A Course in Convexity

Jacob Aguirre *Email: aguirre@gatech.edu*
Last updated: November 6, 2023

1 Introduction and Preliminaries

Convex Sets at Large

1.1 Convex sets. Main definitions, and some interesting examples and problems

First, we will set the stage where the action is "taking place". Much of the action, though definitely not all, happens in Euclidean space $\mathbb{R}^d$.

Definition 1.1. The $d-$dimensional Euclidean space $\mathbb{R}^d$ consists of all $d-$tuples $x = (\xi_1, \ldots, \xi_d)$ of real numbers. We call an element of $\mathbb{R}^d$ a *vector* (or more often) a *point*. We can add points, for example

$$z = x + y \quad \text{for } x = (\xi_1, \ldots, \xi_d), \quad y = (\eta_1, \ldots, \eta_d) \quad \text{and } z = (\zeta_1, \ldots, \zeta_d)\,,$$

provided

$$\zeta_i = \xi_i + \eta_i \quad \text{for } i = 1, \ldots, d.$$

We can multiply a point by a real number

$$\text{if } x = (\xi_1, \ldots, \xi_d) \quad \text{and} \quad \alpha \text{ is a real number}\,,$$

then

$$\alpha x = (\alpha\xi_1, \ldots, \alpha\xi_d)$$

is a point from $\mathbb{R}^d$. We consider the *scalar product* in $\mathbb{R}^d$ as

$$\langle x, y\rangle = \xi_1\eta_1 + \ldots + \xi_d\eta_d\,, \qquad \text{where} \quad x = (\xi_1, \ldots, \xi_d) \quad \text{and} \quad y = (\eta_1, \ldots, \eta_d).$$

We define the (Euclidean) norm

$$\|x\| = \sqrt{\xi_1^2 + \ldots + \xi_d^2}$$

of a point $x = (\xi_1, \ldots, \xi_d)$ and the *distance* between two points x and y

$$d(x, y) = \|x - y\| \quad \text{for } x, y \in \mathbb{R}^d$$

We denote the volume of a set $A \subset \mathbb{R}^d$ as $vol_d A$.

1.2 Convex sets, convex combinations, and convex hulls

Let $\{x_1, \ldots, x_m\}$ be a finite set of points from $\mathbb{R}^d$. A point

$$x = \sum_{i=1}^{m} \alpha_i x_i, \quad \text{where } \sum_{i=1}^{m} \alpha_i = 1 \quad \text{and } \alpha_i \geq 0 \text{ for } i = 1, \ldots, m$$

is called a *convex combination* of $x_1, \ldots, x_m$. Given two distinct points $x, y \in \mathbb{R}^d$, the set

$$[x, y] = \Big\{ \alpha x + (1 - \alpha) y : \ 0 \leq \alpha \leq 1 \Big\}$$

of all convex combinations of x and y is called the *interval* with endpoints x and y. A set $A \subset \mathbb{R}^d$ is called *convex*, provided $[x, y] \subset A$ for any two $x, y \in A$, or in words: a set is convex if and only if for every two points it contains the interval that connects them. Obviously, the empty set $\emptyset$ is convex. For $A \subset \mathbb{R}^d$, the set of all convex combinations of points from A is called the *convex hull* of A and denoted $\operatorname{conv}(A)$. We will see that $\operatorname{conv}(A)$ is the smallest convex set containing A. We will show this point later. Now as an early peovlem, we consider the proving the following theorem.

Schur-Horn Theorem For an $n \times n$ real symmetirc matrix $A = (\alpha_{ij})$, let $diag(A) = (\alpha_{11}, \ldots, \alpha_{nn})$ be the diagonal of A, considered as a vector from $\mathbb{R}^n$. Let us fix real numbers $\lambda_1, \ldots, \lambda_n$. Consider the set $X \subset \mathbb{R}^n$ of all diagonals of $n \times n$ real symmetric matrices with the eigenvalues $\lambda_1, \ldots, \lambda_n$. Prove that X is a convex set. Furthermore, let $l = (\lambda_1, \ldots, \lambda_n)$ be the vector of eigenvalues, so $l \in \mathbb{R}^n$. For a permutation σ of the set $\{1, \ldots, n\}$, let $l^\sigma = (\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(n)})$ be the vector with the permuted coordinates. Prove that

$$X = \operatorname{conv}\Big(l^\sigma : \sigma \quad \text{ranges over all } n! \text{ permutations of the set } \{1, \ldots, n\} \Big).$$

1.3 Convex sets in vector spaces.

We recall that a set V with the operations $+, \cdot$ is called a real vector space provided the following eight axioms are satisfied,

1. $u + v = v + u \quad$ for any two $u, v \in V$
2. $u + (v + w) = (u + v) + w \quad$ for any three $u, v, w \in V$
3. $(\alpha\beta)v = \alpha(\beta v)$ for any $v \in V$ and any $\alpha, \beta \in \mathbb{R}$
4. $1v = u$ for any $v \in V$
5. $(\alpha + \beta)v = \alpha v + \beta v$ for any $v \in V$ and any $\alpha, \beta \in \mathbb{R}$
6. $\alpha(v + u) = \alpha v + \alpha u$ for any $\alpha \in \mathbb{R}$ and any $u, v \in V$
7. there exists a *zero vector* $\mathbf{0} \in V$ such that $v + \mathbf{0} = v$ for each $v \in V$

8. for each $v \in V$ there exists a vector $-v \in V$ such that $v + (-v) = \mathbf{0}$.

Operations with convex sets. Let V be a vector space and let $A, B \subset V$ be (convex) sets. The *Minkowski sum* $A + B$ is a subset in V defined by

$$A + B = \Big\{ x + y : \quad x \in A, y \in B \Big\}.$$

In particular if $B = \{b\}$ is a point, the set

$$A + b = \Big\{ x + b : x \in A \Big\}$$

is a *translation* of A. For any number α and a subset $X \subset V$, the set

$$\alpha X = \Big\{ \alpha x : x \in X \Big\}$$

is called a *scaling* of X. This is quite an obvious property of convex sets, but many are not so obvious.

1.4 Properties of the Convex Hull. Carathéodory's Theorem

Theorem 2.1 *Let V be a vector space and let $S \subset V$ be a set. Then the convex hull of S is a convex set and any convex set containing S also contains conv(S). Thus, conv(S) is the smallest convex set containing S.*

Proof: First, we'll prove that conv(S) is a convex set. Let us choose two convex combinations, $u = \alpha_1 u_1 + \ldots + \alpha_m u_m$ and $v = \beta_1 v_1 + \ldots + \beta_n v_n$ of points from S. The interval $[u, v]$ consists of the points $\gamma u + (1 - \gamma) v$ for $0 \leq \gamma \leq 1$. Each such point $\gamma \alpha_1 u_1 + \ldots + \gamma \alpha_m u_m + (1 - \gamma) \beta_1 v_1 + \ldots + (1 - \gamma) \beta_n v_n$ is a convex combination of points $u_1, \ldots, u_m, v_1, \ldots, v_n$ from S since

$$\sum_{i=1}^{m} \gamma \alpha_i + \sum_{i=1}^{n} (1 - \gamma) \beta_i = \gamma \sum_{i=1}^{m} \alpha_i + (1 - \gamma) \sum_{i=1}^{n} \beta_i = \gamma + (1 - \gamma) = 1.$$

Thus, conv(S) is convex.

Now, we prove that for any convex set A such that $S \subset A$, we have conv$(S) \subset A$. Take a convex combination

$$u = \alpha_1 u_1 + \ldots + \alpha_m u_m$$

of points $u_1, \ldots, u_m$ from S. We must prove $u \in A$. Without loss of generality, we may assume that $\alpha_i > 0$ for $i = 1, \ldots, m$. Proceeding by induction on m, if $m = 1$, then $u = u_1$ and $u \in A$ since $S \subset A$. Suppose that $m > 1$. Then $\alpha_m < 1$ and we may write

$$u = (1 - \alpha_m) w + \alpha_m u_m, \qquad \text{where } w = \frac{\alpha_1}{1 - \alpha_m} u_1 + \ldots + \frac{\alpha_{m-1}}{1 - \alpha_m} u_{m-1}.$$

Now, we know w is a convex combination of $u_1, \ldots, u_{m-1}$ because

$$\sum_{i=1}^{m-1} \frac{\alpha_i}{1-\alpha_m} = \frac{1}{1-\alpha_m} \sum_{i=1}^{m-1} \alpha_i = \frac{1-\alpha_m}{1-\alpha_m} = 1.$$

Therefore, by the induction hypothesis, we have $w \in A$. Since A is convex, $[w, u_m] \in A$ so $u \in A$.

□

2.2 Definitions. The convex hull of a finite set of points in $\mathbb{R}^d$ is called a *polytope.*

Let $c_1, \ldots, c_m$ be vectors from $\mathbb{R}^d$ and let $\beta_1, \ldots, \beta_m$ be numbers. The set

$$P = \left\{ x \in \mathbb{R}^d : \langle c_i, x \rangle \leq \beta_i \quad \text{for } i = 1, \ldots, m \right\}$$

is called a *polyhedron.*

Examples:

1. We consider the set

$$\Delta = \left\{ (\xi_1, \ldots, \xi_{d+1}) \in \mathbb{R}^{d+1} : \quad \xi_1 + \ldots + \xi_{d+1} = 1 \text{ and } \xi_i \geq 0 \quad \text{for } i = 1, \ldots, d+1 \right\}$$

   as a polytope in $\mathbb{R}^{d+1}$. This polytope is called the *standard $d-$dimensional simplex.*

2. We consider the set

$$I = \left\{ (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d : \quad 0 \leq \xi_i \leq 1 \quad \text{for } i = 1, \ldots, d \right\}$$

   this set is commonly known as the *$d-$dimensional cube.*

3. Consider the set

$$\Theta = \left\{ (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d : \quad |\xi_1| + \ldots + |\xi_d| \leq 1 \right\}$$

   as the *(hyper)octahedron or crosspolytope.*

   We leave proving these statements to be true as an exercise to the reader.

Now, it might seem intuitively obvious that in the space of a small dimension, to represent a given point x from the convex hull of a set A as a convex combination, we would need to use only a few points of A, although the choie will, of course, depend on x. This fact is well known as Carathéodory's Theorem, which we prove below.

2.3 Carathéodory's Theorem. *Let $S \subset \mathbb{R}^d$ be a set. Then every point $x \in$conv(S) can be represented as a convex combination of $d+1$ points from S:*

$$x = \alpha_1 y_1 + \ldots + \alpha_{d+1} y_{d+1}, \quad \text{where } \sum_{i=1}^{d+1} \alpha_i = 1, \quad \alpha_i \geq 0$$

and $y_i \in S$ for $i = 1, \ldots, d+1$.

Proof: Consider that every point $x \in$conv(S) can be written as a convex combination

$$x = \alpha_1 y_1 + \ldots + \alpha_m y_m$$

of some points $y_1, \ldots, y_m \in S$. We can assume that $\alpha_i > 0$ for all $i = 1, \ldots, m$. If $m < d+1$, we can add terms $0y_1$, say, to get a convex combination with $d+1$ terms. Suppose that $m > d+1$. Then we can construct a convex combination with fewer terms. Consider a system of linear homogeneous equations in m real variables $\gamma_1, \ldots, \gamma_m$:

$$\gamma_1 y_1 + \ldots + \gamma_m y_m = 0 \quad \text{and} \quad \gamma_1 + \ldots + \gamma_m = 0.$$

The first vector equation reads as d real linear equations

$$\gamma_1 \eta_{1j} + \ldots + \gamma_m \eta_{mj} = 0 : \quad j = 1, \ldots, d$$

in the coordinates η_{ij} of $y_i : (\eta_{i1}, \ldots, \eta_{id})$. Altogether, we have $d+1$ linear homogeneous equations in m variables $\gamma_1, \ldots, \gamma_m$. Since $m > d+1$, there must exist a non-trivial solution $\gamma_1, \ldots, \gamma_m$. Since $\gamma_1, \ldots, \gamma_m = 0$, some γ_i are strictly positive and some are strictly negative. Let

$$r = \min\{\alpha_i/\gamma_l : \quad \gamma_i > 0\} = \alpha_{i_0}/\gamma_{i_0}.$$

Now, let $\tilde{\alpha}_i = \alpha_i - r\gamma_i$ for $i = 1, \ldots, m$. Then $\tilde{\alpha}_i \geq 0$ for all $i = 1, \ldots, m$ and $\alpha_{i_0} = 0$. Furthermore,

$$\tilde{\alpha_1} + \ldots + \tilde{\alpha_m} = (\alpha_1 + \ldots + \alpha_m) - r(\gamma_1 + \ldots + \gamma_m) = 1$$

and

$$\tilde{\alpha}_1 y_1 + \ldots + \tilde{\alpha}_m y_m = \alpha_1 y_1 + \ldots + \alpha_m y_m - r(\gamma_1 y_1 + \ldots + \gamma_m y_m) = x.$$

Therefore, we represented x as a convex combination

$$x = \sum_{i \neq i_0} \tilde{\alpha}_i y_i$$

of $m-1$ points $y_1, \ldots, y_m$.

So, if x is a convex combination of $m > d+1$ points, it can be written as a convex combination of fewer points. Iterating this procedure, we get x as a convex combination of $d+1$ (or fewer) points from S.

□

2.4 Corollary. *If $S \subset \mathbb{R}^d$ is a compact set, then conv(S) is a compact set.*
Proof: Let $\Delta \subset \mathbb{R}^{d+1}$ be the standard $d-$dimensional simplex; see our example discussed earlier. We represent this $d-$dimensional simplex as

$$\Delta = \left\{ (\alpha_1, \ldots, \alpha_{d+1}) : \sum_{i=1}^{d+1} \alpha_i = 1 \quad \text{and} \quad \alpha_i \geq 0 \text{ for } i = 1, \ldots, d+1 \right\}.$$

Then Δ is compact and so is the direct product

$$S^{d+1} \times \Delta = \left\{ (u_1, \ldots, u_{d+1}; \alpha_1, \ldots, \alpha_{d+1}) : \quad u_i \in S \quad \text{and } (\alpha_1, \ldots, \alpha_{d+1}) \in \Delta \right\}.$$

Let us consider the map $\Phi : S^{d+1} \times \Delta \to \mathbb{R}^d$,

$$\Phi(u_1, \ldots, u_{d+1}; \alpha_1, \ldots, \alpha_{d+1}) = \alpha_1 u_1 + \ldots + \alpha_{d+1} u_{d+1}.$$

Theorem 2.3 implies that the image of Φ is $\text{conv}(S)$. since Φ is continuous, the image of Φ is compact, which completes the proof.

□

Practice problems.

1. Give an example of a closed set in $\mathbb{R}^2$ whose convex hull is not closed.
2. Prove that the convex hull of an open set in $\mathbb{R}^d$ is open.

1.5 An Application: Positive Polynomials

In this section, we demonstrate a somewhat unexpected application of Carathéodory's Theorem (2.3 Theorem). We will use Carathéodory's Theorem in the space of (homogeneous) polynomials.

Let us fix positive integers k and n and let $H_{2k,n}$ be the real vector space of all homogeneous polynomials $p(x)$ of degree $2k$ in n real variables $x = (\xi_1, \ldots, \xi_n)$. We choose a bssis of $H_{2k,n}$ consisting of the monomials

$$e_1 = \xi_n^{\alpha_n}, \ldots, \xi_n^{\alpha_n} \quad \text{for} \quad a = (\alpha_1, \ldots, \alpha_n) \quad \text{where} \quad \alpha_1 + \ldots + \alpha_n = 2k.$$

Hence, dim $H_{2k,n} = \binom{n+2k-1}{2k}$. At this point, we are not particularly concerned with choosing the "correct" scalar product in $H_{2k,n}$. Instead, we declare $\{e_a\}$ the orthonormal basis of $H_{2k,n}$ hence identifying $H_{2k,n} = \mathbb{R}^d$ with $d = \binom{n+2k-1}{2k}$.

We can change variables in polynomials.

3.1 Definition. Let $U : \mathbb{R}^n \to \mathbb{R}^n$ be an orthogonal transformation and let $p \in H_{2k,n}$ be a polynomial. We define $q = U(p)$ by

$$q(x) = p(U^{-1}x) \quad \text{for} \quad x = (\xi_1, \ldots, \xi_n).$$

Clearly, q is a homogeneous polynomial of degree $2k$ in $\xi_1, \ldots, \xi_n$.

3.2 Lemma. *Let $p \in H_{2k,n}$ be a polynomial such that $U(p) = p$ for every orthogonal transformation U. Then*

$$p(x) = \gamma \|x\|^{2k} = \gamma(\xi_1^2 + \ldots + \xi_n^2)^k \quad \text{for some} \quad \gamma \in \mathbb{R}.$$

Proof: Let us choose a point $y \in \mathbb{R}^d$ such that $\|y\| = 1$ and let $\gamma = p(y)$. Let us consider then

$$q(x) = p(x) - \gamma \|x\|^{2k}.$$

Thus q is a homogeneous polynomial of degree $2k$ and $q(Ux) = q(x)$ for any orthogonal transformation U and any vector x. Moreover, $q(y) = 0$. Since for every vector x such that $\|x\| = 1$ there is an orthogonal transformation U_x such that $U_x y = x$, we have $q(x) = q(U_x y) = q(y) = 0$ and hence $q(x) = 0$ for all x such that $\|x\| = 1$. Since 1 is a homogeneous polynomial, we have $q(x) = 0$ for all $x \in \mathbb{R}^n$. Therefore, $p(x) = \gamma \|x\|^{2k}$ as claimed.

□

We are now going to invoke Theorem 2.3 to deduce the existence of an interesting identity.

3.3 Proposition. *Let k and n be positive integers. Then there exist vectors $c_1, \ldots, c_m \in \mathbb{R}^n$ such that*

$$\|x\|^{2k} = \sum_{i=1}^{m} \langle c_i, x \rangle^{2k} \quad \text{for all} \quad x \in \mathbb{R}^n.$$

In words: the $k-$th power of the sum of squares of n real variables is a sum of $2k$-th powers of linear forms in the variables.

Proof: We are going to apply Carathéodory's Theorem in the space $H_{2k,n}$.

Let

$$\mathbb{S}^{n-1} = \left\{ c \in \mathbb{R}^n : \quad \|c\| = 1 \right\}$$

be the unit sphere in $\mathbb{R}^n$. For a $c \in \mathbb{S}^{n-1}$, let

$$p_c(x) = \langle c, x \rangle^{2k} \quad \text{where} \quad x = (\xi_1, \ldots, \xi_n).$$

Hence, we have $p_c \in H_{2k,n}$. Let

$$K = conv\left(p_c : \quad c \in \mathbb{S}^{n-1}\right)$$

be the convex hull of all polynomials p_c. Since the sphere $\mathbb{S}^{n-1}$ is a compact and the map $c \longmapsto p_c$ is continuous, the set $\{p_c : \quad c \in \mathbb{S}^{n-1}\}$ is a compact subset of $H_{2k,n}$. Therefore, by Corollary 2.4, we conclude that K is compact.

Now, let us prove that $\gamma \|x\|^{2k} \in K$ for some $\gamma > 0$. The idea is to average the polynomials p_c over all possible vectors $c \in \mathbb{S}^{n-1}$ and let

$$p(x) = \int_{\mathbb{S}^{n-1}} p_c(x) dc = \int_{\mathbb{S}^{n-1}} \langle c, x \rangle^{2k} dc$$

be the average of all polynomials p_c. We observe that $p \in H_{2k,n}$. Moreover, since dc is a rotation invariant measure, we have $U(p) = p$ for any orthogonal transformation U of $\mathbb{R}^n$ and hence by Lemma 3.2 we must have

$$p(x) = \gamma \|x\|^{2k} \quad \text{for some} \quad \gamma \in \mathbb{R}.$$

We observe that $\gamma > 0$. Indeed, for any $x \neq 0$, we have $p_c(x) > 0$ for all $c \in \mathbb{S}^{n-1}$ except from a set of measure 0 and hence $p(x) > 0$.

Now, the integral above can be approximated with arbitrary precision by a finite Riemann sum

$$p(x) \approx \frac{1}{N} \sum_{i=1}^{N} p_{c_i}(x) \quad \text{for some} \quad c_i \in \mathbb{S}^{n-1}.$$

Therefore, p lines in the closure of K. Since K is closed, $p \in K$. By Theorem 2.3, we can write $p(x) = \gamma \|x\|^{2k}$ as a convex combination of some $\binom{n+2k-1}{2k} + 1$ polynomials $p_{c_i}(x) = \langle c_i, x \rangle^{2k}$. Dividing by γ, we complete the proof.

□

It is not always to come up with a particular choice of c_i in the identity of Proposition 3.3. We will now apply Proposition 3.3 to study positive polynomials.

3.4 Definition. Let $p \in H_{2k,n}$ be a polynomial. We say that p is positive provided $p(x) > 0$ for all $x \neq 0$. Equivalently, $p \in H_{2k,n}$ is positive provided $p(x) > 0$ for all $x \in \mathbb{S}^{n-1}$. Similarly, a polynomial $p \in H_{2k,n}$ is *non-negative* if $p(x) \geq 0$ for all x.

Practice Problem. Prove that the set of all positive polynomials in a non-empty open convex set in $H_{2k,n}$ and that the set of all non-negative polynomials a non-empty closed convex set in $H_{2k,n}$.

$\square$

3.5 Proposition. *Let $p \in H_{2k,n}$ be a positive polynomial. Then there exists a positive integer s and vectors $c_1, \ldots, c_m \in \mathbb{R}^n$ such that*

$$\|x\|^{2s-2k} p(x) = \sum_{i=1}^{m} \langle c_i, x \rangle^{2s} \quad \text{for all} \quad x \in \mathbb{R}^n.$$

Proof: Consider for a polynomial $f \in H_{2k,n}$,

$$f(x) = \sum_{a=(\alpha_1,\ldots,\alpha_n)} \lambda_a \xi_1^{\alpha_1}, \ldots, \xi_n^{\alpha_n} ,$$

let us formally define the differential operator as

$$f(\partial) = \sum_{a=(\alpha_1,\ldots,\alpha_n)} \lambda_a \frac{\partial^{\alpha_1}}{\partial \xi_1^{\alpha_1}}, \ldots, \frac{\partial^{\alpha_n}}{\partial \xi_n^{\alpha_n}}.$$

Let us choose a positive integer $s > 2k$ and the corresponding identity of Proposition 3.3 :

$$\|x\|^{2s} = \sum_{i=1}^{m} \langle c_i, x \rangle^{2s}.$$

Let us see what happens if we apply $f(\partial)$ to both sides of the identity. It is not very hard to see that

$$f(\partial)\Big(\langle c, x \rangle^{2s} \Big) = \frac{(2s)!}{(2s-2k)!} f(c) \cdot \langle c, x \rangle^{2s-2k}.$$

It suffices to check the earlier identity when f is a monomial and then it is straightforward. One can also see that

$$f(\partial)\Big(\|x\|^{2s} \Big) = \frac{2^{2k} s!}{(s-2k)!} g(x) \cdot \|x\|^{2s-2k} \quad \text{for some} \quad g \in H_{2k,n}.$$

The correspondence $f \mapsto g$ defines a linear transformation

$$\Phi_s : H_{2k,n} \longmapsto H_{2k,n}$$

and the crucial observation is that Φ_s converges to the identity operator I as s grows. Then again, it suffices to check this when f is a monomial in which case $\Phi_s(f) = f + O(1/s)$ by the repeated application of the chain rule.

Since $I^{-1} = I$, for all sufficiently large s the operator Φ_s is invertible and Φ_s^{-1} converges to the identity operator I as s grows. Now, we note that the set of positive polynomials is open. Therefore, for a sufficiently large s the polynomial $q = \Phi_s^{-1}(p)$ lies in a sufficiently small neighborhood of $p = I(p)$ and hence is positive. Applying $q(\partial)$ to both sides we get

$$\frac{2^{2k} s!}{(s-2k)!} \Phi_s(q) \cdot \|x\|^{2s-2k} = \frac{(2s)!}{(2s-2k)!} \sum_{i=1}^{m} q(c_i) \langle c_i, x \rangle^{2s-2k}.$$

Now, $\Phi_s(q) = p$ and $q(c_i) > 0$ for $i = 1, \ldots, m$. Rescaling, we obtain a representation of $p \cdot \|x\|^{2s-2k}$ as a sum of powers of linear forms.

□

Later on in Chapter 2, we will discuss the structure of the set of *non*-homogeneous non-negative univariate polynomials in Chapter 2. The results Some interesting metric properties of the set of non-negative multivariate polynomials are discussed in Chapter 5.

1.6 Theorems of Radon and Helly

The following very useful result was first stated in 1921 by J. Radon as a lemma.

4.1 Radon's Theorem. *Let $S \subset \mathbb{R}^d$ be a set containing at least $d+2$ points. Then, there exist two non-intersecting subsets $R \subset S$ and $B \subset S$ such that*

$$conv(R) \cap conv(B) \neq \emptyset.$$

Proof: Let $v_1, \ldots, v_m, m \geq d+2$, be distinct points from S. Consider the following system of $d+1$ linear homogeneous equations in variables $\gamma_1, \ldots, \gamma_m$

$$\gamma_1 v_1 + \ldots + \gamma_m v_m = 0 \quad \text{and} \quad \gamma_1 + \ldots + \gamma_m = 0.$$

Since $m \geq d+2$, there exists a non-trivial solution to this system. Then let

$$R = \{v_i : \gamma_i > 0\} \quad \text{and} \quad B = \{v_i : \gamma_i < 0\}.$$

Then $R \cap B = \emptyset$.

Now, let $\beta = \sum_{i:\gamma_i>0} \gamma_i$. Then, $\beta > 0$ and $\sum_{i:\gamma_i<0} \gamma_i = -\beta$, since γ's sum up to zero. Now, since $\gamma_1 v_1 + \ldots + \gamma_m v_m = 0$, we have

$$\sum_{i:\gamma_i>0} \gamma_i v_i = \sum_{i:\gamma_i<0} (-\gamma_i) v_i.$$

Now let

$$v = \sum_{i:\gamma_i>0} \frac{\gamma_i}{\beta} v_i = \sum_{i:\gamma_i<0} \frac{-\gamma_i}{\beta} v_i.$$

Hence, v is a convex combination of points from R and a convex combination of points from B. In other words, $v \in \text{conv }(R)$ and $v \in \text{conv }(B)$.

4.2 Helly's Theorem. *Let $A_i, \dots, A_m$, $m \geq d+1$, be a finite family of convex sets in $\mathbb{R}^d$. Suppose that every $d+1$ of the sets have a common point:*

$$A_{i_1} \cap \dots \cap A_{i_{d+1}} \neq \emptyset.$$

Then all the sets have a common point,

$$A_1 \cap \dots \cap A_m \neq \emptyset.$$

Proof: The proof is by induction on m(starting with $m = d+1$). Suppose that $m > d+1$. Then, by the induction hypothesis, for every $i = 1, \dots, m$ there is a point p_i in the intersection $A_1 \cap \dots \cap A_{i-1} \cap A_{i+1} \cap \dots \cap A_m$ (A_i is missing). Altogether, we have that $m > d+1$ points p_i, each of which belongs to all the sets, except perhaps A_i. If two of these points happened to coincide, we get a point which belong sto all the sets $A_i's$. Otherwise, by Radon's Theorem, there are non-intersecting subsets $R = \{p_i : i \in I\}$ and $B = \{p_j : j \in J\}$ such that there is a point

$$p \in \text{conv(R)} \cap \text{conv(B)}.$$

We claim that p is a common point of $A_1, \dots, A_m$. Indeed, all the points $p_i : i \in I$ of R belongs to the sets $A_i : i \notin I$. All the points $p_j : j \in J$ of B belong to the sets $A_j : j \notin J$. Since the sets A_i are convex, every point from conv(R) belongs to the sets $A_i : i \notin I$. Similarly, every point from conv(B) belongs to the sets $A_j : j \notin J$. Therefore,

$$p \in \bigcap_{i \notin I} A_i \quad \text{and} \quad p \in \bigcap_{j \notin J} A_j.$$

Since $I \cap J = \emptyset$, we have

$$p \in \bigcap_{i=1}^{m} A_i$$

and the proof follows.

□

2 Faces and Extreme Points

We take a closer look at convex sets. In this chapter, we are interested in local properties of closed convex sets in Euclidean space. A finite-dimensional closed convex set always has an interior when considered in a proper ambient space and, therefore, has a non-trivial boundary. We explore the structure of the boundary and define and study the faces and extreme points. We look at the structure of some particular convex sets: the Birkhoff polytope, transportation polyhedra, the moment cone, the cone of non-negative univariate polynomials and the cone of positive semidefinite matrices. Our main tools are the Isolation Theorem in a general vector space and the Krein-Milman Theorem in Euclidean space. Applications include the Schur-Horn Theorem describing the set of possible diagonals of a symmetric matrix having prescribed eigenvalues, efficient formulas for numerical integration, a characterization of the polynomials that are non-negative on the interval and numerous quadratic convexity results, including the Brickman Theorem, which describes various situations when the image of a quadratic map turns out to be convex. Quadratic convexity allows us to visualize often counter-intuitive results about the facial structure of the cone of positive semidefinite matrices through the existence and rigidity properties of configurations of points in Euclidean space.

2.1 The Isolation Theorem

In this section, we develop one of the most useful and universal tools to explore the structure of a convex set, both in finite and infinite dimensions. We will assume the reader is well-versed in linear algebra, and so I will omit the linear algebra review.

1.6 Theorem *Let V be a vector space, let $A \subset V$ be an algebraically open convex set and let $u \notin A$ be a point. Then there exists an affine hyperplane H which contains u and strictly isolates A.*

Proof: Without loss of generality, we may assume that $u = 0$ is the origin. First, we'll prove the result in the case of $V = \mathbb{R}^2$. Let $S = \{x \in \mathbb{R}^2 : \|x\| = 1\}$ be the circle of radius 1 centered at the origin. Let us project A radially into $S : v \mapsto \frac{v}{\|v\|}$. Since A is convex, it is connected, and therefore, the image of this projection is a connected arc Γ of S. Furthermore, since A is algebraically open, Γ must be an open arc

$$\Gamma = \left\{(\cos\phi, \sin\phi) : \quad \alpha < \phi < \beta\right\}$$

of the circle S. Indeed, let $x \in \Gamma$ be a point. Then $x = \frac{v}{\|v\|}$ for some $v \in A$, so we can choose a straight line L through v parallel to the tangent line to S at x. Then the intersection $L \cap A$ will be an open interval containing v, so the radial projection of A will contain an open arc combining x.

Next, we observe that the length of Γ cannot be greater than π, because otherwise Γ would have contained two antipodal points x and $-x$ and 0 would have been in A, since A is convex. Now, let v be an endpoint of Γ (which is not in Γ, since Γ is open). The straight line through 0 and v is the desired hyperplane containing 0 and strictly isolating A.

Next, suppose that dim $V \geq 2$. We can prove that there is a straight line L such that $0 \in L$ and $L \cap A = \emptyset$. To prove this, let us consider any 2−dimensional plane P containing 0. The intersection $B = P \cap A$ is a convex algebraically open subset of P (possibly empty) and as we've proved, there is a line $L \subset P$ such that $0 \in L$ and $L \cap B = \emptyset$. Then L is the desired straight line.

Now, we prove the theorem. Let $H \subset V$ be the maximal affine subspace such that $0 \in H$ and $H \cap A = \emptyset$. By maximal, we mean a subspace which has these properties and is not contained in a larger subspace with the same properties. If V is finite-dimensional, we choose H to be a subspace of the largest possible dimension such that $0 \in H$ and $h \cap A = \emptyset$. If V is arbitrary, the existence of such an H is ensured by Zorn's Lemma. We claim that H is a hyperplane. To prove this, consider the quotient V/H and let $pr : V \mapsto V/H$ be the projection. If H is not a hyperplane, then dim $V/H \geq 2$ and $pr(A)$ is an algebraically open subset in V/H. Then as we know, there exists a straight line $L \subset V/H$ such that $0 \in L$ and $L \cap pr(A) = \emptyset$. Then the preimage $G = pr^{-1}(L) = \{x : pr(X) \in L\}$ is a subspace in V such that $0 \in G, G \cap A = \emptyset, H \subset G, G$ is strictly larger than H. This contradiction shows that H must be a hyperplane.

□

2.2 Convex Sets in Euclidean Space

In this section, we will explore the consequences of the Isolation Theorem for convex sets in Euclidean space. For finite-dimensional convex sets, there is no difficulty in recognizing their interior and boundary.

2.1 Definitions. Let $A \subset \mathbb{R}^d$ be a set. A point $u \in A$ is called an interior point of A if there exists $\epsilon > 0$ such that the (open) ball $B(u_\epsilon) = \{x : \|x - u\| \leq \epsilon\}$ centered at u with radius ϵ is contained in A: $B(u, \epsilon) \subset A$. The set of al interior points of A is called the interior of A and denoted int(A). The set of all non-interior points of A is called the *boundary* of A and denoted ∂A.

Now, we prove that if, starting from any point in the convex set, we move towards an interior point of the set, and we immediately get into the interior of the set.

2.2 Lemma. *Let $A \subset \mathbb{R}^d$ be a convex set and let $u_0 \in int(A)$ be an interior point of A. Then, for any $u_1 \in A$ and any $0 \leq \alpha < 1$, the point $u_\alpha = (1 - \alpha)u_0 + \alpha u_1$ is an interior point of A.*

Proof: Let $B(u_0, \epsilon) \subset A$ be a ball centered at u_0 and contained in A. Then simple elementary geometry shows that $B(u_\alpha, (1-\alpha)\epsilon) \subset A$. Thus, we're done.

□

2.3 Corollary. *Let $A \subset \mathbb{R}^d$ be a convex set. Then int(A) is a convex set.*

Proof: Consider $u, v \in$ int(A) be points and let $w = \alpha u + (1-\alpha)v$ for $0 \leq \alpha \leq 1$. If $\alpha < 1$, we apply Lemma 2.2 with $u_0 = v, u_1 = u$ and $w = u_\alpha$ to show that $w \in$int(A). If $\alpha = 1$, then $w = u \in int(A)$.

Consider that we want to show that if a non-empty convex set in Euclidean space has empty interior, then we can pass to a smaller ambient space, where the set acquires an interior. This property makes the finite-dimensional situation radically different from the infinite-dimensional case.

2.4 Theorem. *Let $A \subset \mathbb{R}^d$ be a convex set. If int $A = \emptyset$, then there exists an affine subspace $L \subset \mathbb{R}^d$ such that $A \subset L$ and dim $L < d$.*

Proof: First, suppose there are no $d+1$ affinely independent points $v_1, \ldots, v_{d+1}$ in A. For if there were no such points, then $\Delta =$conv$(v_1, \ldots, v_{d+1}) \subset A$. Then, section 2.3 would imply that Δ contains an interior point. Let $k < d+1$ be the maximum number of affinely independent points in A and let $v_1, \ldots, v_k$ be such points. Then, for each point $v \in A$, there is a solution to the system

$$\gamma_1 v_1 + \ldots + \gamma_k v_k + \gamma v = 0\,,$$
$$\gamma_1 + \ldots + \gamma_k + \gamma = 0$$

such that $\gamma \neq 0$. Then $v \in A$ can be expressed as an affine combination of $v_1, \ldots, v_k$,

$$v = \sum_{i=1}^{k} (-\frac{\gamma_1}{\gamma}) v_k.$$

Therefore, A is contained in the affine subspace L that is the affine hull of $v_1, \ldots, v_k$. So, dim $L = k-1 < d$.

□

2.5 Definition. The dimension of a convex set $A \subset \mathbb{R}^d$ is the dimension of the smallest affine subspace that contains A. By convention, the dimension of the empty set is -1 (obviously).

2.6 Definition. Let $K \subset \mathbb{R}^d$ be a closed convex set. A (possibly empty) set $F \subset K$ is called a face of K if there exists an affine hyperplane H which isolates K and such that $F = K \cap H$. If F is a point, then F is called an exposed point of K. A nonempty face

$F \neq K$ is called a *proper* face of K.

Next, we prove that a boundary point lies in some face of a closed convex set.

2.7 Theorem. Let $K \subset \mathbb{R}^d$ be a convex set with a non-empty interior and let $u \in \partial K$ be a point. Then there exists an affine hyperplane H, called a supprot hyperplane at u, such that $u \in H$ and H isolates K.

Proof: By Corollary 2.3, int(K) is a nonempty convex open set. Therefore, int(K) is a convex, algebraically open set such that $u \notin int(K)$. Therefore, by Theorem 1.6, there is an affine hyperplane H containing u and isolating $int(K)$. Then by Problem 2, Section 2.3, H isolates K, so H is a support hyperplane at u.

□

2.8 Corollary. Let $K \subset \mathbb{R}^d$ be a closed convex set with non-empty interior and let $u \in \partial K$ be a point. Then, there is a proper face F of K such that $u \in F$.

Proof: Let H be a support of K at u. Let $F = H \cap K$.

□

2.9 Theorem. *Let $A \subset \mathbb{R}^d$ be a non-empty convex set and let $u \notin A$ be a point. Then, there is an affine hyperplane $H \subset \mathbb{R}^d$ such that $u \in H$ and H isolates A.*

Proof: Let us choose the minimal affine subspace $L \subset \mathbb{R}^d$ such that $A \subset L$. Theorem 2.4 implies that A has a non-empty interior as a subset of L. If $u \notin L$, we can choose H disjoint from L. Hence we may assume that $u \in L$. Thus, restricting ourselves to L, we see that int(A)$\neq \emptyset$ (in L) and that $u \in L$. Then, by Theorem 2.7, there is an affine hyperplane $\widehat{H}$ in L, such that $u \in \hat{H}$ and $\widehat{H}$ isolates A. Then, we can choose any hyperplane H such that $H \cap L = \widehat{H}$.

□

Practice Problem

1. Let $A \subset \mathbb{R}^d$ be a convex set and let $L \subset \mathbb{R}^d$ be an affine subspace such that $L \cap A = \emptyset$. Prove that there exists an affine hyperplane H such that $L \subset H$ and H isolates A.

2.3 Extreme Points. The Krein-Milman Theorem for Euclidean Space

Certain points on the boundary of a convex set capture a lot of information about the set both in finite and infinite dimensions. Here I present the central definition for this chapter.

3.1 Definition. Let V be a vector space and let $A \subset V$ be a set. A point $a \in A$ is called an *extreme* point of A provided for any two points $b, c \in A$ such that $(b+c)/2 = a$ one must have $b = c = a$. The set of all extreme points of A is denoted $ex(A)$.

3.2 Theorem. *Let V be a vector space, let $A \subset V$ be a non-empty set and let $f : V \to \mathbb{R}$ be a linear functional.*

1. Suppose that f attains its maximum (resp. minimum) on A at a unique point $u \in A$, that is, $f(u) > f(v) \forall v \neq u, v \in A$ (resp. $f(u) < f(v) \forall v \neq u, v \in A$). Then u is an extreme point of A.

2. Suppose that f attains its maximum (minimum) α on A and suppose that $B = \{x \in A : f(x) = \alpha\}$ is the set where the maximum (minimum) is attained. Let u be an extrme point of B. Then, u is an extreme point of A.

Proof: First, let us discuss the maximum case. If $u = (a+b)/2$, then $f(u) = (f(a)+f(b))/2$, where $f(a) \leq f(u)$ and $f(b) \leq f(u)$. Therefore, $f(a) = f(b) = f(u)$ and we must have $a = b = u$, because the maximum is unique. For the second part, suppose that $u = (a+b)/2$ for $a, b \in A$. Then $\alpha = f(u) = (f(a) + f(b))/2$ and $f(a), f(b) \leq \alpha$. Thus we must have $f(a) = f(b) = \alpha$, so $a, b \in B$. Then, $a = b = u$ since u is an extreme point of B.

$\square$

Practice Problems

1. Let $K \subset \mathbb{R}^d$ be a closed convex set and let $F \subset K$ be a face. Prove that if $u \in F$ is an extreme point of F, then u is an extreme point of K.

2. Let $K \subset \mathbb{R}^d$ be a compact convex set and let $u \in K$ be a point such that $\|u\| \geq \|v\|$ for each $v \in K$. Prove that u is an extreme point of K.

We now prove a finite-dimensional version of a *quite general* and powerful result obtained by Krein and Milman in 1940.

3.3 Minkowski's Theorem. *Let $K \subset \mathbb{R}^d$ be a compact convex set. Then K is the convex hull of the set of its extreme points: K =conv(ex(K)).*

Proof: We begin by proof of induction in dimension d. Suppose $d = 0$. Then K is a point and the result clearly follows. Now, suppose $d > 0$. Without loss of generality, we may assume that $int(K) \neq \emptyset$. Otherwise, K lies in an affine subspace of a smaller dimension (recall Theorem 2.4) and the result follows by the induction hypothesis. We must show that every point $u \in K$ can be represented as a convex combination of extreme points of K. If $u \in \partial K$, then by Corollary 2.8, there exists a face F of K such that $u \in F$. Then F lies in an affine subspace of a smaller dimension, and by the induction hypothesis $u \in \text{conv}(\text{ex}(F))$, so the result follows since $\text{ex}(F) \subset \text{ex}(K)$.

Now, suppose that $u \in \text{int}(K)$. We can then draw a straight line L through u. The intersection $L \cap K$ is an interval $[a, b] \in \partial K$ and u is an interior point of $[a, b]$. As we already proved, $a, b \in \text{conv}(\text{ex}(K))$. Since u is a convex combination of a and b, the result follows.

□

The following corollary underscores the importance of extreme points for optimization.

3.4 Corollary. *Let $K \subset \mathbb{R}^d$ be a compact convex set and let $f : \mathbb{R}^d \to \mathbb{R}$ be a linear functional. Then there exists an extreme point u of K such that $f(u) \geq f(x)$ $\forall x \in K$.*

Proof: Clearly, we see f attains its maximum value, say, α on K. Then let $F = \{x \in K \mid f(x) = \alpha\}$ be the corresponding face of K. Then $\text{ext}(F) \neq \emptyset$ and any $u \in \text{ext}(F)$ is an extreme point of K.

□

We now prove another useful result whose proof resembles that of Theorem 3.3.

3.5 Lemma. *Let $A \subset \mathbb{R}^d$ be a non-empty closed convex set which does not contain straight lines. Then, A has an extreme point.*

Proof: Proceed again by induction on d. Clearly, the result obviously holds when $d = 0$. Now, suppose $d > 0$. Without loss of generality, we may assume A has a non-empty interior. Otherwise, using Theorem 2.4, we reduce the dimension d. Let us choose a point $a \in A$ and L to be any straight line passing through a. The intersection $L \cap A$ is a non-empty, closed interval (bounded or unbounded) that cannot be the whole line L. Let b be a boundary point of this interval. Now clearly, let $b \in \partial K$ and by Corollary 2.8, there exists a proper face F of K containing b. We observe that F is a closed convex set which does not contain straight lines and that $\dim F < d$. Applying the induction hypothesis, we conclude that F has an extreme point u. Then, problem one in example problems in section 3.2 implies that u is an extreme point of A.

□

2.4 Extreme points of polyhedra

For most of the rest of this chapter, we will be looking at the extreme points of various closed convex sets in Euclidean space. We start with a polyhedron, the set of solutions to finitely many linear inequalities in $\mathbb{R}^d$.

4.1 Definition. An extreme point of a polyhedron is called a *vertex*. Let us describe the vertices of a polyhedron.

4.2 Theorem. *Let $P \subset \mathbb{R}^d$ be a polyhedron*

$$P = \left\{ x \in \mathbb{R}^d : \langle c_i, x \rangle \leq \beta_i \quad \text{for} \quad i = 1, \ldots, m \right\},$$

where $c_i \in \mathbb{R}^d$ *and* $\beta_i \in \mathbb{R}$ *for* $i = 1, \ldots, m$.
For $u \in P$, *let*

$$I(u) = \{i : \langle c_i, x \rangle = \beta_i\}$$

be the set of the inequalities that are active on c. *Then* u *is a vertex of* P *if and only if the set of vectors* $\{c_i : i \in I(u)\}$ *linearly spans* $\mathbb{R}^d$. *In particular, if* u *is a vertex of* P, *the set* $I(u)$ *contains at least* d *indices:* $|I(u)| \geq d$.

Proof: Suppose that the vectors c_i with $i \in I(u)$ do not span $\mathbb{R}^d$. Then there is a nonzero $y \in \mathbb{R}^d$ such that $\langle y, c_i \rangle = 0$ for all $i \in I(u)$. We note that $\langle c_i, u \rangle < \beta_i$ for $i \notin I(u)$. For $\epsilon > 0$ let $u_+ = u + \epsilon y$ and let $u_- = u - \epsilon y$. Then $u = (u_+ + u_-)/2$, $u_+ \neq u_-$ and for sufficiently small $\epsilon > 0$ the points u_- and u_+ belong to the polyhedron P. Hence u is not an extreme point of P.

Suppose now that $u \in P$ and the vectors c_i with $i \in I(u)$ span $\mathbb{R}^d$. Suppose that $u = (v + w)/2$ for $v, w \in P$. Then $\langle c_i, v \rangle \leq \beta_i$ and $\langle c_i, w \rangle \leq \beta_i$. Since $\langle c_i, u \rangle = \beta_i$ for $i \in I(u)$ span $\mathbb{R}^d$, the system $\langle c_i, x \rangle = \beta_i, i \in I(u)$, of linear equations must have a unique solution. Therefore, $v = w = u$ and u is an extreme point.

□

4.3 Corollary. *A bounded polyhedron is a polytope, that is, the convex hull of finitely many points.*

Proof: By Theorem 4.2, every vertex v of a polyhedron is a solution to a system $\langle c_i, x \rangle = \beta_i, i \in I(u)$, of linear equations where the vectors $c_i : i \in I(v)$ span $\mathbb{R}^d$. Every such system has at most one solution. Therefore, the number of vertices of a polyhedron in $\mathbb{R}^d$, defined by a set of m inequalities, does not exceed $\binom{m}{d}$ and, hence, is finite. By Theorem 3.3, P is the convex hull of the set of its extreme points and the result follows.

□

Practice Problems.

1. Prove that a polyhedron has finitely many faces.

2. Prove that a face of a polyhedron is a polyhedron.

3. Prove that polytopes have finitely many faces.

4. Let $A \subset \mathbb{R}^d$ be a closed convex set. Prove that A has finitely many faces if and only if A is a polyhedron.

The effect of "unrealistic solutions" in linear programming

Consider $P \subset \mathbb{R}^d$ to be a polyhedron defined by a system of m linear inequalities. Suppose we want to solve a linear program (LP)

$$y = \min\langle c, x\rangle$$

$$\text{such that} \quad x \in P\,,$$

where $c \in \mathbb{R}^d$ is the given vector of the objective function and $x \in P$ is a vector of variables. If the point $u \in P$ where the minimum is attained is unique, then by Part 1, Theorem 3.2, u must be a vertex of P. Then by Theorem 4.2 this implies that at least d of the m inequality constraints are satisfied with equalities at u. This sometimes is not at all desired.

I will assume that the reader has background in basic linear programming, and so I will omit examples of basic LP programs such as the "diet problem" or "tea blending".

2.5 The Birkhoff Polytope

In this section, we describe the vertices of an interesting polyhedron.

5.1 Definition. Let σ be a permutation of the set $\{1, \ldots, n\}$. The permutation matrix X^σ is the $n \times n$ matrix $X^\sigma = (\xi^\sigma_{ij}) : i, j = 1, \ldots, n$, defined as follows

$$\xi^\sigma_{ij} = \begin{cases} 1 & \text{if } \sigma(j) = i, \\ 0 & \text{otherwise.} \end{cases}$$