I Convexity

We begin with a review of basic geometric notions such as hyperplanes and affine subspaces in \mathbf{R}^d , and we spend some time by discussing the notion of general position. Then we consider fundamental properties of convex sets in \mathbf{R}^d , such as a theorem about the separation of disjoint convex sets by a hyperplane and Helly's theorem.

1.1 Linear and Affine Subspaces, General Position

Linear subspaces. Let \mathbf{R}^d denote the *d*-dimensional Euclidean space. The points are *d*-tuples of real numbers, $x = (x_1, x_2, \ldots, x_d)$.

The space \mathbf{R}^d is a vector space, and so we may speak of linear subspaces, linear dependence of points, linear span of a set, and so on. A linear subspace of \mathbf{R}^d is a subset closed under addition of vectors and under multiplication by real numbers. What is the geometric meaning? For instance, the linear subspaces of \mathbf{R}^2 are the origin itself, all lines passing through the origin, and the whole of \mathbf{R}^2 . In \mathbf{R}^3 , we have the origin, all lines and planes passing through the origin, and \mathbf{R}^3 .

Affine notions. An arbitrary line in \mathbb{R}^2 , say, is not a linear subspace unless

it passes through 0. General lines are what are called *affine subspaces*. An affine subspace of \mathbf{R}^d has the form x + L, where $x \in \mathbf{R}^d$ is some vector and L is a linear subspace of \mathbf{R}^d . Having defined affine subspaces, the other "affine" notions can be constructed by imitating the "linear" notions. What is the *affine hull* of a set $X \subseteq \mathbf{R}^d$? It is the intersection of all affine subspaces of \mathbf{R}^d containing X. As is well known, the linear span of a set X can be described as the set of all linear combinations of points of X. What is an *affine combination* of points $a_1, a_2, \ldots, a_n \in \mathbf{R}^d$ that would play an analogous role? To see this, we translate the whole set by $-a_n$, so that a_n becomes the origin, we make a linear combination, and we translate back by $+a_n$. This yields an expression of the form $\beta_1(a_1 - a_n) + \beta_2(a_2 - a_n) + \cdots + \beta_n(a_n - a_n) + a_n = \beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_{n-1} a_{n-1} + (1 - \beta_1 - \beta_2 - \cdots - \beta_{n-1})a_n$, where β_1, \ldots, β_n are arbitrary real numbers. Thus, an affine combination of points $a_1, \ldots, a_n \in \mathbf{R}^d$ is an expression of the form

$$\alpha_1 a_1 + \cdots + \alpha_n a_n$$
, where $\alpha_1, \ldots, \alpha_n \in \mathbf{R}$ and $\alpha_1 + \cdots + \alpha_n = 1$.

Then indeed, it is not hard to check that the affine hull of X is the set of all affine combinations of points of X.

The affine dependence of points a_1, \ldots, a_n means that one of them can be written as an affine combination of the others. This is the same as the existence of real numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$, at least one of them nonzero, such that both

$$\alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_n a_n = 0$$
 and $\alpha_1 + \alpha_2 + \dots + \alpha_n = 0$.

(Note the difference: In an affine *combination*, the α_i sum to 1, while in an affine *dependence*, they sum to 0.)

Affine dependence of a_1, \ldots, a_n is equivalent to linear dependence of the n-1 vectors $a_1-a_n, a_2-a_n, \ldots, a_{n-1}-a_n$. Therefore, the maximum possible number of affinely independent points in \mathbf{R}^d is d+1.

Another way of expressing affine dependence uses "lifting" one dimension higher. Let $b_i = (a_i, 1)$ be the vector in \mathbf{R}^{d+1} obtained by appending a new coordinate equal to 1 to a_i ; then a_1, \ldots, a_n are affinely dependent if and only if b_1, \ldots, b_n are linearly dependent. This correspondence of affine notions in \mathbf{R}^d with linear notions in \mathbf{R}^{d+1} is quite general. For example, if we identify \mathbf{R}^2 with the plane $x_3 = 1$ in \mathbf{R}^3 as in the picture,



then we obtain a bijective correspondence of the k-dimensional linear subspaces of \mathbf{R}^3 that do not lie in the plane $x_3 = 0$ with (k-1)-dimensional affine subspaces of \mathbf{R}^2 . The drawing shows a 2-dimensional linear subspace of \mathbf{R}^3 and the corresponding line in the plane $x_3 = 1$. (The same works for affine subspaces of \mathbf{R}^d and linear subspaces of \mathbf{R}^{d+1} not contained in the subspace $x_{d+1} = 0$.)

This correspondence also leads directly to extending the affine plane \mathbf{R}^2 into the *projective plane*: To the points of \mathbf{R}^2 corresponding to nonhorizontal lines through 0 in \mathbb{R}^3 we add points "at infinity," that correspond to horizontal lines through 0 in \mathbb{R}^3 . But in this book we remain in the affine space most of the time, and we do not use the projective notions.

Let $a_1, a_2, \ldots, a_{d+1}$ be points in \mathbb{R}^d , and let A be the $d \times d$ matrix with $a_i - a_{d+1}$ as the *i*th column, $i = 1, 2, \ldots, d$. Then a_1, \ldots, a_{d+1} are affinely independent if and only if A has d linearly independent columns, and this is equivalent to $\det(A) \neq 0$. We have a useful criterion of affine independence using a determinant.

Affine subspaces of \mathbf{R}^d of certain dimensions have special names. A (d-1)dimensional affine subspace of \mathbf{R}^d is called a *hyperplane* (while the word *plane* usually means a 2-dimensional subspace of \mathbf{R}^d for any d). One-dimensional subspaces are lines, and a k-dimensional affine subspace is often called a kflat.

A hyperplane is usually specified by a single linear equation of the form $a_1x_1 + a_2x_2 + \cdots + a_dx_d = b$. We usually write the left-hand side as the scalar product $\langle a, x \rangle$. So a hyperplane can be expressed as the set $\{x \in \mathbf{R}^d : \langle a, x \rangle = b\}$ where $a \in \mathbf{R}^d \setminus \{0\}$ and $b \in \mathbf{R}$. A (closed) half-space in \mathbf{R}^d is a set of the form $\{x \in \mathbf{R}^d : \langle a, x \rangle \ge b\}$ for some $a \in \mathbf{R}^d \setminus \{0\}$; the hyperplane $\{x \in \mathbf{R}^d : \langle a, x \rangle \ge b\}$ is its boundary.

General k-flats can be given either as intersections of hyperplanes or as affine images of \mathbf{R}^k (parametric expression). In the first case, an intersection of k hyperplanes can also be viewed as a solution to a system Ax = b of linear equations, where $x \in \mathbf{R}^d$ is regarded as a column vector, A is a $k \times d$ matrix, and $b \in \mathbf{R}^k$. (As a rule, in formulas involving matrices, we interpret points of \mathbf{R}^d as column vectors.)

An affine mapping $f: \mathbb{R}^k \to \mathbb{R}^d$ has the form $f: y \mapsto By + c$ for some $d \times k$ matrix B and some $c \in \mathbb{R}^d$, so it is a composition of a linear map with a translation. The image of f is a k'-flat for some $k' \leq \min(k, d)$. This k' equals the rank of the matrix B.

General position. "We assume that the points (lines, hyperplanes,...) are in general position." This magical phrase appears in many proofs. Intuitively, general position means that no "unlikely coincidences" happen in the considered configuration. For example, if 3 points are chosen in the plane without any special intention, "randomly," they are unlikely to lie on a common line. For a planar point set in general position, we always require that no three of its points be collinear. For points in \mathbb{R}^d in general position, we assume similarly that no unnecessary affine dependencies exist: No $k \leq d+1$ points lie in a common (k-2)-flat. For lines in the plane in general position, we postulate that no 3 lines have a common point and no 2 are parallel. The precise meaning of general position is not fully standard: It may depend on the particular context, and to the usual conditions mentioned above we sometimes add others where convenient. For example, for a planar point set in general position we can also suppose that no two points have the same x-coordinate.

What conditions are suitable for including into a "general position" assumption? In other words, what can be considered as an unlikely coincidence? For example, let X be an n-point set in the plane, and let the coordinates of the *i*th point be (x_i, y_i) . Then the vector $v(X) = (x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)$ can be regarded as a point of \mathbb{R}^{2n} . For a configuration X in which $x_1 = x_2$, i.e., the first and second points have the same x-coordinate, the point v(X)lies on the hyperplane $\{x_1 = x_2\}$ in \mathbb{R}^{2n} . The configurations X where some two points share the x-coordinate thus correspond to the union of $\binom{n}{2}$ hyperplanes in \mathbb{R}^{2n} . Since a hyperplane in \mathbb{R}^{2n} has (2*n*-dimensional) measure zero, almost all points of \mathbb{R}^{2n} correspond to planar configurations X with all the points having distinct x-coordinates. In particular, if X is any n-point planar configuration and $\varepsilon > 0$ is any given real number, then there is a configuration X', obtained from X by moving each point by distance at most ε , such that all points of X' have distinct x-coordinates. Not only that: Almost all small movements (*perturbations*) of X result in X' with this property.

This is the key property of general position: Configurations in general position lie arbitrarily close to any given configuration (and they abound in any small neighborhood of any given configuration). Here is a fairly general type of condition with this property. Suppose that a configuration Xis specified by a vector $t = (t_1, t_2, \ldots, t_m)$ of m real numbers (coordinates). The objects of X can be points in \mathbf{R}^d , in which case m = dn and the t_i are the coordinates of the points, but they can also be circles in the plane, with m = 3n and the t_j expressing the center and the radius of each circle, and so on. The general position condition we can put on the configuration X is $p(t) = p(t_1, t_2, \ldots, t_m) \neq 0$, where p is some nonzero polynomial in m variables. Here we use the following well-known fact (a consequence of Sard's theorem; see, e.g., Bredon [Bre93], Appendix C): For any nonzero m-variate polynomial $p(t_1, \ldots, t_m)$, the zero set $\{t \in \mathbf{R}^m : p(t) = 0\}$ has measure 0 in \mathbf{R}^{m} .

Therefore, almost all configurations X satisfy $p(t) \neq 0$. So any condition that can be expressed as $p(t) \neq 0$ for a certain polynomial p in m real variables, or, more generally, as $p_1(t) \neq 0$ or $p_2(t) \neq 0$ or ..., for finitely or countably many polynomials p_1, p_2, \ldots , can be included in a general position assumption.

For example, let X be an n-point set in \mathbb{R}^d , and let us consider the condition "no d+1 points of X lie in a common hyperplane." In other words, no d+1 points should be affinely dependent. As we know, the affine dependence of d+1 points means that a suitable $d \times d$ determinant equals 0. This determinant is a polynomial (of degree d) in the coordinates of these d+1 points. Introducing one polynomial for every (d+1)-tuple of the points, we obtain $\binom{n}{d+1}$ polynomials such that at least one of them is 0 for any configuration X with d+1 points in a common hyperplane. Other usual conditions for general position can be expressed similarly.

In many proofs, assuming general position simplifies matters considerably. But what do we do with configurations X_0 that are not in general position? We have to argue, somehow, that if the statement being proved is valid for configurations X arbitrarily close to our X_0 , then it must be valid for X_0 itself, too. Such proofs, usually called *perturbation arguments*, are often rather simple, and almost always somewhat boring. But sometimes they can be tricky, and one should not underestimate them, no matter how tempting this may be. A nontrivial example will be demonstrated in Section 5.5 (Lemma 5.5.4).

Exercises

- 1. Verify that the affine hull of a set $X \subseteq \mathbf{R}^d$ equals the set of all affine combinations of points of X. \square
- 2. Let A be a 2×3 matrix and let $b \in \mathbb{R}^2$. Interpret the solution of the system Ax = b geometrically (in most cases, as an intersection of two planes) and discuss the possible cases in algebraic and geometric terms. 2
- 3. (a) What are the possible intersections of two (2-dimensional) planes in \mathbb{R}^4 ? What is the "typical" case (general position)? What about two hyperplanes in \mathbf{R}^4 ? 3 (b) Objects in \mathbb{R}^4 can sometimes be "visualized" as objects in \mathbb{R}^3 moving

in time (so time is interpreted as the fourth coordinate). Try to visualize the intersection of two planes in \mathbb{R}^4 discussed (a) in this way.

1.2 Convex Sets, Convex Combinations, Separation

Intuitively, a set is convex if its surface has no "dips":

not allowed in a convex set

5



1.2.1 Definition (Convex set). A set $C \subseteq \mathbb{R}^d$ is convex if for every two points $x, y \in C$ the whole segment xy is also contained in C. In other words, for every $t \in [0, 1]$, the point tx + (1 - t)y belongs to C.

The intersection of an arbitrary family of convex sets is obviously convex. So we can define the *convex hull* of a set $X \subseteq \mathbf{R}^d$, denoted by $\operatorname{conv}(X)$, as the intersection of all convex sets in \mathbf{R}^d containing X. Here is a planar example with a finite X:



An alternative description of the convex hull can be given using convex combinations.

1.2.2 Claim. A point x belongs to conv(X) if and only if there exist points $x_1, x_2, \ldots, x_n \in X$ and nonnegative real numbers t_1, t_2, \ldots, t_n with $\sum_{i=1}^n t_i = 1$ such that $x = \sum_{i=1}^n t_i x_i$.

The expression $\sum_{i=1}^{n} t_i x_i$ as in the claim is called a *convex combination* of the points x_1, x_2, \ldots, x_n . (Compare this with the definitions of linear and affine combinations.)

Sketch of proof. Each convex combination of points of X must lie in conv(X): For n = 2 this is by definition, and for larger n by induction. Conversely, the set of all convex combinations obviously contains X, and it is convex.

In \mathbb{R}^d , it is sufficient to consider convex combinations involving at most d+1 points:

1.2.3 Theorem (Carathéodory's theorem). Let $X \subseteq \mathbb{R}^d$. Then each point of $\operatorname{conv}(X)$ is a convex combination of at most d+1 points of X.

For example, in the plane, conv(X) is the union of all triangles with vertices at points of X. The proof of the theorem is left as an exercise to the subsequent section.

A basic result about convex sets is the separability of disjoint convex sets by a hyperplane.

1.2.4 Theorem (Separation theorem). Let $C, D \subseteq \mathbf{R}^d$ be convex sets with $C \cap D = \emptyset$. Then there exists a hyperplane h such that C lies in one of the closed half-spaces determined by h, and D lies in the opposite closed half-space. In other words, there exist a unit vector $a \in \mathbf{R}^d$ and a number $b \in \mathbf{R}$ such that for all $x \in C$ we have $\langle a, x \rangle \geq b$, and for all $x \in D$ we have $\langle a, x \rangle \leq b$.

If C and D are closed and at least one of them is bounded, they can be separated strictly; in such a way that $C \cap h = D \cap h = \emptyset$.

In particular, a closed convex set can be strictly separated from a point.

This implies that the convex hull of a closed set X equals the intersection of all closed half-spaces containing X.

Sketch of proof. First assume that C and D are compact (i.e., closed and bounded). Then the Cartesian product $C \times D$ is a compact space, too, and the distance function $(x, y) \mapsto ||x - y||$ attains its minimum on $C \times D$. That is, there exist points $p \in C$ and $q \in D$ such that the distance of C and D equals the distance of p and q.

The desired separating hyperplane h can be taken as the one perpendicular to the segment pq and passing through its midpoint:



It is easy to check that h indeed avoids both C and D.

If D is compact and C closed, we can intersect C with a large ball and get a compact set C'. If the ball is sufficiently large, then C and C' have the same distance to D. So the distance of C and D is attained at some $p \in C'$ and $q \in D$, and we can use the previous argument.

For arbitrary disjoint convex sets C and D, we choose a sequence $C_1 \subseteq$ $C_2 \subseteq C_3 \subseteq \cdots$ of compact convex subsets of C with $\bigcup_{n=1}^{\infty} C_n = C$. For example, assuming that $0 \in C$, we can let C_n be the intersection of the closure of $(1-\frac{1}{n})C$ with the ball of radius *n* centered at 0. A similar sequence $D_1 \subseteq D_2 \subseteq \cdots$ is chosen for D, and we let $h_n = \{x \in \mathbf{R}^d : \langle a_n, x \rangle = b_n\}$ be a hyperplane separating C_n from D_n , where a_n is a unit vector and $b_n \in \mathbf{R}$. The sequence $(b_n)_{n=1}^{\infty}$ is bounded, and by compactness, the sequence of (d+1)component vectors $(a_n, b_n) \in \mathbf{R}^{d+1}$ has a cluster point (a, b). One can verify, by contradiction, that the hyperplane $h = \{x \in \mathbf{R}^d : \langle a, x \rangle = b\}$ separates C and D (nonstrictly).

The importance of the separation theorem is documented by its presence in several branches of mathematics in various disguises. Its home territory is probably functional analysis, where it is formulated and proved for infinitedimensional spaces; essentially it is the so-called Hahn–Banach theorem. The usual functional-analytic proof is different from the one we gave, and in a way it is more elegant and conceptual. The proof sketched above uses more special properties of \mathbf{R}^d , but it is quite short and intuitive in the case of compact C and D.

Connection to linear programming. A basic result in the theory of linear programming is the Farkas lemma. It is a special case of the duality of linear programming (discussed in Section 10.1) as well as the key step in its proof.

1.2.5 Lemma (Farkas lemma, one of many versions). For every $d \times n$ real matrix A, exactly one of the following cases occurs:

(i) The system of linear equations Ax = 0 has a nontrivial nonnegative solution $x \in \mathbf{R}^n$ (all components of x are nonnegative and at least one of them is strictly positive).

(ii) There exists a $y \in \mathbf{R}^d$ such that $y^T A$ is a vector with all entries strictly negative. Thus, if we multiply the *j*th equation in the system Ax = 0 by y_j and add these equations together, we obtain an equation that obviously has no nontrivial nonnegative solution, since all the coefficients on the left-hand sides are strictly negative, while the right-hand side is 0.

Proof. Let us see why this is yet another version of the separation theorem. Let $V \subset \mathbf{R}^d$ be the set of *n* points given by the column vectors of the matrix *A*. We distinguish two cases: Either $0 \in \operatorname{conv}(V)$ or $0 \notin \operatorname{conv}(V)$.

In the former case, we know that 0 is a convex combination of the points of V, and the coefficients of this convex combination determine a nontrivial nonnegative solution to Ax = 0.

In the latter case, there exists a hyperplane strictly separating V from 0, i.e., a unit vector $y \in \mathbf{R}^d$ such that $\langle y, v \rangle < \langle y, 0 \rangle = 0$ for each $v \in V$. This is just the y from the second alternative in the Farkas lemma.

Bibliography and remarks. Most of the material in this chapter is quite old and can be found in many surveys and textbooks. Providing historical accounts of such well-covered areas is not among the goals of this book, and so we mention only a few references for the specific results discussed in the text and add some remarks concerning related results.

The concept of convexity and the rudiments of convex geometry have been around since antiquity. The initial chapter of the *Handbook* of Convex Geometry [GW93] succinctly describes the history, and the handbook can be recommended as the basic source on questions related to convexity, although knowledge has progressed significantly since its publication.

For an introduction to functional analysis, including the Hahn-Banach theorem, see Rudin [Rud91], for example. The Farkas lemma originated in [Far94] (nineteenth century!). More on the history of the duality of linear programming can be found, e.g., in Schrijver's book [Sch86].

As for the origins, generalizations, and applications of Carathéodory's theorem, as well as of Radon's lemma and Helly's theorem discussed in the subsequent sections, a recommendable survey is Eckhoff [Eck93], and an older well-known source is Danzer, Grünbaum, and Klee [DGK63].

Carathéodory's theorem comes from the paper [Car07], concerning power series and harmonic analysis. A somewhat similar theorem, due to Steinitz [Ste16], asserts that if x lies in the interior of conv(X)for an $X \subseteq \mathbb{R}^d$, then it also lies in the interior of conv(Y) for some $Y \subseteq X$ with $|Y| \leq 2d$. Bonnice and Klee [BK63] proved a common generalization of both these theorems: Any k-interior point of X is a k-interior point of Y for some $Y \subseteq X$ with at most max(2k, d+1) points, where x is called a k-interior point of X if it lies in the relative interior of the convex hull of some k+1 affinely independent points of X.

Exercises

- 1. Give a detailed proof of Claim 1.2.2. $\boxed{2}$
- 2. Write down a detailed proof of the separation theorem. \exists
- 3. Find an example of two disjoint closed convex sets in the plane that are not strictly separable. 1
- 4. Let f: R^d → R^k be an affine map.
 (a) Prove that if C ⊆ R^d is convex, then f(C) is convex as well. Is the preimage of a convex set always convex? 2
 - (b) For $X \subseteq \mathbf{R}^d$ arbitrary, prove that $\operatorname{conv}(f(X)) = \operatorname{conv}(f(X))$.
- 5. Let $X \subseteq \mathbf{R}^d$. Prove that diam $(\operatorname{conv}(X)) = \operatorname{diam}(X)$, where the diameter diam(Y) of a set Y is $\sup\{||x y||: x, y \in Y\}$. 3
- 6. A set $C \subseteq \mathbf{R}^d$ is a *convex cone* if it is convex and for each $x \in C$, the ray $\overrightarrow{0x}$ is fully contained in C.

(a) Analogously to the convex and affine hulls, define the appropriate "conic hull" and the corresponding notion of "combination" (analogous to the convex and affine combinations). 3

(b) Let C be a convex cone in \mathbb{R}^d and $b \notin C$ a point. Prove that there exists a vector a with $\langle a, x \rangle \geq 0$ for all $x \in C$ and $\langle a, b \rangle < 0$.

7. (Variations on the Farkas lemma) Let A be a d×n matrix and let b ∈ R^d.
(a) Prove that the system Ax = b has a nonnegative solution x ∈ Rⁿ if and only if every y ∈ R^d satisfying y^TA ≥ 0 also satisfies y^Tb ≥ 0. 3
(b) Prove that the system of inequalities Ax ≤ b has a nonnegative

(b) Prove that the system of inequalities $Ax \leq b$ has a nonnegative solution x if and only if every nonnegative $y \in \mathbf{R}^d$ with $y^T A \geq 0$ also satisfies $y^T b \geq 0$. 3

8. (a) Let $C \subset \mathbf{R}^d$ be a compact convex set with a nonempty interior, and let $p \in C$ be an interior point. Show that there exists a line ℓ passing through p such that the segment $\ell \cap C$ is at least as long as any segment parallel to ℓ and contained in C.

(b) Show that (a) may fail for C compact but not convex.

1.3 Radon's Lemma and Helly's Theorem

Carathéodory's theorem from the previous section, together with Radon's lemma and Helly's theorem presented here, are three basic properties of convexity in \mathbf{R}^d involving the dimension. We begin with Radon's lemma.

1.3.1 Theorem (Radon's lemma). Let A be a set of d+2 points in \mathbb{R}^d . Then there exist two disjoint subsets $A_1, A_2 \subset A$ such that

 $\operatorname{conv}(A_1) \cap \operatorname{conv}(A_2) \neq \emptyset.$

A point $x \in \operatorname{conv}(A_1) \cap \operatorname{conv}(A_2)$, where A_1 and A_2 are as in the theorem, is called a *Radon point* of A, and the pair (A_1, A_2) is called a *Radon partition* of A (it is easily seen that we can require $A_1 \cup A_2 = A$).

Here are two possible cases in the plane:



Proof. Let $A = \{a_1, a_2, \ldots, a_{d+2}\}$. These d+2 points are necessarily affinely dependent. That is, there exist real numbers $\alpha_1, \ldots, \alpha_{d+2}$, not all of them 0, such that $\sum_{i=1}^{d+2} \alpha_i = 0$ and $\sum_{i=1}^{d+2} \alpha_i a_i = 0$.

Set $P = \{i: \alpha_i > 0\}$ and $N = \{i: \alpha_i < 0\}$. Both P and N are nonempty. We claim that P and N determine the desired subsets. Let us put $A_1 = \{a_i: i \in P\}$ and $A_2 = \{a_i: i \in N\}$. We are going to exhibit a point x that is contained in the convex hulls of both these sets.

Put $S = \sum_{i \in P} \alpha_i$; we also have $S = -\sum_{i \in N} \alpha_i$. Then we define

$$x = \sum_{i \in P} \frac{\alpha_i}{S} a_i.$$
(1.1)

Since $\sum_{i=1}^{d+2} \alpha_i a_i = 0 = \sum_{i \in P} \alpha_i a_i + \sum_{i \in N} \alpha_i a_i$, we also have

$$x = \sum_{i \in N} \frac{-\alpha_i}{S} a_i \,. \tag{1.2}$$

The coefficients of the a_i in (1.1) are nonnegative and sum to 1, so x is a convex combination of points of A_1 . Similarly, (1.2) expresses x as a convex combination of points of A_2 .

Helly's theorem is one of the most famous results of a combinatorial nature about convex sets.

1.3.2 Theorem (Helly's theorem). Let C_1, C_2, \ldots, C_n be convex sets in \mathbb{R}^d , $n \ge d+1$. Suppose that the intersection of every d+1 of these sets is nonempty. Then the intersection of all the C_i is nonempty.

The first nontrivial case states that if every 3 among 4 convex sets in the plane intersect, then there is a point common to all 4 sets. This can be proved by an elementary geometric argument, perhaps distinguishing a few cases, and the reader may want to try to find a proof before reading further. In a contrapositive form, Helly's theorem guarantees that whenever C_1, C_2, \ldots, C_n are convex sets with $\bigcap_{i=1}^n C_i = \emptyset$, then this is witnessed by some at most d+1 sets with empty intersection among the C_i . In this way, many proofs are greatly simplified, since in planar problems, say, one can deal with 3 convex sets instead of an arbitrary number, as is amply illustrated in the exercises below. It is very tempting and quite usual to formulate Helly's theorem as follows: "If every d+1 among n convex sets in \mathbb{R}^d intersect, then all the sets intersect." But, strictly speaking, this is false, for a trivial reason: For $d \ge 2$, the assumption as stated here is met by n = 2 disjoint convex sets.

Proof of Helly's theorem. (Using Radon's lemma.) For a fixed d, we proceed by induction on n. The case n = d+1 is clear, so we suppose that $n \ge d+2$ and that the statement of Helly's theorem holds for smaller n. Actually, n = d+2 is the crucial case; the result for larger n follows at once by a simple induction.

Consider sets C_1, C_2, \ldots, C_n satisfying the assumptions. If we leave out any one of these sets, the remaining sets have a nonempty intersection by the inductive assumption. Let us fix a point $a_i \in \bigcap_{j \neq i} C_j$ and consider the points $a_1, a_2, \ldots, a_{d+2}$. By Radon's lemma, there exist disjoint index sets $I_1, I_2 \subset \{1, 2, \ldots, d+2\}$ such that

$$\operatorname{conv}(\{a_i: i \in I_1\}) \cap \operatorname{conv}(\{a_i: i \in I_2\}) \neq \emptyset.$$

We pick a point x in this intersection. The following picture illustrates the case d = 2 and n = 4:



We claim that x lies in the intersection of all the C_i . Consider some $i \in \{1, 2, ..., n\}$; then $i \notin I_1$ or $i \notin I_2$. In the former case, each a_j with $j \in I_1$ lies in C_i , and so $x \in \text{conv}(\{a_j: j \in I_1\}) \subseteq C_i$. For $i \notin I_2$ we similarly conclude that $x \in \text{conv}(\{a_j: j \in I_2\}) \subseteq C_i$. Therefore, $x \in \bigcap_{i=1}^n C_i$.

An infinite version of Helly's theorem. If we have an infinite collection of convex sets in \mathbb{R}^d such that any d+1 of them have a common point, the

entire collection still need not have a common point. Two examples in \mathbb{R}^1 are the families of intervals $\{(0, 1/n): n = 1, 2, ...\}$ and $\{[n, \infty): n = 1, 2, ...\}$. The sets in the first example are not closed, and the second example uses unbounded sets. For *compact* (i.e., closed and bounded) sets, the theorem holds:

1.3.3 Theorem (Infinite version of Helly's theorem). Let C be an arbitrary infinite family of compact convex sets in \mathbb{R}^d such that any d+1 of the sets have a nonempty intersection. Then all the sets of C have a nonempty intersection.