ and facets of $\mathcal{A}(\rho_o(\vec{l}(P)))$. If these steps were performed over all cells of the arrangement, then the total time required to execute Step 3 is of the same order as the total number of incidences between cells and facets. In Section 6.2, we saw that the size of the incidence graph of a homogeneous arrangement in $E^m$ was $O(n^{m-1})$. Therefore, Step 3 requires only $O(n^d)$ total time.

The weighted and unweighted results are both summarized in the following theorem:

**Theorem 7.6** *Let $P$ be a set of $n$ distinct labeled points of $E^d$, for some fixed dimension $d$. Let $k$ be the interpenetration of the points of $P$, and let each of the points of $P$ be associated with some positive weight. Then a widest weak strict linear separator of $P$ may be found, or its non-existence determined, in $O((k+1)n^d)$ time and and $O(n^d)$ space. Furthermore, if the weights are identical, then the time required drops to $O(n^d)$.*

If the points of $P$ are in general position, the topological sweep method outlined in Section 6.5 allows the determination of the strict linear interpenetration of the red and green points of $P$ in $O(n^d)$ time, but using only $O(n)$ space. In a second pass, knowing this interpenetration value (call it $k$), each cell $f$ of $\mathcal{A}(\rho_o(\vec{l}(P)))$ associated with maximal components of $P$ may be enumerated. Unfortunately, the facets of $f$ are not available using the sweepline method, even though the maximal component corresponding to $f$ may be produced in linear time, as well as a widest strong linear separator of this component. If this is performed for every such cell $f$, we may determine a widest weak linear separator of $P$, in $O(n^{d+1})$ time. We conclude this discussion with the following theorem:

**Theorem 7.7** *Let $P$ be a set of $n$ distinct labeled points of $E^d$ in general position, for some fixed dimension $d$. Let each of the points of $P$ be associated with some positive weight. Then a widest weak strict linear separator of $P$ may be found, or its non-existence determined, in $O(n^{d+1})$ time and $O(n)$ space.*

## 7.4 Linear Separation of Hyperspheres

The problem of finding a widest strong linear separator for a set of unweighted labeled points $P$ is related to the problem of finding a strong linear separator for a set of hyperspheres of varying radii, as we shall show in this section.

Let $S = \{s_1, s_2, \ldots, s_n\}$ be a set of labeled hyperspheres in $\boldsymbol{E}^d$, where each $s_i$ is defined by

1. its center $ctr(s_i) = (c_{i1}, c_{i2}, \ldots, c_{id}) \in \boldsymbol{E}^d$, and

2. its positive radius $rad(s_i) \in \boldsymbol{R}$,

and the label of $s_i$ is either *red* or *green*. Formally,

$$s_i = \{x \in \boldsymbol{E}^d | \, \delta(x, ctr(c_i)) = rad(s_i)\}.$$

If hyperplane $h = (h_1, h_2, \ldots, h_{d+1})$ is a strong non-strict linear separator of $S$, then the minimum orthogonal distance between the hyperspheres of $S$ and $h$ is given by

$$\delta(s_i, h) = \min_i \frac{|ctr(s_i) \cdot \tilde{h} + h_{d+1}|}{\|\tilde{h}\|} \geq rad(s_i).$$

Furthermore, the sign of $ctr(s_i) \cdot \tilde{h} + h_{d+1}$ depends on whether $s_i$ is labeled red or green.

Consider the set of labeled points $P = \{p_1, p_2, \ldots, p_n\}$ with associated weights $W = \{\omega_1, \omega_2, \ldots, \omega_n\}$, such that $p_i = ctr(s_i)$, with the same label as $s_i$, and $\omega_i = \frac{1}{rad(s_i)}$, for all $i = 1, 2, \ldots, n$. Then the expression

$$\min_i \frac{\mu(p_i) \cdot h}{\|\tilde{h}\|} \geq 1$$

holds if and only if

$$\begin{aligned} \min_i \omega_i \frac{p_i \cdot \tilde{h} + h_{d+1}}{\|\tilde{h}\|} &\geq 1 \text{ if } p_i \text{ is red, and} \\ -\min_i \omega_i \frac{p_i \cdot \tilde{h} + h_{d+1}}{\|\tilde{h}\|} &\geq 1 \text{ if } p_i \text{ is green.} \end{aligned}$$

This in turn is equivalent to the following:

$$\begin{aligned} \min_i \frac{ctr(s_i) \cdot \tilde{h} + h_{d+1}}{\|\tilde{h}\|} &\geq rad(s_i) \text{ if } s_i \text{ is red, and} \\ -\min_i \frac{ctr(s_i) \cdot \tilde{h} + h_{d+1}}{\|\tilde{h}\|} &\geq rad(s_i) \text{ if } s_i \text{ is green.} \end{aligned}$$

From this we may conclude that the problem of finding a strong non-strict linear separator for $S$ is reduced to the problem of finding a $(d+1)$-tuple $h$ satisfying

$$\min_i \frac{\mu(p_i) \cdot h}{\|\tilde{h}\|} \geq 1$$

where $P$ is defined as above. Similarly, the problem of finding a strong strict linear separator for $S$ is reduced to the problem of finding a $(d+1)$-tuple satisfying the strict version of the above inequality. The problem of finding a strong linear separator of a set of labeled hyperspheres is simply a special case of problem of finding a widest strong strict linear separator of a set of labeled points. Naturally, this problem may be solved using the same techniques as in Section 7.2, which leads us to the following result:

**Theorem 7.8** *Let $S$ be a set of $n$ labeled hyperspheres of $\boldsymbol{E}^d$, for some fixed dimension $d$. Then a strong strict or non-strict linear separator of $S$ may be found, or its non-existence determined, in $O(n)$ time.*

We shall now tackle the problem of finding weak linear separators for $S$. If $h$ is a $(d+1)$-tuple representing a hyperplane of $\boldsymbol{E}^d$, by Observation 7.1, we have the freedom to restrict $\|\tilde{h}\|$ to be always equal to 1. Hence any $h$ with $\|\tilde{h}\| = 1$ satisfying

$$\mu(p_i) \cdot h \quad \geq \quad 1 \text{ for all } i = 1, 2, \ldots, n$$

corresponds to a strong non-strict linear separator of $S$, and any such $h$ also satisfying the strict inequalities is a strong strict separator.

Now, let $h$ be a $(d+1)$-tuple representing a hyperplane that is not a strong (non-strict) separator of the hyperspheres of $S$. Let $P'$ be the subset of all points $p$ of $P$ such that

$$\mu(p) \cdot h \geq 1,$$

and let $S'$ be the set of hyperspheres of $S$ corresponding to the points of $P'$. Clearly, $h$ is a strong separator of the hyperspheres of $S'$.

The constraints given above define closed (and open) half-spaces in $\boldsymbol{E}^{d+1}$. We shall denote the sets of these closed and open half-spaces by $K_{\geq}$ and $K_{>}$, respectively; the set of bounding hyperplanes of these half-spaces shall be called $K$. Consider the non-homogeneous half-space arrangement $\mathcal{A}(K_{>})$ in $\boldsymbol{E}^{d+1}$. Each face $f$ of

$\mathcal{A}(K_>)$ intersecting the hypercylinder $\{h \in E^{d+1} \mid \|\tilde{h}\| = 1\}$ corresponds to a strict and non-strict linearly separable component of the hyperspheres of $S$, depending on the number of half-spaces of $K_>$ and $K_\geq$ containing $f$, respectively. As with homogeneous arrangements, we will denote these numbers of half-spaces by $\chi_o(f)$ and $\chi_c(f)$, respectively. Hence we may obtain weak strict and non-strict linear separators of $S$ by constructing the arrangement $\mathcal{A}(K_>)$, and visiting the faces of the arrangement one by one. Of the faces intersecting the hypercylinder, those attaining the maximum value of $\chi_c$ correspond to maximal non-strict linearly separable components of $S$; those attaining the maximum value of $\chi_o$ correspond to maximal strict components of $S$.

Since we have already developed the tools that allow us to find these weak separators in the previous chapter, we will content ourselves with only a brief overview of the algorithm. The arrangement may be constructed using the original incremental construction algorithm of Edelsbrunner, O'Rourke, and Seidel [EOS86], modified to allow the calculation of differential values. When visiting the faces of the arrangement, the values of $\chi_o$ and $\chi_c$ may be computed for every face based on the differential values, as in the algorithms of the previous chapter.

To test whether the face $f$ intersects the hypercylinder, the vertices in the closure of $f$ may be examined. The most effective way of performing this is by first testing the vertices of the arrangement, then the edges, 2-faces, and so on, up until the cells are tested. The result of the test (inside, outside, or intersecting the hypercylinder) may be stored with each face for use when testing its superface. The amount of time required to test all faces in this manner is proportional to the size of the arrangement.

Once the desired face $f$ has been isolated, any point of $f$ contained in the hypercylinder may be selected as a tuple corresponding to the weak (strict or non-strict) linear separator of $S$. The worst-case time and space complexity of the algorithm is bounded by the worst-case time and space required to build the non-homogeneous arrangement in $E^{d+1}$:

**Theorem 7.9** *Let $S$ be a set of $n$ distinct labeled hyperspheres in $E^d$. A weak strict or non-strict linear separator of $S$ may be obtained in $O(n^{d+1})$ time and space.*

Unfortunately, there seems to be no elegant way to reduce the space complexity of this solution through the use of the topological sweep method. The intersection test described above for $f$ requires the knowledge of the faces in the closure of $f$ – faces which cannot be obtained during the sweep in dimensions higher than two.

# Chapter 8

# Conclusion

In this thesis, we have explored the relationship between separable subsets of point sets and homogeneous arrangements, from both algorithmic and combinatorial viewpoints, concentrating on the object class of point sets, and the separator classes of hyperplanes and hyperspheres. With the notion of weak separation, a combinatorial measure of the separability of two sets has been introduced, and the separability issue is no longer a binary one. However, there exist many questions related to this topic that are yet unanswered.

Kirchberger's theorem concerning the non-strict linear separability of finite sets is as follows:

**Theorem 8.1 (Kirchberger)** *Two finite subsets $P$ and $Q$ of $E^d$ are non-strictly linearly separable if and only if, for each set $T$ consisting of at most $2d+2$ points of $P \cup Q$, the sets $T \cap P$ and $T \cap Q$ are non-strictly linearly separable.*

An open problem that remains is whether the number of subsets required to test for non-strict separability may be reduced as in the strict case. If so, techniques different to those of Chapter 3 must be employed. Because the transformation used in the proof of Theorem 3.7 has a singularity at the origin, some way must be found to avoid it. This was accomplished for the strict case by "covering" the origin with a distinguished member set from one of the two families. A strict separator, in avoiding this set, would also avoid the origin. A non-strict separator, however, would not be constrained to avoid this singularity. This difficulty also prevents the extension of Kirchberger's non-strict separability theorem to finite families of arbitrary sets via the approach of Chapter 3.

In [Lay82], Lay states a quantitative form of Kirchberger's theorem that is closely related to wide linear separation. He defines a *slab* as a closed connected region bounded by two distinct parallel hyperplanes, and the *width* of a slab as the distance between its bounding hyperplanes.

**Theorem 8.2** (Lay) *Two non-empty compact subsets $P$ and $Q$ of $E^d$ are strictly separable by a slab of width $w > 0$ if and only if, for each set $T$ consisting of at most $d+2$ points of $P \cup Q$, there exists a slab of width $w$ that strictly separates $T \cap P$ and $T \cap Q$.*

An interesting extension of this result would be the development of quantitative versions of the theorems of Chapter 3.

Another question that is still very much open is whether the problem of finding separating surfaces other than hyperplanes and hyperspheres may be cast into the setting of hyperplane arrangements. Even relatively simple quadratic separating surfaces such as ellipses in $E^2$ seem to resist such transformations. A possible solution is to abandon the hyperplane arrangement approach in favour of arrangements of more complex surfaces. At this time, very little is known about these arrangements.

The upper bound on the number of separable components of a range of cardinalities is almost certainly not tight. The gap between this upper bound and the lower bound for $k$-sets due to Clarkson [Cla88] is yet to be closed. Also, an interesting open question is whether or not the worst-case number of $k$-sets of $n$ points is of the same order as the worst-case number of separable components of size $k$.

A very difficult question that still remains to be answered is whether there is a firm link between the linear or spherical interpenetration of two point sets $R$ and $G$, and the number of separable subsets of $R \cup G$. Naturally, the number of separable subsets rises as the interpenetration diminishes - interpenetration of zero implies that every subset is separable. The interpenetration provides a bound on the minimum and maximum cardinalities of the appropriate components of $R$ and $G$, which of course transforms into upper and lower bounds on $\chi_o$ and $\chi_c$ in the dual arrangement. Because the expressions of Corollaries 5.18 and 5.19 for the number of separable subsets are in terms of $\chi_o$ and $\chi_c$, one would hope to be able to bound the number of these subsets in terms of the interpenetration. So far, no non-trivial bounds have been found. This question is particularly interesting in light of Kirchberger's theorem and its extensions: we know that if all subsets of a certain size are separable, all the points are separable. If only some proportion of these subsets are separable, what is the size of the largest separable subset, in terms of this proportion?

An open problem of an algorithmic nature concerns the time complexity of finding weak separators. It is not clear whether the expensively-obtained information inherent in an arrangement of hyperplanes is wholly required to determine weak linear and spherical separators of point sets. It seems that a reduction in the time complexities of most of the separation algorithms presented in this thesis would entail the abandonment of the hyperplane arrangement as an algorithmic tool.

# Bibliography


[Bha88] Bhattacharya, B. K. Circular separability of planar point sets, *in Computational Morphology*, G. T. Toussaint, ed., North-Holland, Amsterdam, 1988.

[BFL89] Bárány, I., Füredi, Z., and Lovász, L. On the number of halving planes, *in* "Proc. $5^{th}$ Ann. ACM Symp. Comp. Geom., 1989", pp. 140-144.

[Bar76] Bartle, R. G. *The Elements of Real Analysis*, John Wiley and Sons, New York, 1976.

[BM76] Bondy, J. A. and Murty, U. S. R. *Graph Theory With Applications*, Maxmillan Press, Hong Kong, 1976.

[Bor69] Borsuk, K. *Multidimensional Analytic Geometry*, Polish Scientific Publishers, Warsaw, 1969.

[Bro79] Brown, K. Q. Voronoi diagrams from convex hulls, *Inform. Process. Lett.* **9**, 1979, pp. 223-228.

[Bro80] Brown, K. Q. Geometric Transforms for Fast Geometric Algorithms, Ph.D. Thesis, Rep. CMU-CS-80-101, Dept. Comp. Sci., Carnegie-Mellon Univ., Pittsburgh, PA, 1980.

[Buc43] Buck, R. C. Partition of space, *Amer. Math. Monthly* **50**, 1943, pp. 541-544.

[CP86] Chazelle, B. and Preparata, F. P. Halfspace range search: an algorithmic application of $k$-sets, *Discrete Comp. Geom.* **1**, 1986, pp. 76-90.

[CW82] Chin, F. and Wang, C. A. Optimal algorithms for the intersection and the minimum distance problems between planar polygons, Technical Report TR 82-8, Dept. of Computer Science, Univ. of Alberta, Edmonton, Alberta, Canada, Aug. 1982.

[Chv83] Chvátal, V. *Linear Programming*, W. H. Freeman and Co., New York, 1983.

[Cla86] Clarkson, K. L. Linear programming in $O(n \cdot 3^{d^2})$ time, *Inform. Process. Lett.* **22**, 1986, pp. 21-24.

[Cla88] Clarkson, K. L. Applications of random sampling in computational geometry II, *in* "Proc. $4^{th}$ Ann. ACM Symp. Comp. Geom., 1988", pp. 1-11.

[CSY87] Cole, R., Sharir, M., and Yap, C. K. On $k$-hulls and related problems, *SIAM J. Comput.* **16**, 1987, pp. 61-77.

[Cov65] Cover, T. M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, *IEEE Trans. Math. Comput.* EC-14, 1965, pp. 326-334.

[DGK63] Danzer, L., Grünbaum, B. and Klee, V. Helly's Theorem and its Relatives, in *Convexity, AMS Proc. Symp. Pure Math.* **7**, 1963, pp. 101-181.

[DK85] Dobkin, D. P. and Kirkpatrick, D. G. A linear algorithm for determining the separation of convex polyhedra, *J. Algorithms* 6(3), 1985, pp. 381-392.

[DH73] Duda, R. O. and Hart, P. E. *Pattern Classification and Scene Analysis,* Wiley-Interscience, New York, 1973.

[Dye86] Dyer, M. E. On a multidimensional search technique and its application to the Euclidean one-center problem, *SIAM J. Comput.* **15**, 1986, pp. 725-738.

[Ede82] Edelsbrunner, H. On computing the extreme distances between two convex polygons, Report No. 96, Institüt für Informationsverarbeitung, Technical Univ. of Graz, Graz, Austria, Aug 1982.

[Ede87] Edelsbrunner, H. *Algorithms in Combinatorial Geometry,* Brauer, W., Rozenberg, G., and Salomaa, A., eds., Springer-Verlag, Berlin, 1987.

[EG86] Edelsbrunner, H. and Guibas, L. J. Topologically sweeping an arrangement, *in* "Proc. $18^{th}$ Ann. ACM Symp. Theory Comput. 1986", pp. 389-403.

[EOS86] Edelsbrunner, H., O'Rourke, J. and Seidel, R. Constructing arrangements of lines and hyperplanes with applications, *SIAM J. Comput.* **15**, 1986, pp. 341-363.

[ES86] Edelsbrunner, H. and Seidel, R. Voronoi diagrams and arrangements, *Discrete Comput. Geom.* **1**(1), 1986, pp. 25-44.

[ELSS73] Erdös, P., Lovász, L., Simmons, A., and Strauss, E. G. Dissection graphs of planar point sets, in *A Survey of Combinatorial Theory*, J. N. Srivastava *et al.*, eds., North-Holland, Amsterdam, 1973, pp. 139-149.

[Eul1] Euler, L. Elementa doctrinae solidorum, *Novi Comm. Acad. Sci. Imp. Petropol.* **4**, 1752/53, pp. 109-140.

[Eul2] Euler, L. Demonstratio nonnullarum insignium proprietatum, quibus solida hedris planis inclusa sunt praedita, *Novi Comm. Acad. Sci. Imp. Petropol.* **4**, 1752/53, pp. 140-160.

[Eve72] Eves, H. W. *A Survey of Geometry*, rev. ed., Allyn and Bacon, Boston, 1972.

[Fis86] Fisk, S. Separating point sets by circles and the recognition of digital disks, *Pattern Analysis and Machine Intelligence* 8(4), 1986, pp. 554-6.

[Grü67] Grünbaum, B. *Convex Polytopes*, John Wiley and Sons, New York, 1967.

[Grü72] Grünbaum, B. *Arrangements and Spreads,* Regional. Conf. Ser. Math., Amer. Math. Soc., Providence, RI, 1972.

[Har69] Harary, F. *Graph Theory*, Addison-Wesley, Reading, MA, 1969.

[Hel23] Helly, E. Über Mengen konvexer Körper mit gemeinschaftlichen Punkten, *Jber. Deutsch. Math.-Verein.* **32**, 1923, pp. 175-6.

[Iag62] Iaglom, I. M. *Geometric Transformations*, tr. A. Shields, Random House, New York, 1962.

[Kir03] Kirchberger, P. Über Tchebychefsche Annäherungsmethoden, *Math. Ann.* **57**, 1903, pp. 509-540.

[Lay71] Lay, S. R. On separation by spherical surfaces, *Amer. Math. Monthly* **78**, 1971, pp. 1112-3.

[Lay82] Lay, S. R. *Convex Sets and Their Applications*, John Wiley and Sons, New York, 1982.

[Lee82] Lee, D. T. On $k$-nearest neighbour Voronoi diagrams in the plane, *IEEE Trans. Comput.* C-**31**, 1982, pp. 478-487.

[Lov71] Lovász, L. On the number of halving lines, *Ann. Univ. Sci. Budapest, Eötvös, Sect. Math.* **14**, 1971, pp. 107-108.

[Meg83] Meggido, N. Linear time algorithms for linear programming in $R^3$ and related problems, *SIAM J. Comput.* **12**(4), 1983, pp 759-776.

[Meg84] Meggido, N. Linear programming in linear time when the dimension is fixed, *J. Assoc. Comput. Mach.* **31**, 1984, pp 114-127.

[Nef81] Nef, W. Zur Einführung der Eulerschen Charakteristik, *Monatsh. Math.* **92**, 1981, pp. 41-46.

[Nef84] Nef, W. Ein einfacher Beweis des Satzes von Euler-Schläfli, *Elem. Math.* **39**, 1984, pp. 1-6.

[OKM86] O'Rourke, J., Kosaraju, S. R., and Megiddo, N. Computing circular separability, *Discrete Comput. Geom.* **1**(1), 1986, pp. 105-113.

[PS85] Preparata, F. P. and Shamos, M. I. *Computational Geometry*, Springer-Verlag, New York, 1985.

[RS50] Rademacher, H. and Schoenberg, I. J. Helly's theorem on convex domains and Tchebycheff's approximation problem, *Canad. J. Math.* **2**, 1950, pp. 245-256.

[Rud64] Rudin, W. *Principles of Mathematical Analysis*, McGraw-Hill, 1964.

[Sch01] Schläfli, L. Theorie der vielfachen Kontinuität, *Denkschr. Schweiz. naturf. Ges.* **38**, 1901, pp. 1-237.

[Sch81] Schwartz, J. T. Finding the minimum distance between two convex polygons, *Inform. Process. Lett.* **13**, 1981, pp. 168-170.

[SW70] Stoer, J. and Witzgall, C. *Convexity and Optimization in Finite Dimensions I*, Springer-Verlag, New York, 1970.

[Val64] Valentine, F. A. *Convex Sets,* McGraw-Hill, New York, 1964.

[Wat73] Watson, D. A Refinement of theorems of Kirchberger and Carathéodory, *J. Austral. Math. Soc.***15**, 1973, pp. 190-192.

[YKII88] Yamamoto, P., Kato, K., Imai, K. and Imai, H. Algorithms for vertical and orthogonal $L_1$ linear approximation of points, *in* "Proc. $4^{th}$ Ann. ACM Symp. Comp. Geom., 1988", pp. 352-361.