MT5862
Discrete Geometry

LST / S12023

Module information

0.1. Module goals

This module is an overview of discrete geometry. We'll see a lot of topics and learn what the basic objects are.

0.2. Assessment

100% exam.

0.3. Contact info

Here's how to reach me:

- email/teams: `lst6`
- office: MI 328

My pronouns are the "(he/him)" ones[1]. Let me know yours if you think it will help us communicate.

0.4. Course materials

You will have:

- These lecture notes, which I may update from time to time
- Problems, hints, and solutions, at the end of this file
- A running module diary
- Periodic emails
- My handwritten board notes
- Lecture recordings ("best effort")

Everything is distributed via Moodle.

0.5. Reading List

- Matoušek, *Lectures on Discrete Geometry*
- Lovász, *Graphs and geometry*
- Simon, *Convexity: An analysis approach*

They are all available online via the library.

[1]Interestingly, the "(e/em)" pronouns were invented by Michael Spivak, author of the well-known calculus and differential geometry books.

0.6. Policies

The standard University academic good practice policies apply. Attendance isn't mandatory, but I recommend it.

0.7. Schedule

Here's a graphical representation of the semester:

	Mon	**Tue**	**Wed**	**Thu**	**Fri**
Week 1 (13th Sep)	L		L		L
Week 2 (20th Sep)			L	T	L
Week 3 (27th Sep)	L		L	T	L
Week 4 (4th Oct)			L	T	L
Week 5 (11th Oct)	L		L	T	L
Week 6 (18st Oct)					
Week 7 (25th Oct)	L		L	T	L
Week 8 (1th Nov)			L	T	L
Week 9 (8th Nov)	L		L	T	L
Week 10 (15th Nov)			L	T	L
Week 11 (22th Nov)	L		L	L	R

L: Lecture (12.00 Teams)
T: Tutorial (15.00 PHY Theatre C)
R: Revision (12.00 Teams)

0.8. Doing problems

You don't have to hand anything in, but it's a good idea to do problems as you go. I can look at anything you'd like me to, and encourage you to submit answers to questions you find tricky or interesting.

Contents

1. Affine space

1.1. Linear algebra revision

In this module, we'll be working in finite-dimensional, real vector spaces V with a Euclidean inner product, which we'll write

$$\langle v, w \rangle \tag{1.1}$$

for vectors v and w.

Nearly all the time, this will be $\mathbb{R}^d$, the usual space of column vectors of length d, with real entries.

As in MT2501 and MT3501, we have the standard basis of elementary vectors $e_1, \ldots, e_d$. If

$$v = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_d = \end{pmatrix} = \alpha_1 e_1 + \cdots + \alpha_d e_d \qquad \text{and} \qquad w = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_d \end{pmatrix} = \beta_1 e_1 + \cdots + \beta_d e_d$$

Since our inner product is the standard one, we will write

$$\langle v, w \rangle = v^\top w = \sum_{i=1}^{d} \alpha_i \beta_i \tag{1.2}$$

where we have allowed ourselves to treat the 1×1 matrix $v^\top w$ as a scalar.

Two descriptions of a linear subspace For the moment, let us think about a general d-dimensional F-vector space V. Suppose that $X \subseteq V$ is a subspace of dimension k. From MT2501 we have the very familiar description of X as the linear span of some k vectors in $\mathbb{R}^d$ (which necessarily form an LI set). In symbols:

$$X = \{\alpha_1 x_1 + \cdots + \alpha_k x_k : \alpha_1, \ldots, \alpha_k \in F\} = \operatorname{lin}\{x_i : i \in [k]\} \tag{1.3}$$

However, in MT3501, we learned another description, which is as follows. Consider the "inclusion" linear map

$$\iota : X \hookrightarrow V \qquad \text{defined by} \qquad \iota(x) = x$$

clearly ι has rank k, and so, the dual transformation

$$\iota^* : V^* \twoheadrightarrow X^*$$

will also have rank k, and so, nullity $d - k$. If a covector $\ell \in V^*$ is in the kernel of ι^* and $x \in X$ is arbitrary, we have

$$\mathbf{0} = \iota^*(\ell)(x) = \ell(\iota(x)) = \ell(x)$$

from which we conclude that ℓ vanishes on all of X. Playing this argument in reverse says that ℓ is in the kernel of ι^* iff ℓ vanishes on X. If we now let

$$\ker \iota^* = \operatorname{lin}\{\ell_j : j \in [k]\}$$

we get a different representation for X, which is

$$X = \bigcap_{j \in [k]} \{v \in V : \ell_j(v) = 0\} = \bigcap_{j \in [k]} \ker \ell_j$$

Specializing this discussion to $k = d - 1$, we see the association

$$\{\text{1-dim'l linear subspaces in } V^*\} \quad \longleftrightarrow \quad \{\text{linear hyperplanes in } V\} \tag{1.4}$$

Summarizing what we learned about linear algebra:

Theorem 1.1. *Let V be a d-dimensional vector space. A subset $X \subseteq V$ is a k-dimensional subspace iff*

- $X = \operatorname{lin}\{x_i \in X : i \in [k]\}$ *with the x_i an LI set.*
- $X = \cap_{i=1}^{d-k} H_i$ *with each linear hyperplane $H_i = \ker \ell_i$ for an LI set of covectors ℓ_i.*

Taking one last step, if $X \subseteq V$ is any subset at all, we can look at

$$\operatorname{lin} X \qquad \text{and} \qquad \operatorname{ann} X := \{\ell \in V^* : \ell \text{ vanishes on } X\} = \ker \iota^* \tag{1.5}$$

which are subspaces of V and V^* satisfying

$$d = \dim \operatorname{lin} X + \dim \operatorname{ann} X \tag{1.6}$$

Duality and orthogonality The setup above is very general, and, unless we specialize, about as far as we can go. In general, one can't conceive of V^* as "inside" V, so we can only make an association between lines in V^* and hyperplanes in V.

Here, though, we are intersted in $\mathbb{R}^d$ with the Euclidean inner product. For any $v \in \mathbb{R}^d$, the map

$$\ell_v(x) = \langle v, x \rangle = v^\top x \tag{1.7}$$

has, in standard coordinates, the $1 \times d$ matrix $v^\top$, so we get an identification of $\left(\mathbb{R}^d\right)^*$ with row vectors.

Now, for $X \subseteq V$, we see that the orthogonal complement

$$X^\perp = \{y \in V : \langle x, y \rangle = 0 \text{ for all } x \in X\} \tag{1.8}$$

is identified with the annihilator $\operatorname{ann} X$, and so

$$\mathbb{R}^d = \operatorname{lin} X \oplus X^\perp \tag{1.9}$$

Notice that this strengthens the dimension formula (1.6) in a way that doesn't hold in full generality.

Exercises

Exercise 1.1. Redo the discussion leading to (1.5) in F^d using matrices. What is the "MT2501" statement that corresponds to the ranks of ι and ι^* being equal? ♣

Exercise 1.2. Let $A \in \mathbb{R}^{m \times d}$ be an $m \times d$ matrix, and $b \in \mathbb{R}^m$ vector. Prove that either $Ax = b$ has a solution or there is a vector $\omega \in \ker A^T$ with $\langle \omega, b \rangle = 1$ but not both. ♣

Exercise 1.3. Give an example (necessarily not an inner product space) where (1.9) doesn't hold. ♣

Exercise 1.4. Let V be an d-dimensional vector space and $X \subseteq V$ a non-empty subset of V. Show that the linear span of X is equal to:

- The minimal linear subspace L of V so that $L \supseteq X$.
- The intersection of all linear subspaces L so that $L \supseteq X$.
- The intersection of all linear hyperplanes L so that $L \supseteq X$.

♣

1.2. Affine space

Linear subspaces of $\mathbb{R}^d$ aren't a rich enough class for geometry, since we will want to talk about, e.g., lines that don't go through the origin.

Affine subspaces

Definition 1.2. Let V be a d-dimensional vector space. An *affine subspace* of dimension k of V is a set

$$v + W = \{v + W : w \in W\} \qquad (v \in V \text{ fixed and } W \subseteq V \text{ a } k\text{-dimensional subspace})$$

♡

We set up some terminology:

- Dimension zero takes $W = \{\mathbf{0}\}$, so $v + W = \{v\}$. Affine subspaces of dimension zero are *points*.
- Dimension d has $W = V$. Since $v + V = V$ for any $v \in V$, we get the unique affine subspace of dimension d, which we just call *affine space*.
- Dimension $d - 1$ implies that W is a linear hyperplane. We call $v + W$ an *affine hyperplane*

Remark 1.3. In MT3501, you have seen affine subspaces $v + W$ (with v varying) as the vectors in a quotient V/W. Here we are thinking geometrically, with affine subspaces being the natural setup when we want to "forget the origin".

Two descriptions of an affine subspace The next task is to derive an analogy to the two descriptions of a a linear space. Again, fix a d-dimensional vector spance V and let $A = v + W \subseteq V$ be an affine subspace of dimension k.

Two quick observations:

- If A goes through the origin, it is a linear subspace. Indeed, we would have, for some $w \in W$, $\mathbf{0} = v + w$, so $v = -w \in W$. And then $v + W = W$.
- If A does not contain the origin, then A cannot be closed under linear combinations, since $0v = \mathbf{0} \notin A$.

However, there is a restricted class of linear combinations that *do* preserve an affine subspace.

Definition 1.4. Let V be a vector space and $x_1, \dots, x_k \in V$ vectors. An *affine combination* of $x_1, \dots, x_k$ is a linear combination

$$\alpha_1 x_1 + \cdots + \alpha_k x_k \qquad (\alpha_1 + \cdots + \alpha_k = 1)$$

♡

The check is quick.

Lemma 1.5. *Let V be a vector space and $A = v + W$ an affine subspace of V. If $x_1, \dots, x_k \in A$, then any affine combination of the x_i is also in A.*

Proof. Write $x_i = v + w_i$ with $w_i \in W$ for $i \in [k]$. Let the coefficients α_i of an affine combination be given. We compute

$$\sum_{i=1}^k \alpha_i x_i = \sum_{i=1}^k \alpha_i (v + w_i) = \overbrace{\left(\sum_{i=1}^k \alpha_i\right) v}^{=1v=v} + \overbrace{\left(\sum_{i=1}^k \alpha_i w_i\right)}^{\in W}$$

□

Definition 1.6. Let V be a vector space over a field F, and $X \subseteq V$. The *affine span* $\operatorname{aff} X$ is defined to be the set of all affine combinations of points in X. In symbols:

$$\operatorname{aff} X = \left\{ \alpha_1 x_1 + \cdots + \alpha_k x_k : k \in \mathbb{N}, \alpha_i \in F, \sum_{i=1}^k \alpha_i = 1 \right\}$$

♡

Now we can look at generation. In the affine setting, the familiar statement from Euclid that "two points make a line" does hold.

Lemma 1.7. *Let V be a vector space and $X \subseteq V$ an affine subspace of dimension k. Then there are points $x_0, x_1, \dots, x_k \in A$ so that $A = \operatorname{aff}\{x_0, x_1, \dots, x_k\}$.*

Proof. Let $X = v + W$ with $v \in V$ and $W \subseteq V$ a k-dimensional linear subspace of V. Let $w_1, \dots, w_k$ be a basis of W. Let $x \in X$ be given and write $x = v + w$ with $w \in W$. Then we have, for some scalars $\alpha_1, \dots, \alpha_k$,

$$x = v + \sum_{i=1}^k \alpha_i w_i = (1 - \alpha_1 - \cdots - \alpha_k) v + \sum_{i=1}^k \alpha_i (v + w_i)$$

which is an affine combination of the $k + 1$ points

$$v, v + w_1, \dots, v + w_k$$

□

The generating set we found is built on a basis for the linear space W, so we would like it to be independent in some sense.

Definition 1.8. Let V be a vector space and $x_1, \dots, x_k \in V$ be vectors. An *affine dependence* among the points x_i is a linear combination

$$\alpha_1 x_1 + \cdots + \alpha_k x_k = \mathbf{0} \qquad (\alpha_1 + \cdots + \alpha_k = 0)$$

The points $x_1, \dots, x_k$ are *affinely independent* if the only affine dependence among them has all $\alpha_i = 0$. ♡

The apparent asymmetry between the definitions of affine combination and affine dependence can be explained as follows. Suppose that x is in the affine span $x_1, \dots, x_k$. Then we have

$$x = \alpha_1 x_1 + \cdots + \alpha_k x_k$$

with $\alpha_1 + \cdots + \alpha_k = 1$. Rearranging, we see that

$$\mathbf{0} = -x + \alpha_1 x_1 + \cdots + \alpha_k x_k$$

which implies that the list $x, x_1, \dots, x_k$ is affinely dependent, from the -1 coefficient of x.

Lemma 1.9. *Let V be a vector space. Then $X \subseteq V$ is an affine subspace of dimension k iff there are affinely independent points $x_0, \dots, x_k$ in X so that $X = \operatorname{aff}\{x_i : i \in [k]\}$.*

Proof. Suppose that $X = v + W$ is an affine subspace from dimension k. Keeping the notation from Lemma 1.7 and its proof, we want to check that the points

$$v, v + w_1, v + w_2, \dots, v + w_k$$

are affinely independent. To this end, let an affine dependence

$$\mathbf{0} = \alpha_0 v + \alpha_1(v + w_1) + \cdots + \alpha_k(v + w_k)$$

be given. As an affine dependence this rearranges to

$$\mathbf{0} = 0v + \alpha_1 w_1 + \cdots + \alpha_k w_k$$

and so, since the w_i are an LI set (as a basis for W), all the $\alpha_i = 0$ for $1 \le i \le k$. As the α_i sum to zero, $\alpha_0 = 0$. Hence any affine dependence among these vectors is trivial.

Now suppose that X is the affine span of the affinely independent set $x_0, \dots, x_k$. The collection of vectors

$$x_1 - x_0, \dots, x_k - x_0$$

must be linearly independent, since any linear dependence

$$\mathbf{0} = \beta_1(x_1 - x_0) + \cdots + \beta_k(x_k - x_0)$$

yields an affine dependence

$$\mathbf{0} = (-\beta_1 - \dots - \beta_k)x_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

among the x_i, which, by our affine independence assumption must be trivial.

Now set $W = \operatorname{lin}\{x_i - x_0 : i \in [k]\}$ and let $x \in X$ be given. We write x as an affine combination of the x_i:

$$x = (1 - \alpha_1 - \cdots - \alpha_k)x_0 + \alpha_1 x_1 + \dots + \alpha_k x_k = x_0 + \overbrace{\alpha_1(x_1 - x_0) + \cdots + \alpha_k(x_k - x_0)}^{\in W}$$

so that $X \subseteq x_0 + W$. Playing this argument in reverse gives the reverse inclusion. □

Lemma 1.9 says that affine subspaces of a vector space have a notion of a basis.

Definition 1.10. Let A be a k-dimensional affine subspace of a vector space V. A set of points $x_0, \dots, x_k$ in A is an *affine basis* of A if

$$A = \operatorname{aff}\{x_0, \dots, x_{k+1}\}$$

and the x_i are an affinely independent set. ♡

The upshot is that the results about generation for finite dimensional linear spaces translate over:

- Every affine subspace of a vector space has a basis.
- A basis for an affine subspace is a maximal affinely independent subset.
- A basis for an affine subspace is a minimal generating set.

Next we should look at describing an affine subspace in terms of hyperplanes. We start with a useful observation.

Lemma 1.11. *Let $A = v + W$ be an affine subspace of an F-vector space V. Then a covector $\ell \in V^*$ is constant on A iff ℓ vanishes on W.*

Proof. Let $x \in A$ be given. We may write $x = v + w$, for some $w \in W$. As ℓ is a linear form $\ell : V \to F$, we can compute

$$\ell(x) = \ell(v + w) = \ell(v) + \ell(w) = \ell(v)$$

where we used, in the last step, that ℓ vanishes on W. Since x was arbitrary and v independent of x, ℓ is constant on A.

Now suppose that ℓ is constant on A and let $w, z \in W$ be given. We have

$$\mathbf{0} = \ell(v + w) - \ell(v + z) = \ell(w) - \ell(z)$$

so ℓ is constant on W. As a linear subspace, $\mathbf{0} \in W$, so, for all $w \in W$, $\ell(w) = \ell(\mathbf{0}) = \mathbf{0}$. Hence ℓ vanishes on W. □

What we get from this is that an affine hyperplane in V is of the form

$$\{x \in V : \ell(x) = \alpha\} \qquad (\ell \in V^* \text{ non-zero and } \alpha \in F) \tag{1.10}$$

Note that many pairs of ℓ and α determine the same hyperplane, since you can scale them. However, if $\alpha \neq 0$ there is a canonical choice by scaling to make it one. We then get an association

$$\{\text{non-zero covectors } \ell \in V^*\} \qquad \longleftrightarrow \qquad \{\text{affine hyperplanes in } V \text{ not through the origin}\} \tag{1.11}$$

We will explore this duality map a lot. It's, in a way, nicer than what we had linearly (no scales) but we had to give something up (not all hyperplanes dualize).

Now we can give the second description of an affine subspace.

Lemma 1.12. *Let V be a d-dimensional F-vector space. A non-empty subset $X \subseteq V$ is a k-dimensional affine subspace iff there are LI covectors $\ell_1, \dots, \ell_{d-k}$ and scalars $\alpha_1, \dots, \alpha_k$ so that*

$$X = \bigcap_{i=1}^{d-k} H_i$$

where H_i is the affine hyperplane $\{x : \ell_i(x) = \alpha_i\}$.

Proof. Suppose that $X = v + W$ is an affine subspace of dimension k. From Theorem 1.1, there is an LI set of covectors

$$\ell_1, \dots, \ell_{d-k}$$

so that their vanishing set is W. By Lemma 1.11, each ℓ_i is constant on X. If we take $\alpha_i = \ell_i(v)$, we get that

$$X \subseteq \bigcap_{i=1}^{d-k} H_i$$

For the reverse containment, notice that, if we add another LI covector to the list, the common vanishing set will be of dimension $k-1$ and so strictly smaller than W.

Now suppose that X is the intersection of $d-k$ hyperplanes as in the statement. Set

$$W = \bigcap_{i=1}^{d-k} \ker \ell_i$$

By Theorem 1.1, W is a linear subspace of V of dimension k. By hypothesis, there is a $v \in X$. We claim that $X = v + W$, which shows it is an affine space of dimension k. To this end, let $w \in W$ be given. Then, for each $i \in [d-k]$,

$$\ell_i(v+w) = \ell_i(v) = \alpha_{d-k} \qquad \Longrightarrow \qquad v + w \in X$$

As w was arbitrary $v + W \subseteq X$. For the other inclusion, note that if $x \in X$, then, for each i

$$\ell_i(x - v) = 0$$

which implies that $x - v \in W$. As x was arbitrary, $X \subseteq v + W$. □

Exercises

Exercise 1.5. Show that "two points make a line" in $\mathbb{R}^d$ and that this line is unique. ♣

Exercise 1.6. Show that if A is an affine subspace of a d-dimensional vector space V, and $B \subseteq A$ is an affine subspace of V of the same dimension as A, then $A = B$. ♣

Exercise 1.7. Let V be a d-dimensional vector space and $\{A_i\}_{i \in I}$ an indexed collection of affine subspaces. Show that, if non-empty, the intersection $\cap_{i \in I} A_i$ is also an affine subspace of V. ♣

Exercise 1.8. Explain why the hypothesis in Exercise 1.7 that the intersection is non-empty can't be dropped. ♣

Exercise 1.9. Let V be an d-dimensional vector space and $X \subseteq V$ a non-empty subset of V. Show that the affine span of X is equal to:

- The minimal affine subspace A of V so that $A \supseteq X$.
- The intersection of all affine subspaces A so that $A \supseteq X$.
- The intersection of all affine hyperplanes H so that $H \supseteq X$.

♣

This was an exam question. This statement is knows as "Desargues' Theorem" in classical projective geometry. I've put it here, because it's very robust and works in any vector space.

Exercise 1.10. Let V be a 3-dimensional vector space. Let $\{a, A, b, B, c, C\} \subseteq V$ be six points such that each of the sets

$$\operatorname{aff}\{a, A, b, B\} \quad \operatorname{aff}\{b, B, c, C\} \quad \operatorname{aff}\{a, A, c, C\}$$

has dimension 2 and that no pair of the lines

$$\operatorname{aff}\{a, A\} \quad \operatorname{aff}\{b, B\} \quad \operatorname{aff}\{c, C\}$$

are in the same direction. Show that, if all six points do not lie in a common plane, then

$$\varnothing \neq \operatorname{aff}\{a, A\} \cap \operatorname{aff}\{b, B\} \cap \operatorname{aff}\{c, C\}$$

♣

1.3. First geometric steps

Let us return to affine geometry in $\mathbb{R}^d$. The inner product gives us the *Euclidean norm*

$$\|x\| = \sqrt{\langle x, x\rangle} \qquad (x \in \mathbb{R}^d) \tag{1.12}$$

which gives us the *Euclidean metric*

$$\operatorname{dist}(x, y) = \|x - y\| \qquad (x, y \in \mathbb{R}^d) \tag{1.13}$$

Both of these behave nicely with respect to translation and scaling. From now on, when we make topological statements, they are with respect to the "standard" topology from the Euclidean metric.

Let us consider the following question:

Does the list of distances

$$\operatorname{dist}(x, x_1), \dots, \operatorname{dist}(x, x_{d+1})$$

to a fixed set of $d + 1$ points in $\mathbb{R}^d$ determine x?

Without some restriction on the x_i, the answer is clearly "no". For example, take $d = 1$ and put $x_1 = x_2 = 0$. Then for any $y \in \mathbb{R}^1$, the the distances from y and $-y$ to x_1 and x_2 are the same. What is going wrong is that this choice of the x_i is affinely dependent.

Theorem 1.13. *Let $x_1, \dots, x_{d+1}$ be a set of points that are affinely independent $\mathbb{R}^d$. Then if x and y are points such that*

$$\operatorname{dist}(x, x_i) = \operatorname{dist}(y, x_i) \qquad \textit{(all } i \in [d+1]\textit{)}$$

we have $x = y$.

Proof. Suppose, for a contradiction, that we have points $x \neq y$ in $\mathbb{R}^d$, such that

$$\operatorname{dist}(x, x_i)^2 = \operatorname{dist}(y, x_i)^2 \qquad (\text{all } i \in [d+1]) \tag{1.14}$$

(because distances are positive, this happens iff the unsquared equations in the statement hold). This means that all the x_i lie in the set

$$H = \{z : \operatorname{dist}(x, z)^2 = \operatorname{dist}(y, z)^2\}$$

Expanding out both sides of the equation defining H, we get

$$\langle x, x\rangle + 2\langle x, z\rangle + \langle z, z\rangle = \langle y, y\rangle + 2\langle y, z\rangle + \langle z, z\rangle$$

Rearranging we get

$$\langle x - y, z\rangle = \frac{1}{2}(\langle y, y\rangle - \langle x, x\rangle)$$

Because x and y are fixed, the r.h.s. is a constant and the l.h.s. is a linear form that corresponds to a covector. We conclude that H is an affine hyperplane.

It now follows (cf Exercise 1.9) that the affine span of the x_i has dimension at most $d - 1$. Lemma 1.9 then implies that there is no affinely independent subset of H with cardinality $d + 1$. This contradicts the hypothesis that the x_i are affinely independent, completing the proof. □

The set H that appears in the proof is called the *perpendicular bisector* of x and y, and we will use it again. Another useful idea we have encountered is that a good non-degeneracy assumption is to ask that small subsets of a point set are affinely independent.

Definition 1.14. Let $X = \{x_1, \dots, x_n\}$ be a finite set of points in $\mathbb{R}^d$. We say that X is in *general position* if any subset of X with cardinality at most $d + 1$ is affinely independent. ♡

Geometrically, what this definition says is that no three of the x_i lie on a line, no four of the x_i lie on a plane, and so on.

Remark 1.15 (Unexaminable remark). General position point sets are "typical" in the sense that the ones that aren't satisify a non-trivial polynomial equation, so they are closed and nowhere dense in the set of configurations with the same number of points.

Exercises

Exercise 1.11. Are the corners of a triangle in $\mathbb{R}^2$ in general position? What about in $\mathbb{R}^3$?

Are the corners of a unit square in general position in $\mathbb{R}^2$? What about $\mathbb{R}^3$ (where the square stays flat)? ♣

Exercise 1.12. Say that a set $X = \{x_1, \dots, x_n\}$ of points in $\mathbb{R}^d$ are in *strongly general position* if the affine spans of any two disjoint subsets $A, B \subseteq X$ intersect in the "expected" dimension

$$\dim \operatorname{aff} A + \dim \operatorname{aff} B - d$$

when this is positive and otherwise don't intersect.

Describe a point set that is in general position but not strongly general position. ♣

Exercise 1.13. Let $A \subseteq \mathbb{R}^d$ be an affine subspace. Show that A is closed. ♣

Exercise 1.14. Let A be an affine subspace of $\mathbb{R}^d$, and $x \in \mathbb{R}^d$ a point. Show that there is a unique closest point y in A to x. ♣

2. Convex sets

2.1. Convexity

For points $x, y \in \mathbb{R}^d$, we'll use the notation

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

for the closed segment between x and y.

Definition 2.1. A subset $X \subseteq \mathbb{R}^d$ is *convex* if for all $x, y \in X$, the segment $[x, y] \subseteq X$. ♡

Informally, convex sets are closed under taking segments.

Example 2.2. In $\mathbb{R}^1$, the only convex sets are intervals, including $\mathbb{R}^1$ itself and half-infinite ones. ♠

Example 2.3. These are important examples of convex sets in $\mathbb{R}^d$:

- The empty set. (Later, we have to watch for special cases.)
- $\mathbb{R}^d$ itself.
- Any affine subspace of $\mathbb{R}^d$.
- A *closed half-space* H^- is a set of the form

  $$\{x \in \mathbb{R}^d : \ell(x) \le \alpha\} \qquad (\ell \in (\mathbb{R}^d)^*, \alpha \in \mathbb{R})$$

  Notice that the boundary of a closed half-space is an affine hyperplane.
- A *open half-space* is a set of the form

  $$\{x \in \mathbb{R}^d : \ell(x) > \alpha\} \qquad (\ell \in (\mathbb{R}^d)^*, \alpha \in \mathbb{R})$$

- A segment $[x, y]$.
- The *closed ball* $D(x, \varepsilon) = \{y \in \mathbb{R}^d : \mathrm{dist}(x, y) \le \varepsilon\}$.
- The *open ball* $B(x, \varepsilon) = \{y \in \mathbb{R}^d : \mathrm{dist}(x, y) < \varepsilon\}$.
- The *standard $(d-1)$-simplex*

  $$\Delta_{d-1} = \{x = (\alpha_i)_{i=1}^d : \text{all } \alpha_i \ge 0 \text{ and } \alpha_1 + \cdots + \alpha_d = 1\}$$

- The *d-cube*

  $$\square_d = \{x = (\alpha_i)_{i=1}^d : -1 \le \alpha_i \le 1\}$$

♠

We will do some of these in class, and definitely check the rest (Exercise 2.1) for practice.

Topological preliminaries At this point, we recall some basic notions from metric space topology to set up notation.

Definition 2.4. Let $X \subseteq \mathbb{R}^d$. The set X is said to be *open* if it is a union of open balls; it is *closed* if X^c is open. Alternatively, X is closed if every convergent sequence in X has its limit in X.

The set X is *open as a subspace of its affine span* if $X = U \cap \operatorname{aff} X$ for some open set U; X is *closed as a subspace of its affine span* if $X^c \cap \operatorname{aff} A$ is open in its affine span. ♡

Definition 2.5. Let $X \subseteq \mathbb{R}^d$. The *interior* of X is the inclusion-wise maximal open subset $\operatorname{int} X \subseteq X$ contained in X. The *closure* of X is the inclusion-wise minimal closed set $\operatorname{cl} X \supseteq X$ containing X.

The (topological) *boundary* $\operatorname{bd} X$ is $\operatorname{cl} X \setminus \operatorname{int} X$. When X is closed, it contains its boundary. ♡

In our affine setting we'll want to work with sets of dimension lower than d. Since a proper affine subspace of $\mathbb{R}^d$ is closed and has empty interior, we'll need two topological definitions better adapted to our setting.

Definition 2.6. Let $X \subseteq \mathbb{R}^d$. The *relative interior* $\operatorname{ri} X \subseteq X$ is the interior of X as a subspace of its affine span. The *relative boundary* $\operatorname{rb} X \subseteq X$ is the boundary of X as a subspace of its affine span. ♡

Example 2.7. Let us work out the relative interior of the standard simplex $\Delta_{d-1} \subseteq \mathbb{R}^d$. We are going to show that it is the set of $x \in \Delta_{d-1}$ with all coordinates positive.

Let $\ell : \mathbb{R}^d \to \mathbb{R}$ be the linear map that sums the coordinates. The affine span

$$\operatorname{aff} \Delta_{d-1} = \operatorname{aff}\{e_1, \dots, e_d\} = \{x : \ell(x) = 1\}$$

which is the hyperplane of vectors with coordinates summing to one. (We can see that the elementary vectors are affinely independent, and ℓ vanishes on all of them, so they are an affine base for it.)

Now we suppose that $x \in \Delta_{d-1}$ has a zero coordinate, say the ith. Then, for any $j \neq i$ and $r > 0$, we have that $x - (r/2)e_i + (r/2)e_j$ lies in the intersection $\operatorname{aff} \Delta_{d-1} \cap B(x, r)$, but it has a negative coordinate, so it is not in Δ_{d-1}. As r was arbitrary, x is in the relative boundary of Δ_{d-1}. ♠

We will have occasion to use the following very frequently.

Definition 2.8. A set $X \subseteq \mathbb{R}^d$ is *compact* if it is closed and bounded. ♡

Theorem 2.9 (Extreme Value Theorem). *Let $X \subseteq \mathbb{R}^d$ be compact and $f : X \to \mathbb{R}$ continuous. Then f achieves a maximum over f.*

It is easy to construct counterexamples when X is non-compact, so we will very often be interested in compact sets.

Lemma 2.10. *Let $[x, y] \subseteq \mathbb{R}^d$ be a segment. If $\ell \in \left(\mathbb{R}^d\right)^*$ achieves its maximum on* $\operatorname{ri}[x, y]$*, then ℓ is constant on $[x, y]$.*

Proof. The relative interior of $[x, y]$ is (see Exercise 2.3)

$$\operatorname{ri}[x, y] = \{(1 - \alpha)x + \alpha y : 0 < \alpha < 1\}$$

As $[x, y]$ is compact, the maximum is attained somewhere. Suppose that it is attained at $z = (1 - \lambda)x + \lambda y$ for $0 < \lambda < 1$. Linearity of ℓ implies that

$$\ell(z) = (1 - \lambda)\ell(x) + \lambda\ell(y)$$

We assume that $\ell(x) \leq \ell(z)$ and $\ell(y) \leq \ell(z)$, so, if either $\ell(x) < \ell(z)$ or $\ell(y) < \ell(z)$

$$\ell(z) = (1 - \lambda)\ell(x) + \lambda\ell(y) < (1 - \lambda)\ell(z) + \lambda\ell(z) = \ell(z)$$

which is a contradiction. Hence

$$\ell(x) = \ell(y) = \ell(z)$$

Another application of linearity implies that ℓ is constant over $[x, y]$. For each $0 \leq \beta \leq 1$,

$$\ell((1 - \beta)x + \beta y) = (1 - \beta)\ell(x) + \beta\ell(y) = \ell(z)$$

□

The contrapositive statement: if ℓ is non-constant on $[x, y]$ that, for any $z \in \operatorname{ri}[x, y]$ either

$$\ell(x) < \ell(z) < \ell(y) \qquad \text{or} \qquad \ell(y) < \ell(z) < \ell(x)$$

is also useful.

Exercises

Exercise 2.1. Check that all the sets in Example 2.3 are indeed convex. ♣

Exercise 2.2. Show that the intersection an arbitrary collection of convex sets is convex. ♣

Exercise 2.3. What is the relative interior of a segment $[x, y]$? ♣

Exercise 2.4. What is the relative interior of the cube? ♣

Exercise 2.5. Show that the standard simplex is compact. ♣

Exercise 2.6. Give an example of a set where the relative boundary is different than the boundary. ♣

Exercise 2.7. Is it true that if $X, Y \subseteq \mathbb{R}^d$ are closed convex subsets, and $(\operatorname{ri} X) \cap Y$ is non-empty that $X \cap (\operatorname{ri} Y)$ is non-empty? ♣

Exercise 2.8. Suppose that $X \subseteq \mathbb{R}^d$ is convex and $T : \mathbb{R}^d \to \mathbb{R}^m$ is an *affine map* $T(x) = L(x) + b$, where $L : \mathbb{R}^d \to \mathbb{R}^m$ is linear and $b \in \mathbb{R}^m$. Show that $T(X)$ is convex. ♣

2.2. Convex combinations

Definition 2.11. Let $x_1, \dots, x_n$ be a finite set of points in $\mathbb{R}^d$. A *convex combination* of the x_i is a linear combination

$$x = \alpha_1 x_1 + \dots + \alpha_n x_n \qquad (\text{all } \alpha_i \geq 0 \text{ and } \alpha_1 + \dots + \alpha_n = 1) \tag{2.1}$$

♡

A convex combination is a special kind of affine combination. Notice the similarity between the constraints on the coefficients and the standard simplex.

Lemma 2.12. *A set $X \subseteq \mathbb{R}^d$ is convex if and only if it is closed under taking convex combinations.*

Proof. Notice that every point of the segment $[x, y]$ is a convex combination of x and y so if X is closed under taking convex combinations, it is certainly convex.

Now suppose that X is convex and $x_1, \dots, x_n \in X$. If $n = 1, 2$ certainly any convex combination of the x_i is in X. For $x \geq 3$, we do induction on n. We may assume that al the coefficients $\alpha_i > 0$ (otherwise, we can apply the I.H. right away). Then

$$\alpha_1 x_1 + \dots + \alpha_n x_n = \alpha_1 x_1 + (1 - \alpha_1) \overbrace{\frac{1}{1 - \alpha_1} (\alpha_2 x_2 + \dots + \alpha_n x_n)}^{\text{in } X \text{ by IH}}$$

and then the whole r.h.s. is in X, as it is on the segment between two points in X. □

Definition 2.13. Let X be a non-empty convex subset of $\mathbb{R}^d$. The *dimension* of X is the dimension of aff X. By convention, the dimension of the empty set is -1. ♡

This definition makes sense in light of the following.

Lemma 2.14. *Let X be a non-empty convex subset of $\mathbb{R}^d$. Then X has non-empty relative interior.*

This proof is very topological and not examinable.

Unexaminable proof. Suppose that X has k-dimensional affine span. Pick an affine basis $x_1, \dots, x_{k+1} \in X$ for aff X. Since X is convex, the map $f : \Delta_k \to X$ given by

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{k+1} \end{pmatrix} \mapsto \alpha_1 x_1 + \dots + \alpha_{k+1} x_{k+1}$$

is well-defined, continuous and injective. Since Δ_k is compact, $f(\Delta_k) \subseteq X$ is homeomorphic to it. In particular, $f(\operatorname{ri} \Delta_k)$ will be open in the affine span of $x_1, \dots, x_k$, which is aff X. □

The idea of reducing a general case to a simplex will appear again.

2.3. The convex hull

Definition 2.15. Let $X \subseteq \mathbb{R}^d$. The *convex hull* of X is the inclusion-wise minimal convex set $\operatorname{conv} X$ that contains X. ♡

This definition makes sense in light of Exercise 2.2, since we can get it by intersecting every convex set that contains X. We'll be willing to work hard to reduce "every" to something more manageable. For now, we note that there is a dual formulation using convex combinations.

Lemma 2.16. *The convex hull* $\operatorname{conv} X$ *of a set* $X \subseteq \mathbb{R}^d$ *is the set* $\mathrm{CC}(X)$ *all convex combinations of points in* X*.*

Proof. Since $X \subseteq \operatorname{conv} X$, which is convex, Lemma 2.12 implies that any convex combination of points in X must be in $\operatorname{conv} X$. Hence $\mathrm{CC}(X) \subseteq \operatorname{conv} X$. The reverse inclusion follows from the fact $\mathrm{CC}(X)$ is convex, which is an exercise. □

Example 2.17.
- $\Delta_{d-1} = \operatorname{conv}(\{e_1, \dots, e_d\})$. This is easy to check.
- $D(x, \varepsilon) = \operatorname{conv}(\{y : \operatorname{dist}(y, x) = 1\})$. This is more interesting. If $y \in D(x, \varepsilon)$, pick any line L through y. By compactness, L hits the boundary of the ball in both directions; by convexity $L \cap B(x, \varepsilon)$ is then a closed segment $[x_1, x_2]$ with y in its relative interior and $\operatorname{dist}(x_i, x) = 1$. Hence the closed ball is the convex hull of its boundary sphere.

♠

We introduce another useful example of a convex set.

Definition 2.18. A *simplex* in $\mathbb{R}^d$ is the convex hull of $d + 1$ affinely independent points. ♡

Notice that, unlike the standard simplex, "a simplex" is not flat.

Exercises

Exercise 2.9. Let X be a set in $\mathbb{R}^d$. Show that the set $\mathrm{CC}(X)$ of all convex combinations of points in X is convex. ♣

Exercise 2.10. Show that the cube $\square_d$ is the convex hull of a finite point set. ♣

2.4. Projection and separation

We now introduce two basic geometric tools for dealing with convex sets: the projection theorem and the separation theorem.

Theorem 2.19. *Suppose that* X *is a non-empty closed convex subset of* $\mathbb{R}^d$ *and* $y \in \mathbb{R}^d$*. Then there is a unique* $x \in X$ *that minimizes the distance to* y*.*

Proof. A minimizer exists because X is closed and distance is bounded from below and continuous. Uniqueness is where we need convexity. Suppose that $x_1, x_2 \in X$ are such that $\operatorname{dist}(x_1, y) = \operatorname{dist}(x_2, y)$. Then x_1, x_2 and y make an isosceles triangle, with apex y. Pythagoras then implies that the midpoint x' of $[x_1, x_2]$ is closer to y than x_1 or x_2. Since X is convex $[x_1, x_2] \subseteq X$, and then $x' \in X$. Since

$$\operatorname{dist}(x', y) < \operatorname{dist}(x_1, y) = \operatorname{dist}(x_2, y)$$

x_1 and x_2 must not minimize the distance to y. It follows that the minimizer is unique. □

Now we describe the idea behind separation. Any hyperplane

$$H = \{x : \ell(x) = \alpha\} \qquad (\ell \in (\mathbb{R}^d)^*)$$

in $\mathbb{R}^d$ determines two closed half-spaces

$$H^+ = \{x : \ell(x) \geq \alpha\} \qquad H^- = \{x : \ell(x) \leq \alpha\}$$

Definition 2.20. A hyperplane H in $\mathbb{R}^d$ *separates* sets $X, Y \subseteq \mathbb{R}^d$ if $X \subseteq H^+$, $Y \subseteq H^-$ and at least of one of X or Y is not contained in H. ♡

It will turn out to be very important that we allow either X or Y to lie entirely in H.

Theorem 2.21. *Let X be a closed, convex subset of $\mathbb{R}^d$ and let $y \in \mathbb{R}^d$ lie outside of $\operatorname{ri} X$. Then there is a hyperplane separating y from X. Moreover, if $y \in \operatorname{rb} X$, y will lie in the separating hyperplane.*

This is a geometric theorem that doesn't follow for algebraic reasons. The proof is not that important to us, so I am only going to sketch it. However, the statement is really important.

Sketch. It's very easy if y is outside of $\operatorname{aff} X$. Find a hyperplane H that contains X and not y. This H separates X and y.

The next case is to assume that y is in $\operatorname{aff} X$, but that $y \notin X$, so that the projection x of y onto X is different from y. Take H to be the perpendicular bisector (cf Equation 1.14) of x and y and suppose that $X \subseteq H^+$. Neither x nor y is in H, so, in particular, $X \nsubseteq H$. Suppose, for a contradiction, that some $x' \in X$ is in H. Then the segment $[x, x'] \subseteq X$ by convexity and x, x' and y make an isosceles triangle with apex x.

As the base angles of an isosceles triangle are acute, the segment $[x, x']$ goes through the interior of the circle of radius $\operatorname{dist}(x, y)$ around y. (The tangent line to the circle meets it as a right angle.) This implies that there is a point on $[x, x']$ closer to y than x. The resulting contradiction, shows that X is disjoint from H, giving separation.

The last case is where $y \in \operatorname{rb} X$. (Now we have no hope for $y \notin H$.) A limiting argument (which we skip), using a sequence $y_i \to y$ of points not in X and separators H_i gives a separator which will necessarily contain y. □

When $y \in \operatorname{rb} X$, since $X \subseteq H^+$, y must minimize ℓ over X. This means that X is "sitting on" H, and we say that the hyperplane H *supports* X at y; if we dont' want to talk about y, we simply say that H is a *supporting hyperplane* of X.

The "cutting out" representation of a convex set is now more or less immediate.

Theorem 2.22. *Every non-empty convex set X in $\mathbb{R}^d$ is an intersection of half-spaces (some of which might have to be open). If X is closed it is the intersection of closed half-spaces.*

Here's another proof sketch which is not examinable, because it's more finicky than we need (or like).

Sketch. Let $\operatorname{cl} X$ be the closure of X. For every y outside $\operatorname{ri} \operatorname{cl} X$, there is a hyperplane H_y separating y from $\operatorname{cl} X$. In the case where $y \notin \operatorname{cl} X$, we have that $y \notin H_y^+$. If $y \notin X$, but $y \in \operatorname{cl} X$, we have $y \in H_y^+$ but not in the open half-space $\operatorname{ri} H_y^+$.

We conclude that we get exactly X by intersecting all the closed half-spaces H_y^+ for y outside X or in $X \cap \operatorname{rb} X$ and all the open half-spaces $\operatorname{ri} H_y^+$ for $y \in X \setminus \operatorname{rb} X$.

If X was closed, the second case never happens, so we only needed closed half-spaces. □

We may need a lot of half-spaces to cut out a convex set. This is very different than what we saw with affine subspaces, where finitely many hyperplanes suffices.

Example 2.23. The unit ball centered at the origin is the intersection of the half-spaces

$$H_v = \{x : \langle v, x\rangle \leq 1\} \qquad \text{all } v \text{ such that } \|v\| = 1$$

♠

The simplex is much nicer.

Example 2.24. The standard simplex is the intersection of a finite number of half-spaces. This is, in fact, how we described it. ♠

Exercises

Exercise 2.11. Describe the cube $\square_d$ as an intersection of half-spaces. ♣

Exercise 2.12. Show that you can cover the d-dimensional cube $\square_d$ by $d!$ simplices of the form[2]:

$$\Delta_\pi = \{x : -1 \leq \alpha_{\pi^{-1}(1)} \leq \cdots \leq \alpha_{\pi^{-1}(d)} \leq 1\}$$

where π ranges over all permutations of $\{1, \dots, d\}$. Check that the Δ_π have disjoint interiors. ♣

Exercise 2.13. Use Exercise 2.12 to deduce a formula for the volume of the convex hull of the origin and any other d affinely independent points in $\mathbb{R}^d$. ♣

Exercise 2.14. Let $x_1, \dots, x_{d+1}$ be points in $\mathbb{R}^d$ so that $\mathrm{dist}(x_i, x_j) = 1$. What is the volume of $\mathrm{conv}\{x_1, \dots, x_{d+1}\}$? ♣

2.5. Faces

Convex sets have somewhat boring interiors, but very interesting boundaries. From now on, we're going to look mainly at closed convex sets for this reason.

Definition 2.25. Let X be a closed convex set in $\mathbb{R}^d$. A subset $F \subseteq X$ is a *face* of X if F is closed, convex and whenever $y \in F$ and $y \in \mathrm{ri}[x_1, x_2]$ with $x_1, x_2 \in X$, then $x_1, x_2 \in F$.

A face is called *proper* if it is not all of X or empty. ♡

Informally, faces "absorb segments" or don't have segments "through" them. Notice that $\varnothing$ makes sense as a face vacuously and that X is a face trivially.

Example 2.26.

- The faces of the closed ball are empty, the whole thing, and each boundary point by itself.
- The standard simplex Δ_d has 2^{d+1} faces, which are the convex hulls of subsets of $\{e_1, \dots, e_{d+1}\}$ and empty.
- The boundary of a closed half-space is its only proper face.

♠

[2]The reason for the odd-looking π^{-1} instead of π is that if you write π out as a word in the alphabet $\{1, \dots, d\}$, it says that the location 1 is the smallest, and so on.

Lemma 2.27. *Let X be a closed convex set in $\mathbb{R}^d$ and $F \subseteq X$ be a proper face. Then $F \subseteq \operatorname{rb} X$.*

The idea of this proof is just that if F wasn't in the boundary, then it would be possible to put segments "through" it.

Proof. Let $y \in F$ be arbitrary, and pick $x \in X \setminus F$ (possible, because F is proper). Consider the set

$$S_n = \{\alpha y + (1-\alpha)x \,:\, 0 \leq \alpha \leq 1 + 1/n\}$$

(this is a segment extending beyond the end of $[x, y]$ by $1/n$). By definition $S_n \subseteq \operatorname{aff} X$. Because F is a face, S_n must leave X for every n. Because X is convex, $[x, y] \subseteq X$. Hence, y is the limit of a sequence of points in $\operatorname{aff} X \setminus X$ and so $y \in \operatorname{rb} X$. $\square$

This next lemma is really useful, since it lets us set up inductions on dimension.

Lemma 2.28. *Let X be a closed convex subset of $\mathbb{R}^d$. Then every proper face F of X is of lower dimension.*

Proof. If F has the same dimension as X, then $\operatorname{aff} F = \operatorname{aff} X$. Since F has non-empty relative interior by Lemma 2.14, we must have $F \cap \operatorname{ri} X \neq \emptyset$. This contradicts $F \subseteq \operatorname{rb} X$. $\square$

There's a partial converse.

Lemma 2.29. *Let X be a closed convex subset of $\mathbb{R}^d$ and $x \in \operatorname{rb} X$. Then x lies in a proper face of X.*

Proof. The separation theorem provides a hyperplane H that separates x from X. Since H and X are closed and convex, $F := H \cap X$ is as well. Moreover, since $X \subseteq H^+$, no segment in X can cross H, so F is a face. $\square$

The previous proof suggests a way to produce faces.

Lemma 2.30. *Let X be a closed convex subset and suppose that a linear form ℓ has a finite maximum α on X. Then $F = X \cap \{y \,:\, \ell(y) = \alpha\}$ is a face of X. If ℓ is non-constant on X, then F is proper.*

Proof. Let $H = \{y \,:\, \ell(y) = \alpha\}$. As H is closed and convex, so is $F = X \cap H$. Suppose that $[x_1, x_2] \subseteq X$ is a segment so that $F \cap \operatorname{ri}[x_1, x_2]$ is non-empty. We are now in the situation of Lemma 2.10: ℓ achieves its maximum over $[x_1, x_2]$ in the relative interior, and so it is constant over the segment. We conclude that $[x_1, x_2] \subseteq H$. Hence, $x_1, x_2 \in F$. This shows that F is a face.

If ℓ is not constant over X, then $X \nsubseteq H$, so F is proper. $\square$

Exercises

Exercise 2.15. Show that the intersection of two faces of a convex set X is again a face. ♣

Exercise 2.16. Let $X = \{x_1, \dots, x_{d+1}\}$ and $Y = \{y_1, \dots, y_{d+1}\}$ be sets of points in general position in $\mathbb{R}^d$. Show that if $\operatorname{conv} X$ and $\operatorname{conv} Y$ intersect in their relative interiors, that there is a pair of intersecting faces $F \subseteq \operatorname{conv} X$ and $G \subseteq \operatorname{conv} Y$ so that $\dim F + \dim G \leq d$. ♣

Exercise 2.17. Show that the cube $\square_d$ has $3^d + 1$ faces. ♣

2.6. Extreme points

We're now in a position to prove look at minimal generation for convex sets.

Definition 2.31. Let X be a closed convex subset of $\mathbb{R}^d$. A point $x \in X$ is an *extreme point* of X if it is not in the relative interior of any segment in X. ♡

Notice that this just says that extreme points are the zero-dimensional faces.

Example 2.32.

- Every point on the boundary of a closed ball $B(x, \varepsilon)$ is an extreme point.
- The extreme points of the simplex Δ_d are $\{e_1, \dots, e_{d+1}\}$.
- An affine subspace doesn't have any extreme points.
- A closed half-space doesn't have any extreme points.
- The quadrant
$$\left\{ \begin{pmatrix} x \\ y \end{pmatrix} : x, y \geq 0 \right\} \subseteq \mathbb{R}^2$$
has the origin as its only extreme point.

♠

Notice that the examples with a lot of extreme points are compact.

Lemma 2.33. *Let X be a closed convex subset of $\mathbb{R}^d$ and $F \subseteq X$ a proper face. Then every extreme point of F is an extreme point of X and every extreme point of X in F is an extreme point of F.*

Proof. If x is an extreme point of F, then $x \in F$, so if $x \in \operatorname{ri}[x_1, x_2]$ with $x_1, x_2 \in X$, we have the $x_i \in F$. By hypothesis, there are no such x_i in F, so x is an extreme point of X.

If $x \in F$ is an extreme point of X, then x can't, in particular, lie in the relative interior of a segment with endpoints in F, and hence, x is an extreme point of F too. □

The following result is known as the Minkowski–Carathéodory Theorem.

Theorem 2.34. *Let X be a compact convex subset of $\mathbb{R}^d$. Then for every $x \in X$, there are extreme points $x_1, \dots, x_{d+1}$ (we allow repetition) so that $x \in \operatorname{conv}(\{x_1, \dots, x_{d+1}\})$. Moreover, x_1 may be picked arbitrarily.*

Proof. The proof is by induction on the dimension d of X. In the base case $d = 0$, X is a single point, so this is straightforward. Let now $x \in X$ be given.

If $x \in \operatorname{rb} X$, it lies in a proper face F which is of lower dimension (Lemma 2.28) and also compact (as a closed subset of X). The inductive hypothesis now gives us what we need, since x lies in the convex hull of at most d extreme points of F, which are also extreme points of X by Lemma 2.33. We can take x_1 to be any other extreme point of X.

Now suppose that $x \in \operatorname{ri} X$. From Exercise 2.18 (below), X has an extreme point, so let an extreme point x_1 of X be given. We then get (from compactness of X) that $\operatorname{aff}\{x, x_1\} \cap X$ is a segment $[x_1, y]$ with $y \in \operatorname{rb} X$. Lemma 2.29 says that y lies in a proper face F of X, which has dimension at most $d - 1$ (again, Lemma 2.28). By the inductive hypothesis, y lies in the convex hull of extreme points $x_2, \dots, x_{d+1}$ of F (make some of them the same, if necessary). These x_i are also extreme points of X (Lemma 2.33).

Since $x \in \operatorname{conv}\{x_1, y\}$ and $y \in \operatorname{conv}\{x_2, \dots, x_{d+1}\}$, we get that $x \in \operatorname{conv}\{x_1, \dots, x_{d+1}\}$ (by writing out convex combinations and plugging in). □

Exercises

Exercise 2.18. Show that every compact convex set X in $\mathbb{R}^d$ has an extreme point. ♣

Exercise 2.19. Show that $d + 1$ in Theorem 2.34 can't be improved. Also, give an example where it can. ♣

Exercise 2.20. Show that a convex polygon in the plane can be partitioned into triangles with *disjoint* interiors. ♣

If you have had topology, try this unexaminable exercise.

Exercise 2.21. Show that the convex hull of a compact set is compact. ♣

3. Polar duality

3.1. The duality map

This definition just restates Equation 1.11.

Definition 3.1. Let $x \in \mathbb{R}^d \setminus \{0\}$ be a point other than the origin. We define the hyperplane x^* by

$$x^* := \{y : \langle x, y \rangle = 1\}$$

The hyperplane x^* is called the *dual* of x.

Let $H \subseteq \mathbb{R}^d$ be an affine hyperplane $H = \{x : \ell(x) = 1\}$ for some $\ell \in \left(\mathbb{R}^d\right)^*$. Then $0 \notin H$ and $\ell(x) = \langle y, x \rangle$ for some $y \in \mathbb{R}^d$. We define the point

$$H^* := y$$

The point y is called the *dual* of H. ♡

We didn't define a dual for the origin or any hyperplane containing the origin (which is a *linear* subspace). In the case of points the problem is clear: since $\langle 0, y \rangle \neq 1$ for any $y \in \mathbb{R}^d$, we won't get a hyperplane. The problem for hyperplanes is a little more subtle. If $\mathbf{0} \in H$, then $H = \{x : \ell(x) = 0\}$. We still have $\ell(x) = \langle x, y \rangle$ for some $y \in \mathbb{R}^d$, so we could try to define H^* to be y; this doesn't work because y is only determined up to a non-zero scalar multiple.

Example 3.2. In $d = 2$, we can be more concrete about the duality map. For a point $\begin{pmatrix} A \\ B \end{pmatrix}$ we have

$$\begin{pmatrix} A \\ B \end{pmatrix}^* = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 : Ax + By = 1 \right\}$$

Similarly, for the line $H = \{\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 : Ax + By = 1\}$, we have

$$H^* = \begin{pmatrix} A \\ B \end{pmatrix}$$

So, for example, the point $\begin{pmatrix} 0 \\ 3 \end{pmatrix}$ dualizes to the line $\{y = \frac{1}{3}\}$. ♠

Exercises

Exercise 3.1. Suppose that H_1 and H_2 are parallel hyperplanes. What can you say about the points H_1^* and H_2^*? ♣

Exercise 3.2. Imagine we define a different duality map on $\mathbb{R}^2$ by

$$\begin{pmatrix} a \\ b \end{pmatrix}^\dagger := \{y = ax - b\} \qquad \text{and} \qquad \{y = ax - b\}^\dagger := \begin{pmatrix} a \\ b \end{pmatrix}$$

What are the degenerate cases we can't dualize using the map $x \mapsto x^\dagger$? ♣

3.2. A geometric digression

We defined the duality map in an algebraic way. There's some nice underlying geometry. Define $\mathbb{S}^{d-1} \subseteq \mathbb{R}^d$ to be the *unit sphere*

$$\mathbb{S}^{d-1} = \{x : ||x|| = 1\}$$

A point $x \in \mathbb{R}^d$ is outside $\mathbb{S}^{d-1}$ if $||x|| > 1$. In this case, x^* hits $\mathbb{S}^{d-1}$. We first notice that

$$\left\langle \frac{1}{||x||^2} x, x \right\rangle = \frac{1}{||x||^2} \langle x, x \rangle = 1$$

so $x' := \frac{x}{||x||^2} \in x^*$. Since x' is in the unit ball (which is bounded) any hyperplane (unbounded) containing x' needs to intersect $\mathbb{S}^{d-1}$.

Hence, we know there is an $a \in \mathbb{S}^{d-1} \cap x^*$. We can compute

$$\langle a, x - a \rangle = \langle a, x \rangle - \langle a, a \rangle = 1 - ||a||^2 = 0$$

It now follows that the line $\mathrm{aff}\{x, a\}$ is *tangent*[3] to the unit sphere, and, moreover, that any a where $\mathrm{aff}\{x, a\}$ is tangent to the unit sphere lies in $S^{d-1} \cap x^*$. Hence, x is the tip[4] of a conical cap containing the unit sphere any touching it at the intersection with x^*.

This means that polar duality also gives us an association

$$\{\text{circle patterns on } \mathbb{S}^2\} \qquad \longleftrightarrow \qquad \{\text{points sets outside } \mathbb{S}^2 \text{ in } \mathbb{R}^3\}$$

Exercises

Exercise 3.3. Explain in detail why $\mathrm{aff}\{a, x\}$ is tangent to the unit sphere above. ♣

Exercise 3.4. Suppose that x is on the unit sphere. What is the hyperplane x^*? ♣

Exercise 3.5. Imagine we defined x^* by

$$x^* = \{y : \langle y, x \rangle = \frac{1}{2}\}$$

How does this change the geometric picture we developed above? ♣

3.3. Duals of convex sets

Now we can explore a notion of a dual for a convex set.

Definition 3.3. Let $X \subseteq \mathbb{R}^d$. We define the *(polar) dual* X^* of X as

$$X^* = \left(\bigcap_{x \in (X \setminus \{0\})} (x^*)^- \right) \cup \{0\}$$

♡

[3] Meets at one point.

[4] In the jargon of classical geometry "pole".

Equivalently,

$$X^* = \{y : \langle x, y\rangle \leq 1 \text{ for all } x \in X\}$$

We didn't need to give the origin special treatment here, but we did above, since it doesn't have a dual hyperplane.

Lemma 3.4. *Let $X \subseteq \mathbb{R}^d$. Then X^* is, closed, convex and contains the origin.*

Proof. That X^* contains the origin is in the definition. For convexity, note that it is an intersection of closed convex sets. $\square$

Now we come to an important result that is intuitively apparent but has a subtle proof. This is the "double dual" theorem.

Theorem 3.5. *Let $X \subseteq \mathbb{R}^d$ be a compact, d-dimensional convex set that contains the origin. Then $X^{**} = X$.*

In other words, if we take the dual of the dual, we get back to where we started. The analogous statement in the linear setting is that $(X^\perp)^\perp = X$, which is straightforward. Unsurprisingly, we'll need the separation theorem here.

Proof. That $X \subseteq X^{**}$ needs no hypotheses on X. Let $x \in X$ be given. From the definition of X^*, for all $y \in X^*$,

$$\langle y, x\rangle = \langle x, y\rangle \leq 1$$

Hence $x \in X^{**}$.

The containment $X \supseteq X^{**}$ is more difficult. Since there are a lot of quantifiers, we note that a point $x' \notin X^{**}$ if there is a $y \in X^*$ so that

$$\langle x', y\rangle > 1$$

Now let $x' \in X^c$ be given. Since X is compact and convex, the separation theorem (Theorem 2.21) gives us a hyperplane

$$H = \{z : \ell(z) = \alpha\} \qquad (\ell \in \left(\mathbb{R}^d\right)^*, \alpha \in \mathbb{R})$$

so that $X \subseteq H^-$ and $x' \in H^+$, with both x' and X disjoint from H. Because $\mathbf{0} \in X$, we have $\mathbf{0} \notin H$. Hence, $\alpha \neq 0$. Again, using compactness and d-dimensionality of X, we have that ℓ achieves a non-zero maximum $M > 0$[5] on X.

After a suitable (positive) scaling of ℓ, $M = 1$, so $H^* \in X^*$. By construction $\ell(x') = \langle x', H^*\rangle > 1$, so $x' \notin X^{**}$. $\square$

Exercises

Exercise 3.6. Describe the dual of the cube $\square_d$. ♣

Exercise 3.7. Describe the dual of the unit ball in $\mathbb{R}^d$. ♣

[5] Because X is d-dimensional and $\ell \neq \mathbf{0}$, ℓ is non-constant on X.

4. The main theorem of polytopes

4.1. Polytopes

There are two natural notions of finitely described convex sets.

Definition 4.1. Let X be a closed convex subset of $\mathbb{R}^d$. Then:

- X is a *V-polytope* if it is the convex hull of a finite point set.
- X is an *H-polytope* if it is the intersection of finitely many half spaces.

♡

The definition of a V-polytope implies, from the Minkowski–Carathéodory Theorem, that it has finitely many extreme points. That these two notions are nearly equivalent is a fundamental fact. (It's not obvious but it might seem that way.)

Theorem 4.2. *Every V-polytope is an H-polytope, and every compact H-polytope is a V-polytope.*

We'll prove it in a moment. We should notice that this is consistent with our basic examples.

Example 4.3.
- The standard simplex Δ_d and the cube $\square_d$ are are compact and both H- and V-polytopes, and we worked out explicit descriptions of them both ways.
- The ball is compact and has infinitely many extreme points. The main theorem of polytopes tells us it is not an H-polytope either.
- $\mathbb{R}^d$ itself, affine subspaces, and half-spaces are finite intersections of closed half-spaces. They have no extreme points. Necessarily, from the main theorem of polytopes, they are all non-compact.

♠

The last examples show that we can't avoid requiring compactness for the H-polytopes in Theorem 4.2. It's also true that V-polytopes are compact.

Lemma 4.4. *A V-polytope X in $\mathbb{R}^d$ is compact.*

Unexaminable sketch. The idea is the same as the proof of Lemma 2.14. By definition X is the image of the compact set Δ_n under the continuous map

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \mapsto \alpha_1 x_1 + \cdots + \alpha_n x_n$$

which makes X compact as well. □

4.2. Roadmap to the proof

This is a big theorem, so we'll break up the proof. Here's an overview.

One direction is straightforward: we'll analyze the extreme points of compact H-polytopes and conclude there are finitely many. This gives us the "compact H-polytope $\Longrightarrow$ V-polytope" direction. The necessary global geometry is packaged up in the Minkowski-Carathéodory Theorem 2.34.

The other direction "V-polytope $\Longrightarrow$ compact H-polytope" is the "hard" one. Our approach is to reduce to the easy one via duality. Schematically, we show:

$$X \text{ is a } V\text{-ptpe} \quad \Rightarrow \quad X^* \text{ is a cpt. } H\text{-ptpe} \quad \Rightarrow \quad X^* \text{ is a } V\text{-ptpe} \quad \Rightarrow \quad X^{**} \text{ is a cpt. } H\text{-ptpe}$$

and then we are done with the double dual Theorem.

4.3. Extreme points of H-polytopes

Let's now characterize the extreme points of an H-polytope. We'll say that an H-polytope X is

$$X = \bigcap_{i=1}^{n} H_i^- \qquad (H_i^+ = \{y : \ell_i(y) \le \alpha_i\})$$

and we'll call the boundaries of the H_i^-, H_i.

Lemma 4.5. *Let X be an H-polytope in $\mathbb{R}^d$ and $x \in X$. Let $I \subseteq \{1, \dots, n\}$ be such that $x \in H_i$ iff $i \in I$. Then x is an extreme point of X if and only if $\cap_{i \in I} H_i = \{x\}$.*

Proof. If $i \in I$, then $\ell_i(x) = \alpha_i$, so if u is any vector so that $x \pm u \in X$, we must have

$$\ell_i(x) + \ell_i(u) \le \alpha_i \qquad \text{and} \qquad \ell_i(x) - \ell_i(u) \le \alpha_i$$

from which it follows that $\ell_i(u) = 0$. If $[x_1, x_2] \subseteq X$ is a segment with x its relative interior, we can take $u = \varepsilon(x_2 - x_1)$ with $\varepsilon > 0$ small enough to conclude that $[x_1, x_2] \subseteq H_i$. Since $i \in I$ was arbitrary, we conclude that, in fact,

$$[x_1, x_2] \subseteq \bigcap_{i \in I} H_i$$

If the r.h.s. is $\{x\}$, we have shown that x is extreme.

Otherwise, we can find $y \in \cap_{i \in I} H_i$ different from X. Set $u = y - x$. Certainly the segment $[x - \varepsilon u, x + \varepsilon u] \subseteq H_i \subseteq H_i^-$ for all $\varepsilon > 0$ and $i \in I$. For $j \in I^c$, we have $\ell_j(x) < \alpha_j$. Since I^c is finite, for a small enough $\varepsilon > 0$, the segment $[x - \varepsilon u, x + \varepsilon u]$ remains disjoint from each of the H_j, and hence in X. This shows that x is not extreme. □

Exercises

Exercise 4.1. Show that if X is a compact H-polytope, then any linear form ℓ attains its maximum on X at an extreme point. ♣

4.4. Dual of a V-polytope

Now we need to describe the dual of a V-polytope.

Lemma 4.6. *Let $X \subseteq \mathbb{R}^d$ be a V-polytope with the origin in its interior and d-dimensional affine span. Then X^* is a compact H-polytope with the origin in its interior.*

Proof. The last statement, that $\mathbf{0} \in X^*$ follows from Lemma 3.4 (and didn't need any properties of X). We get that X^* is closed in the same way.

Next we check that X^* is an H-polytope. To this end, suppose that $X = \operatorname{conv}\{x_1, \dots, x_n\}$. Since the origin is in the interior of X, we may assume that none of the x_i are the origin[6] Expanding the definition, we need to show that

$$\{0\} \cup \bigcap_{x \in X \setminus \{0\}} (x^*)^- = \{0\} \cup \bigcap_{i=1}^{n} (x_i^*)^-$$

Since the rhs is a less restrictive intersection, we get

$$\{0\} \cup \bigcap_{x \in X \setminus \{0\}} (x^*)^- \subseteq \{0\} \cup \bigcap_{i=1}^{n} (x_i^*)^-$$

For the other direction, we use the Minkowski-Carathéodory Theorem 2.34. If $\mathbf{0} \neq y \in X^*$, then

$$\langle y, x \rangle \leq 1 \qquad (\text{all } x \in X)$$

However, since $x \mapsto \langle y, x \rangle$ is a non-zero linear form (and X is full-dimensional), it takes its maximum over X on some proper face and hence at an extreme point of X (Exercise 4.1 below).

Hence, for any $x \in X$, and y in the rhs, we have

$$\langle y, x \rangle \leq \left(\max_{1 \leq i \leq n} \langle y, x_i \rangle \right) \leq 1$$

so $y \in X^*$. This is the other containment we needed to prove.

For compactness, we note that if X^* is non-compact, then by closedness and convexity, there is some $y \in X^*$ so that $\alpha y \in X^*$ for all $\alpha \geq 0$. The only way this could happen is if $\langle y, x_i \rangle \leq 0$ for all x_i, contradicting that $\mathbf{0}$ is in the interior of X. □

Exercise 4.2. Compare this proof to how we worked out the dual of the cube. ♣

4.5. Proof of Theorem 4.2

First suppose that X is a compact H-polytope, so it is the intersection of some closed half-spaces $\{H_i^-\}_{i=1}^n$. Theorem 2.34 says that X is the convex hull of its extreme points. We classified the extreme points of H-polytopes in Lemma 4.5, which said an extreme point must be the single point of intersection of some d of the H_i^-. Hence, there are at most $\binom{n}{d}$ extreme points of H. This is a finite number, so X is a V-polytope.

Now suppose that X is a V-polytope that is the convex hull of the points $\{x_1, \dots, x_n\}$. For the moment, assume that X is also d-dimensional and contains the origin in its interior. We showed

[6] Why? The justification is maybe subtle.

that X is compact in Lemma 4.4. Lemma 4.6 implies that X^* is a compact H-polytope with the origin in its interior. Hence, X^* is a V-polytope with the origin its interior by the other direction of theorem. Hence X^{**} is an H-polytope by another use of Lemma 4.6. Finally, $X^{**} = X$ by Theorem 3.5, making X an H-polytope.

If X is not full-dimensional or doesn't contain the origin in its interior, restrict to its affine span so that the relative interior becomes the interior, and then translate so that the origin is in the interior. Now we can repeat the same argument. □□

In the following definition we are now free to just talk about "polytopes". This next definition sets up some terminology about the faces of polytopes.

Definition 4.7. Let X be a k-dimensional polytope. Faces of X with dimension

- $k-1$ are called *facets*
- 0 are called *vertices*
- 1 are called *edges*
- $k-2$ are called *ridges*

♡

Exercises

Exercise 4.3. Describe the number of vertices and facets of

- The cube $\square_d$
- The *cross polytope* $\operatorname{conv}\{\pm e_j : 1 \le j \le d\} \subseteq \mathbb{R}^d$
- The standard simplex Δ_d

♣

Exercise 4.4. Extract from the proof of Theorem 4.2, an upper bound on the number of faces a compact d-dimensional polytope with n vertices can have. ♣

Exercise 4.5. For a vector $x \in \mathbb{R}^n$, say that $x \ge 0$ if x has all non-negative coordinates. Suppose that A is an $m \times n$ matrix. Prove one of

- $Ax = b$ with $x \ge 0$
- $A^\top y \ge 0$ with $\langle b, y \rangle < 0$

has a solution, but not both. ♣

4.6. First applications

Theorem 4.2 answers a lot of natural questions about polytopes that otherwise might seem difficult right away.

Example 4.8. Suppose that X and Y are V-polytopes in $\mathbb{R}^d$. Then, if non-empty, $X \cap Y$ is also a V-polytope. From Theorem 4.2, there are families of hyperplanes H_i and G_i so that

$$X = \bigcap_{i=1}^{n} H_i^- \qquad \text{and} \qquad Y = \bigcap_{j=1}^{m} G_j^-$$

and then

$$X \cap Y = \left(\bigcap_{i=1}^{n} H_i^-\right) \cap \left(\bigcap_{j=1}^{m} G_j^-\right)$$

is an H-polytope. Theorem 4.2 then tells us we get a V-polytope again.

If we had to figure out the vertices, this would have been more complicated. ♠

Example 4.9. Suppose that X is a compact d-dimensional H-polytope. Let $\pi : \mathbb{R}^d \to \mathbb{R}^{d-1}$ be the projection map that forgets the last coordinate, i.e., $\pi((\alpha_i)_{i=1}^d) = (\alpha_i)_{i-1}^{d-1}$. Then $\pi(X)$ is an H-polytope.

Again, this seems difficult directly, but we know from Theorem 4.2 that

$$X = \operatorname{conv}\{x_1, \dots, x_n\} \qquad (\text{some } x_1, \dots, x_n \in \mathbb{R}^d)$$

Since π is linear, we have

$$\pi(X) = \operatorname{conv}\{\pi(x_1), \dots, \pi(x_n)\}$$

which is a V-polytope and then, by Theorem 4.2, again an H-polytope. ♠

Like most interesting results, there are a number of ways to look at Theorem 4.2. We've used geometry and optimization, in the form of the separation theorem and Minkowski-Carathéodory. Alternatively, we could have proceeded in analogy to quantifier elimination in mathematical logic. Along this line, the difficult steps would be to establish the above facts via a direct argument.

Let's finish this section with a first step towards connecting to combinatorics.

Theorem 4.10. *Let X be a k-dimensional compact polytope in $\mathbb{R}^d$. Then X has faces of every dimension between -1 and k.*

Proof. As usual, the substantial case if when k is equal to d, and if not, we simply restrict to the affine span, and play the argument.

We argue by induction on d. The base $d = 0$ is easy because X consists of a single point. For the inductive step, let $d > 0$ and X a compact polytope with d-dimensional affine span in $\mathbb{R}^d$ be given. We have

$$X = \operatorname{conv}\{x_1, \dots, x_n\}$$

where each x_i is a vertex. Note that $n \geq d + 1 > 2$. As an extreme point, x_1 lies outside of

$$X' = \operatorname{conv}\{x_2, \dots, x_n\}$$

which is non-empty. Hence there is a hyperplane H disjoint from x_1 and X' that separates them. The hyperplane H has to meet the interior of X, since every neighborhood of x_1 contains interior points of X. It now follows that

$$P = X \cap H$$

has dimension $d-1$ (it contains an open subset of H). As X is compact, so is P, and it is also a polytope (as an intersection of a hyperplane and and H-polytope).

By the IH, P has faces of all dimensions $-1, \ldots, d-1$. Because H can't contain any face of X[7] each face F of P of dimension $0 \le j \le d-1$ must arise from the intersection of H with a face $\widehat{F}$ of X with dimension $j+1$. This closes the induction, since the collection of $\widehat{F}$ have all the dimensions except for 0 and -1, which clearly X has. □

Exercises

Exercise 4.6. Prove that every vertex of a d-dimensional compact polytope is is at least d facets. ♣

Exercise 4.7. Suppose that X is a d-dimensional compact polytope such that every vertex x lies in exactly d facets. These are called *simple polytopes*. Show that there are exactly $d-1$ edges that contain x. ♣

Exercise 4.8. Let X be a d-dimensional compact polytope and F_0 a vertex of X. Show that there is a sequence of faces F_i so that

$$\varnothing = F_{-1} \subseteq F_0 \subseteq F_1 \subseteq \cdots \subseteq F_{d-1} \subseteq F_d = X$$

where each F_i has dimension i. ♣

Exercise 4.9. A polytope is called *simplicial* if every facet is a simplex. Show that if the vertices of a polytope are in general position, then it is simplicial. ♣

Exercise 4.10. Do the vertices of a simplicial polytope need to be in general position? ♣

Exercise 4.11. Show that if X is a simplicial polytope and F is a facet of X, then every subset of vertices of F generates a face of X. ♣

Exercise 4.12. Show that a hyperplane H contains a facet of a k-dimensional polytope X if and only if H does not contain X and supports X at at least k affinely independent vertices. ♣

4.7. An algorithmic digression

The modern interest in polytopes arises, in no small part, from the following task:

Maximize a linear form ℓ over an H-polytope X

The usual presentation is as a *standard form linear program*

$$\text{maximize } \langle c, x\rangle, \text{ subject to } Ax = b \text{ and } x \ge \mathbf{0} \qquad (A \in \mathbb{R}^{m\times n}, c \in \mathbb{R}^n, b \in \mathbb{R}^m) \tag{4.1}$$

where "$x \ge \mathbf{0}$" is taken coordinate-wise.

The set of "feasible" x is an H-polytope, and, when it is compact, this is at least a finite problem. By the Main Theorem, we convert the feasible set to a V-polytope, list out the vertices, and then pick one with largest $\langle c, x\rangle$ value.

Two problems arise:

[7] Where did we just use the Main Theorem?

- We don't know how to do the conversion.
- The example of the cube shows that the number of vertices might be huge relative to the sizes of A, b, and c.

The *simplex method* is the following algorithm:

- Find any v vertex of the feasible set X.
- Enumerate the other vertices $v_1, \ldots, v_k$ connected to v via an edge in X.
- If $\ell(v) \geq \ell(v_i)$ for all $i \in [k]$ stop. Otherwise pick the j so that $\ell(v_j)$ is largest and continue the process from v_j.

The enumeration steps can be done efficiently by manipulating the matrix A, and we won't go into it here[8]. For us, what is more interesting is the proof of correctness, which is encapsulated in the following result. (We will need it later in a different context.)

Theorem 4.11. *Let X be a compact polytope in $\mathbb{R}^d$, $\ell \in \left(\mathbb{R}^d\right)^*$ a linear form. If x is a vertex of X so that $\ell(x) < \max_{y \in X} \ell(y)$, then there vertex w so that $\ell(w) > \ell(x)$ and $[x, w]$ is an edge of X.*

Proof. To set the notation, let x^* be a vertex of X where ℓ is maximized. One exists by the main theorem. Let x be a vertex as in the statement, so that $\ell(x) < \ell(x^*)$.

As in the proof of Theorem 4.10, we find a hyperplane H that separates x from all the other vertices of X and meets the interior. The vertices

$$y_1, \ldots, y_m$$

of $P = X \cap H$ correspond to the edges of X, which we denote

$$[x, w_1], \ldots, [x, w_m]$$

For a contradiction, suppose that $\ell(x) \geq \ell(w_i)$ for each $1 \leq i \leq m$. As y_i is in the relative interior of $[x, w_i]$, Lemma 2.10 implies that $\ell(x) \geq \ell(y_i)$ as well. For any $z \in P$, we then have

$$\ell(z) = \alpha_1 \ell(y_1) + \cdots + \alpha_m \ell(y_m) \leq (\alpha_1 + \cdots + \alpha_m)\ell(x) = \ell(x)$$

where we have used that the α_i are a convex combination.

Now consider the segment $[x, x^*]$. It must meet P because H separates x from x^* and X is convex. Let z^* be the point of intersection. A last application of Lemma Lemma 2.10 tells us that

$$\ell(z^*) \leq \ell(x) < \ell(z^*) < \ell(x^*)$$

which is the desired contradiction. □

[8] There are probably hundreds of books on this topic.

Exercises This series of exercises will be about spanning trees of the complete graph K_n. Fix $n \in \mathbb{N}$ and set $N = \binom{n}{2}$. We make an association between orderd pairs $1 \le i < j \le n$ and the set E of edges of K_n and also (by fixing an order) the set $E = \{1, 2, \dots, N\}$. For a subset $F \subseteq E$ define $\rho(F)$ to be one less than the number of vertices with an endpoint in F.

Exercise 4.13. Show that a set $F \subseteq E$ is acyclic iff for every

$$|F'| \le \rho(F') \qquad (\text{all } F' \subseteq F) \tag{4.2}$$

Conclude that F is a spanning tree in K_n iff (4.2) and, in addition, $|F| = n - 1$. ♣

Fixing an ordering of the edges gives an association between the N edges ij of K_n and elementary vectors / coordinates of $\mathbb{R}^N$. So, for a vector $x \in \mathbb{R}^N$ we can write

$$x = \sum_{ij \in E} \alpha_{ij} e_{ij}$$

For $\emptyset \ne F \subseteq E$, we now define the linear form $\ell_F \in \left(\mathbb{R}^N\right)^*$

$$\ell_F(x) = \sum_{ij \in F} \alpha_{ij} \tag{4.3}$$

(this adds up the coordinates of x corresponding to F) and

$$\mathbf{1}_F = \sum_{ij \in F} e_{ij} \tag{4.4}$$

(this is the "characteristic vector" of F).

We are now going to define

$$T = \operatorname{conv}\{\mathbf{1}_F : F \text{ is a spanning tree of } K_n\} \subseteq \mathbb{R}^N \tag{4.5}$$

and investigate its properties. Here are some basic ones in terms of defining inequalities.

Exercise 4.14. Show that, for all $x \in T$, $\ell_E(x) = n - 1$; i.e., that the coordinates sum to $n - 1$ and $x \ge \mathbf{0}$. ♣

Exercise 4.15. Show that, for all $x \in T$, and $F \subseteq E$,

$$\ell_F(x) \le \rho(F) \tag{4.6}$$

♣

We can get an upper bound on the dimension as follows.

Exercise 4.16. For each $2 \le i < j \le n$, define

$$F_{ij} = \{1i, 1j, ij\} \subseteq E \tag{4.7}$$

Show that the linear forms $\ell_{F_{ij}}$ are linearly independent. Conclude that the dimension of T is at most $n - 1$. ♣

Next we show that the characteristic vectors of spanning trees are vertices of T and get a lower bound on the dimension.

Exercise 4.17. Let F be a spanning tree of K_n. Show that $\mathbf{1}_F$ is a vertex of T. ♣

Exercise 4.18. Define a family of spanning trees $F_1, \ldots, F_n$ by

$$F_i = \{ij \in E \,:\, j \in [n] \setminus \{i\}\}$$

(so F_i has all $n-1$ the edges that i is an endpoint of). Show that the vectors

$$\mathbf{1}_{F_1}, \ldots, \mathbf{1}_{F_n}$$

are affinely independent, and conclude that the dimension of T is at least $n-1$. ♣

At this point we remark that the tree polytope T is "very flat". It has dimension $n-1$ in an ambient space that is $\binom{n}{2} \approx n^2$ dimensional. This often happens in combinatorial applications.

We will finish this set of exercises by working out the edges of T. Let F and G be spanning trees of K_n. We say that F and G are *related by a flip* if there are edges $ij \in F$ and $k\ell \in G$ so that

$$G = (F \setminus \{ij\}) \cup \{kl\} \tag{4.8}$$

Exercise 4.19. Show that, if F and G are related by a flip, then there is a unique cycle $C \subseteq F \cup \{kl\}$ and that $ij \in C$. ♣

Exercise 4.20. Let F be a spanning tree of K_n. For each $ij \in E \setminus F$ (the edges not in F), let $P_{ij} \subseteq F$ be the path between i and j in F. Show that the linear forms $\ell_{P_{ij}}$ are linearly independent and maximized over T at $\mathbf{1}_F$. ♣

Exercise 4.21. Show that, if F and G are spanning trees of K_n then $[\mathbf{1}_F, \mathbf{1}_G]$ is an edge of T iff F and G are related by a flip. ♣

Exercise 4.22. Conclude that any two spanning trees of K_n can be transformed into each other by a sequence of flips. ♣

As a last remark, the polytope T is an example of the "base polytope" of a matroid.

5. Two interesting polytopes

5.1. Cyclic polytopes

Cyclic polytopes turn out to maximize the number of faces in terms of vertices. To define them, we need start with a description of the vertices.

Definition 5.1. The *moment curve* in $\mathbb{R}^d$ is the image of the map $m_d : \mathbb{R} \to \mathbb{R}^d$ given by

$$m_d(t) = \begin{pmatrix} t \\ t^2 \\ \vdots \\ t^d \end{pmatrix}$$

♡

The moment curve gives an explicit construction of a general position point set.

Lemma 5.2. *Let $n \geq d+1$ and suppose that*

$$x_1 = m_d(t_1), \quad x_2 = m_d(t_2), \quad \dots, \quad x_n = m_d(t_n)$$

are points in $\mathbb{R}^d$ along the moment curve with $t_1 < t_2 < \cdots < t_n$. Then the x_i are in general position.

Proof. Supposing the contrary, then there is a hyperplane

$$H = \{y : \ell(y) = \alpha\} \qquad (\ell \in \left(\mathbb{R}^d\right)^* \text{ and } \alpha \in \mathbb{R})$$

and indices $1 \leq i_1 < i_2 < \cdots < i_{d+1} \leq n$ so that

$$\{x_{i_1}, \dots, x_{i_{d+1}}\} \subseteq H$$

In this case, the single-variable polynomial

$$P(t) = \ell(m_d(t)) - \alpha$$

has degree at most d and at least $d+1$ distinct roots. The fundamental theorem of algebra[9] then implies that P must be identically zero. But then ℓ and α are both zero, and so H is not a hyperplane. The resulting contradiction completes the proof. □

In fact, we can improve this with a more detailed analysis. First we need the idea of the orientation of a set of points.

Definition 5.3. Let

$$x = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{pmatrix} \in \mathbb{R}^d$$

[9] Here we just need that if P has degree $k \geq 0$ there are complex numbers $\beta_0, \dots, \beta_k$ so that $P(x) = \beta_0(x - \beta_1)\cdots(x - \beta_k)$

be a point. The *homogeneous coordinates* $\hat{x}$ of x is the vector

$$\hat{x} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_d \\ 1 \end{pmatrix} \in \mathbb{R}^{d+1}$$

♡

Homogeneous coordinates give a linear model of affine d-space by putting it at "height one" in $\mathbb{R}^{d+1}$. We can quickly check that

$$x_1, \dots, x_k \text{ are aff. indep.} \qquad \Longleftrightarrow \qquad \hat{x}_1, \dots, \hat{x}_k \text{ are lin. indep.} \tag{5.1}$$

The connection between determinants and linear independence lets us define an orientation on ordered lists of $d+1$ points in $\mathbb{R}^d$.

Definition 5.4. Let $x_1, \dots, x_{d+1}$ be points in $\mathbb{R}^d$. The *orientation* of $x_1, \dots, x_{d+1}$ is

$$\operatorname{sign}(-1)^d \det \begin{pmatrix} \hat{x}_1 & \cdots & \hat{x}_{d+1} \end{pmatrix}$$

where $\hat{x}_i$ are the homogeneous coordinates of the points x_i. If the x_i have orientation $+$ we say they are *positively oriented* and otherwise *negatively oriented*. If the x_i are not affinely independent, they don't have an orientation. ♡

Notice that the orientation depends on the order of the points. If we swap the order of x_1 and x_2 we will reverse the orientation.

Example 5.5. The points $0, e_1, \dots, e_d$ are positively oriented. We can check this by noting that in this case

$$\det \begin{pmatrix} \hat{\mathbf{0}} & \hat{e}_1 & \cdots & \hat{e}_d \end{pmatrix} = \det \begin{pmatrix} \mathbf{0} & I_d \\ 1 & \mathbf{1}^\top \end{pmatrix} = (-1)^d \det I_d = (-1)^d$$

where $\mathbf{0}$ denotes a column vector of d zeros and $\mathbf{1}$ a column vector of d ones. (This explains the $(-1)^d$ in the definition of orientation. It is there to make this example work out.) ♠

Example 5.6. In $\mathbb{R}^2$, positive orientation amounts to x_1, x_2, x_3 being a left turn. We can check this directly with $0, e_1, e_2$. This turns out to be the only substantial case, since the sign of the determinant used to compute the orientation isn't changed by translating, rotating, and scaling by a positive number.

For general x_1, x_2, x_3, we can translate so that $x_1 = 0$, and then rotate so that x_2 is on the positive x-axis. At this point, we can scale so that $x_2 = e_1$. Now, if x_3 has been transformed into $\begin{pmatrix} a \\ b \end{pmatrix}$ we have the orientation given by the sign of

$$\det \begin{pmatrix} 0 & 1 & a \\ 0 & 0 & b \\ 1 & 1 & 1 \end{pmatrix} = b$$

♠

Example 5.7. To understand the d-dimensional analogue of the previous example, consider the case where we have $x_1, \dots, x_d$ some fixed set of affinely independent points in $\mathbb{R}^d$ and $\mathbf{x}$ a vector of variables.

We know that there is a hyperplane $H = \operatorname{aff}\{x_1, \dots, x_d\}$, and so

$$\det\begin{pmatrix}\hat{x}_1 & \cdots & \hat{x}_d & \mathbf{x}\end{pmatrix} = 0 \qquad \iff \qquad \mathbf{x} \in H$$

and then the orientation determines when $\mathbf{x}$ is "above" or "below" H when the determinant does not vanish.

The key point about using the determinant is that it gives us a specific choice of which side of H is "up" that doesn't change if we translate or rotate and does change if we reflect or invert the x_i. A more naive approach, like saying that "below" is the side that contains the origin, runs in to trouble in this regard. ♠

Definition 5.8. Let $x_1, \dots, x_n$ be points in $\mathbb{R}^d$. We say that the x_i are *acyclic*[10] if for all orderd lists of indices $1 \le i_1 < i_2 < \cdots < i_{d+1} \le n$, the points $x_{i_1}, \dots, x_{i_{d+1}}$ are positively oriented (or they are all negatively oriented—the important thing is that the orientation is the same when the indices are in order). ♡

Lemma 5.9. *Take d, n, and the x_i as in Lemma 5.2. Then the point set x_i is acyclic.*

Proof assuming the Vandermonde determinant. We already proved that for indices $1 \le i_1 < i_2 < \cdots < i_{d+1} \le n$ that

$$\det(\hat{x}_{i_1} \cdots \hat{x}_{i_{d+1}}) \neq 0$$

and we need to show that they all have the same sign. More explicitly, we are interested in

$$\det\begin{pmatrix} t_{i_1} & \cdots & t_{i_{d+1}} \\ t_{i_1}^2 1 & \cdots & t_{i_{d+1}}^2 \\ \vdots & \vdots & \vdots \\ t_{i_1}^d & \cdots & t_{i_{d+1}}^d \\ 1 & \cdots & 1 \end{pmatrix} = (-1)^d \prod_{1 \le k < \ell \le d+1} (t_{i_\ell} - t_{i_k})$$

by the property of the Vandermonde determinant in the handout, since it's the Vandermonde matrix rotated by a cycle of length $d+1$. Because $(t_{i_\ell} - t_{i_k}) > 0$ for all k and ℓ, these determinants all have the same sign. Hence the x_i are acyclic. □

Definition 5.10. Let $x_1, \dots, x_n$ be $n \ge d+1$ points $x_i = m_d(t_i)$ be points along the moment curve in $\mathbb{R}^d$ with $t_i < t_{i+1}$ for all $1 \le i \le n-1$. We define the *cyclic polytope*

$$\operatorname{Cyc}(d, n) := \operatorname{conv}\{x_1, \dots, x_n\}$$

♡

Our next task is to investigate the face structure of $\operatorname{Cyc}(d, n)$.

[10]Sometimes called "pointed", but we will stick to "acyclic".

Definition 5.11. Let $X = \{x_1, \dots, x_n\}$ be points along the moment curve as in Definition 5.10. Let $F = \{x_{i_1}, x_{i_2}, \dots, x_{i_d}\} \subseteq X$ with $i_1 < i_2 < \cdots < i_d$. Then F satisfies the *evenness condition* if for all $x_k, x_\ell \in X \setminus F$ with $k < \ell$,

$$|\{x_{i_j} : k < i_j < \ell\}| \mod 2 = 0$$

i.e., the number of points in F between any two points not in F is even. ♡

Lemma 5.12. *Let $X = \{x_1, \dots, x_n\}$ be $n \geq d+1$ points along the moment curve in $\mathbb{R}^d$ as in Definition 5.10 and $\mathrm{Cyc}(d, n)$. Let $F = \{x_{i_1}, x_{i_2}, \dots, x_{i_d}\} \subseteq X$ with $i_1 < i_2 < \cdots < i_d$. Then $\operatorname{conv} F$ is a facet of $\mathrm{Cyc}(d, n)$ iff F satisfies the evenness condition.*

Proof. Because X is in general position, every facet will be the convex hull of a set of d points from X (Exercise 4.9). Hence, we need to show that $H = \operatorname{aff} F$ (which is a hyperplane by general position) supports $\operatorname{conv} X$ iff F satisfies the evenness condition. Since both of the half-spaces H^- and H^+ are convex and contain F, it will be sufficient to check that one of them contains X iff F satisfies the evenness condition.

To see this, we consider how the moment curve interacts with H. By construction, it meets it in the d points x_{i_j} of F. By the fundamental theorem of algebra, if the moment curve met H in any other point, it would be contained in H, which is a contradiction. Hence, H divides the moment curve into $d + 1$ "segments" which alternate between being "above" or "below" H and contain all the points on $X \setminus F$ in their interiors. For all of X to be on one side of H, the points of $X \setminus F$ need to be in either only "above" segments or "below" segments. This means that in between any two points of $X \setminus F$, if the moment curve passes one x_{i_j} in F, it must pass another one before meeting another points of $X \setminus F$. This is the evenness condition. □

Example 5.13. In $d = 2$ the only possibilities are $F = \{x_1, x_n\}$ and $F = \{x_i, x_{i+1}\}$.

We can also check directly that, in $2d$, the edges of a polygon are between "cyclically adjacent" vertices, so this definition generalizes the $2d$ case. ♠

Example 5.14. In the case of $d = 4$ and $n = 7$, we can highlight in blue some examples of facets, as seen by checking the evenness condition:

$$\begin{array}{ccccccc}
\color{blue}{x_1} & \color{blue}{x_2} & \color{blue}{x_3} & \color{blue}{x_4} & x_5 & x_6 & x_7 \\
\color{blue}{x_1} & \color{blue}{x_2} & \color{blue}{x_3} & x_4 & x_5 & x_6 & \color{blue}{x_7} \\
x_1 & x_2 & \color{blue}{x_3} & \color{blue}{x_4} & \color{blue}{x_5} & \color{blue}{x_6} & x_7 \\
x_1 & \color{blue}{x_2} & \color{blue}{x_3} & x_4 & \color{blue}{x_5} & \color{blue}{x_6} & x_7
\end{array}$$

Here are some failures of the evenness condition highlighted in red:

$$\begin{array}{ccccccc}
\color{red}{x_1} & \color{red}{x_2} & \color{red}{x_3} & x_4 & \color{red}{x_5} & x_6 & x_7 \\
\color{red}{x_1} & \color{red}{x_2} & x_3 & \color{red}{x_4} & x_5 & \color{red}{x_6} & x_7 \\
x_1 & x_2 & \color{red}{x_3} & \color{red}{x_4} & \color{red}{x_5} & x_6 & \color{red}{x_7}
\end{array}$$

♠

Our next goal is to count the number of facets of $\mathrm{Cyc}(d, n)$.

Theorem 5.15. *A cyclic polytope $\mathrm{Cyc}(d, n)$ has*

$$\binom{n - \lfloor d/2 \rfloor}{\lfloor d/2 \rfloor} + \binom{n - \lfloor d/2 \rfloor - 1}{\lfloor d/2 \rfloor - 1} \qquad \textit{facets if } d \textit{ is even}$$

and

$$2\binom{n-\lfloor d/2\rfloor-1}{\lfloor d/2\rfloor} \qquad \text{facets if } d \text{ is odd}$$

Before we give the proof, we notice that, in either case, the number of facets is $O(n^{\lfloor d/2\rfloor})$, which is *much* smaller than the naive n^d bound we got from our proof of the Main Theorem 4.2. Something much harder, that bears mentioning, even though we won't prove or use it, is:

Theorem 5.16. *Suppose that $X \subseteq \mathbb{R}^d$ is any polytope with n vertices. Then, for each $1 \le k \le d$ the number of faces of X of dimension k is at most the number of dimension k faces of* $\mathrm{Cyc}(d,n)$.

To prove Theorem 5.15, we start with a combinatorial result.

Lemma 5.17. *Let $n \ge 2k$ and arrange n black circles on a line. The number of ways to color $2k$ of the circles blue so that consecutive runs of blue circles are of even length is $\binom{n-k}{k}$.*

Proof. Since the runs of blue circles have even length, we can pair each blue circle with another one next to it. Replace each pair of blue circles with a single green one. This process is also reversible, so there is a bijection between the blue/black colourings we want to count and arrangements of $n-2k$ black circles and k green ones, arranged in any way.

This last counting problem is easy. There are $n-k$ total circles and we can pick any k of them to turn green for a total of $\binom{n-k}{k}$ ways to do it. □

Example 5.18. Two examples, with $n = 7$ and $k = 2$ of the transformation in the proof above are:

$$\color{blue}{\bullet\ \bullet\ \bullet\ \bullet}\ \color{black}{\bullet\ \bullet\ \bullet} \qquad \longleftrightarrow \qquad \color{green}{\bullet\ \bullet}\ \color{black}{\bullet\ \bullet\ \bullet}$$

and

$$\bullet\ \bullet\ \color{blue}{\bullet\ \bullet}\ \color{black}{\bullet}\ \color{blue}{\bullet\ \bullet} \qquad \color{black}{\longleftrightarrow} \qquad \bullet\ \bullet\ \color{green}{\bullet}\ \color{black}{\bullet}\ \color{green}{\bullet}$$

♠

Proof of Theorem 5.15 Combinatorially, the evenness condition translates into the following question

> *Let $n \ge d+1$ and arrange n black circles on a line. How many ways are there to color d of the circles blue so that the number of blue circles between any two black ones is even?*

This is similar to what we did in Lemma 5.17, except that runs of blue circles at the very beginning or end can now be odd. We are going to reduce this question to the other one.

First, let's assume that $d = 2k+1$ is odd. Since we need an odd number of blue circles but can only have odd length runs at the beginning / end, we have two types of configurations: those with an odd run at the beginning or at the end. In the case where the odd run is at the beginning, we remove the first (necessarily blue) circle to get all even runs. We conclude that there are $\binom{n-k-1}{k}$ of these. Symmetrically, if the odd run is at the end, we remove the last circle. This gives us $\binom{n-k-1}{k}$ more configurations. Finally, we notice that $k = \lfloor d/2\rfloor$ and get the formula.

If $d = 2k$ is even, there are now two possibilities, either both blue runs at the beginning are even length or both are odd length. If they are both even, we can use Lemma 5.17 right away to get $\binom{n-k}{k}$ possibilities. If they are both odd, we remove the first and last circles (necessarily blue). What's left is an arrangement of $n-2$ circles, $2(k-1)$ of which are blue with all even blue runs. From Lemma 5.17 there are $\binom{n-2-k+1}{k-1} = \binom{n-\lfloor d/2\rfloor-1}{\lfloor d/2\rfloor-1}$ possibilities for this. The two cases are disjoint, and the formula follows. □□

So far, we've used properties of the moment curve extensively. What about acyclic point sets that are not on the moment curve? They are still vertices of a polytope, which is combinatorially equivalent to a cyclic polytope.

Theorem 5.19. *Let $n \geq d+1$ and $X = \{x_1, \dots, x_n\}$ an acyclic point set in $\mathbb{R}^d$. Then* $\operatorname{conv} X$ *is combinatorially equivalent to* $\mathrm{Cyc}(d, n)$.

In other words, a subset of d of the x_i span a facet of the convex hull iff they satisfy the evenness condition. In the proof, we will use properties of determinants and orientation instead of the moment curve.

Proof. Let a set $F \subseteq X$

$$x_{i_1}, x_{i_2}, \dots, x_{i_d} \qquad \text{with} \qquad i_1 < i_2 < \cdots < i_d$$

of the points be given. By general position, $H = \operatorname{aff} F$ is a hyperplane. In Exercise 4.12, we showed that F will generate a fact of $\operatorname{conv} X$ iff all of X is in one of the half-spaces determined by H. As we discussed in Example 5.7, this happens iff the sign of

$$\det \begin{pmatrix} \hat{x}_{i_1} & \cdots & \hat{x}_{i_d} & \hat{x}_k \end{pmatrix}$$

is the same for all $x_k \in X \setminus F$. (Whether it is always positive or always negative doesn't matter to us.)

First assume that F satisfies the evenness condition. Suppose that, for some fixed k, we have

$$i_1 < i_2 < \cdots < i_\ell < k < i_{\ell+1} < \cdots < i_d$$

Because X is acyclic, we have

$$(-1)^d \det \begin{pmatrix} \hat{x}_{i_1} & \hat{x}_{i_2} & \cdots & \hat{x}_{i_\ell} & \hat{x}_k & \hat{x}_{\ell+1} & \cdots & \hat{x}_d \end{pmatrix} > 0$$

Since F satisfies the evenness condition, to get to

$$(-1)^d \det \begin{pmatrix} \hat{x}_{i_1} & \cdots & \hat{x}_d & \hat{x}_k \end{pmatrix}$$

x_k has to move to the right past some number of even length runs of points in F and possibly one odd-length run that must occur because $i_d = n$ (so every x_k will have to skip it). Since skipping past an even length run preserves the determinant, and skipping past and odd length run negates it, the sign of the above expression is does not depend on k. We conclude that F spans a face of $\operatorname{conv} X$.

On the other hand, if F does *not* satisfy the evenness condition, there are choices of x_k before and after some odd run. These choices will have different signs in the orientation determinant, and so be separated by the affine span of F. Hence F does not span a facet. □

Before we leave this section, a combinatorial remark. One reason to be interested in acyclic point sets is that they are "unavoidable" in the following way.

Lemma 5.20. *For any $k, d \in \mathbb{N}$, there is an $N \in \mathbb{N}$ so that if $n \geq N$, any point set $x_1, \dots, x_n$ in general position $\mathbb{R}^d$ contains an acyclic subset $x_{i_1}, \dots, x_{i_k}$ for some $1 \leq i_1 < \cdots < i_k \leq n$.*

Proof. This is an application of Ramsey's Theorem (see the handout). If we take a subset $I \subseteq \{1, \dots, n\}$, with $|I| = d + 1$, there is a unique way to order it so that the $i \in I$ are increasing. With this order, general position assigns I a "color" $+$ or $-$ based on its orientation. According to Ramsey's Theorem, if n is big enough, we can find a "homogeneous" subset $J \subseteq \{1, \dots, n\}$ with $|J| = k$ and all the cardinality $d+1$ subsets of J oriented the same way (when put in sorted order). This J is the acyclic set we want. Take N to be the lower bound from Ramsey's Theorem. □

The N we get here is not optimized in any way.

Exercises

Exercise 5.1. Using the left turn / right turn interpretation of orientation in dimension two, show that $x_1, \dots, x_n \in \mathbb{R}^2$ are the vertices of a convex polygon in order if and only if they are acyclic. ♣

Exercise 5.2. Show that every facet of $\mathrm{Cyc}(d, n)$ is a simplex. ♣

Exercise 5.3. Show that every subset of at most $\lfloor d/2 \rfloor$ vertices of $\mathrm{Cyc}(d, n)$ span a face of $\mathrm{Cyc}(d, n)$. ♣

5.2. The Permutohedron

The symmetric group $\mathrm{Sym}(d)$ acts by bijections on $\{1, 2, \dots, d\}$. We can define a polytope in $\mathbb{R}^d$ using the symmetric group.

Definition 5.21. Let $\sigma \in \mathrm{Sym}(d)$. We define the point

$$x_\sigma = \begin{pmatrix} \sigma(1) \\ \sigma(2) \\ \vdots \\ \sigma(d) \end{pmatrix}$$

The *permutohedron* Π_d is the polytope $\mathrm{conv}\{x_\sigma : \sigma \in \mathrm{Sym}(d)\}$. ♡

In this section, we'll look at the structure of the vertices, facets and faces of Π_d, which have nice combinatorial descriptions. We should start by computing the dimension.

Lemma 5.22. *The permutohedron Π_d is $(d-1)$-dimensional.*

Proof. The coordinates of x_σ add up to

$$1 + 2 + \cdots + d - 1 + d = \binom{d+1}{2}$$

for every sigma, which implies that the vertices all lie in a hyperplane, so the dimension is not more than $d - 1$.

For the other direction, we can do induction on dimension, since Π_1 consists of one point it has dimension zero. Now notice that $X = \mathrm{conv}\{x_\sigma : \sigma(d) = d\}$ has an affine image, by forgetting the last coordinate, that is bijective with Π_{d-1}. But no x_σ with $\sigma(d) = 1$ lies in X, so the dimension of Π_d must be strictly greater than that of Π_{d-1}. □

Now we can check that every x_σ is a vertex.

Theorem 5.23. *Every x_σ, with $\sigma \in \mathrm{Sym}(d)$, is a vertex of Π_d.*

The proof is conceptually nice, but has several steps.

Setup for Theorem 5.23 By the Main Theorem of polytopes, it is sufficient (and also necessary) to find a linear form that achieves its maximum on Π_d at x_σ (and only x_σ). We'll need to work with coordinates, so take a linear form ℓ to be

$$\ell(x) = \langle y, x \rangle \qquad y = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_d \end{pmatrix}$$

so that

$$\ell(x_\sigma) = \sum_{i=1}^{d} \sigma(i)\beta_i = \sum_{i=1}^{d} i\beta_{\sigma^{-1}(i)}$$

We say that ℓ is *sorted* (with respect to σ) if $\beta_{\sigma^{-1}(1)} \leq \dots \leq \beta_{\sigma^{-1}(d)}$ and *strictly sorted* (with respect to σ) if $\beta_{\sigma^{-1}(1)} < \dots < \beta_{\sigma^{-1}(d)}$. The meaning of being strictly sorted is that d is multiplied by the largest β_j in $\ell(x_\sigma)$, $d-1$ by the second largest, and so on.

To complete the proof, we will take ℓ as above and prove the key claim: *ℓ is maximized over Π_d at exactly x_σ iff is ℓ is strictly sorted.*

Strictly sorted $\Rightarrow$ unique maximizer x_σ Fix x_σ and suppose that ℓ is strictly sorted with respect to σ. We're going to show that $\ell(x_\sigma) > \ell(x)$ for any $x \in \Pi_d$ different from x_σ. By the Main Theorem, we only need to consider x_τ with $\mathrm{Sym}(d) \ni \tau$ different from σ.

The idea is simply to show that, starting from such a τ, we can increase the value of ℓ by multiplying τ by a transposition that makes it more like σ.

First, assume that $\tau^{-1}(d) \neq \sigma^{-1}(d)$. Now define $\rho = \tau(\tau^{-1}(d)\ \sigma^{-1}(d))$. We can compute

$$\rho^{-1}(d) = \left((\tau^{-1}(d)\ \sigma^{-1}(d))\tau^{-1}\right)(d) = \sigma^{-1}(d)$$

and

$$\rho^{-1}(\tau(\sigma^{-1}(d))) = \tau^{-1}(d)$$

otherwise, ρ^{-1} and τ^{-1} act the same way.

Now we compute

$$\ell(x_\rho) - \ell(x_\tau) = \sum_{i=1}^{d} i(\beta_{\rho^{-1}(i)} - \beta_{\tau^{-1}(i)})$$

all the terms on the r.h.s. are zero except for

$$d(\beta_{\rho^{-1}(d)} - \beta_{\tau^{-1}(d)}) + \tau(\sigma^{-1}(d))(\beta_{\rho^{-1}(\tau(\sigma^{-1}(d)))} - \beta_{\tau^{-1}(\tau(\sigma^{-1}(d)))})$$

by the computations we did above, this is equal to

$$d(\beta_{\sigma^{-1}(d)} - \beta_{\tau^{-1}(d)}) + \tau(\sigma^{-1}(d))(\beta_{\tau^{-1}(d)} - \beta_{\sigma^{-1}(d)}) = (d - \tau(\sigma^{-1}(d)))(\beta_{\sigma^{-1}(d)} - \beta_{\tau^{-1}(d)})$$

Because ℓ was strictly sorted with respect to σ,

$$\beta_{\sigma^{-1}(d)} - \beta_{\tau^{-1}(d)} > 0$$

and because $\tau^{-1}(d) \neq \sigma^{-1}(d)$

$$d \neq \tau(\sigma^{-1}(d)) \qquad \Longrightarrow \qquad d > \tau(\sigma^{-1}(d))$$

Since both terms are strictly positive, we showed that x_τ isn't a maximizer of ℓ.

It now follows that any maximizer $x_{\tau'}$ of ℓ over Π_d puts d in the same place σ does. By induction, x_σ is the unique maximizer.

Unique maximizer $x_\sigma \Rightarrow$ strictly sorted Now we fix ℓ and suppose that x_σ is the unique point of Π_d maximizing ℓ. To see that ℓ must be strictly sorted with respect to σ we first suppose, for contradiction, that there is some $1 \leq j < d$ so that $\beta_{\sigma^{-1}(j)} \geq \beta_{\sigma^{-1}(d)}$. Our strategy is similar to the other direction, except now σ plays the role of τ: we will show we can get a non-negative increase in ℓ by multiplying σ by an appropriate transposition.

This time, we define $\rho = \sigma(\sigma^{-1}(d)\ \sigma^{-1}(j))$. We check that

$$\rho^{-1}(d) = \sigma^{-1}(j) \qquad \text{and} \qquad \rho^{-1}(j) = \sigma^{-1}(d)$$

at which point similar computations to the ones above tell us that

$$\ell(x_\rho) - \ell(x_\sigma) = (\beta_{\sigma^{-1}(j)} - \beta_{\sigma^{-1}(d)})(d - j) \geq 0$$

where we get only a non-negative change because we have allowed $\beta_{\sigma^{-1}(j)} = \beta_{\sigma^{-1}(d)}$.

It now follows that for x_σ to be the unique maximizer, $\beta_{\sigma^{-1}(d)}$ must be the largest β_i, and the result follows by induction $\square\square$

Our proof has actually showed us quite a bit more. Since any face F of a polytope X can be described at the set of maximizers of some linear form ℓ_F over X, we have a recipe for constructing all the possible ℓ_F and, hence, describing the faces.

As a warmup, let's describe the edges of Π_d.

Lemma 5.24. *Let $\sigma, \tau \in \mathrm{Sym}(d)$. Then $\mathrm{conv}\{x_\sigma, x_\tau\}$ is an edge of the permutohedron Π_d if and only if $\tau = \sigma(\sigma^{-1}(i)\ \sigma^{-1}(i+1))$ for some $1 \leq i \leq d-1$.*

In words, this says that τ and σ are related by swapping the positions of consecutive values. For example if σ 13452 and $i = 1$, then τ would be 23451.

Proof. An argument nearly identical to the proof of Theorem 5.23 tells us that an ℓ that is sorted with respect to both σ and τ and no other x_ρ is maximized on $\mathrm{conv}\{x_\sigma, x_\tau\}$. If σ and τ are related as in the lemma, it is easy to construct such an ℓ. Pick the β_i so that

$$\beta_{\sigma^{-1}(1)} < \cdots < \beta_{\sigma^{-1}(i)} = \beta_{\sigma^{-1}(i+1)} < \beta_{\sigma^{-1}(i+2)} < \cdots < \beta_{\sigma^{-1}(d)}$$

which is clearly sorted with respect to σ. We also have $\tau^{-1}(j) = \sigma^{-1}(j)$ except for $\tau^{-1}(i+1) = \sigma^{-1}(i)$ and $\tau^{-1}(i) = \sigma^{-1}(i+1)$. So

$$\beta_{\tau^{-1}(1)} < \cdots < \beta_{\tau^{-1}(i+1)} = \beta_{\tau^{-1}(i)} < \beta_{\tau^{-1}(i+2)} < \cdots < \beta_{\sigma^{-1}(d)}$$

and hence ℓ is also sorted with respect to τ.

For the other direction, notice that the requirement of ℓ being sorted with respect to some ρ determines all the values of ρ except for $\rho(i)$ and $\rho(i+1)$. This leaves only σ and τ as the possible choices. $\square$

To give a general description of the face, we need a definition.

Definition 5.25. Let $\mathcal{P} = (P_1, \dots, P_k)$ be an *ordered partition* of $\{1, \dots, d\}$ into $1 \le k \le d$ parts. This means that

$$\{1, \dots, d\} = \bigcup_{i=1}^{k} P_i \qquad \text{with } P_i \cap P_j = \varnothing \text{ for } i \neq j$$

and that we regard $(P_1, P_2, \dots, P_k)$ as different than, e.g., $(P_2, P_1, \dots, P_k)$.

Keep the notation for ℓ and β_i of this section. We say that ℓ is *compatible* with an ordered parition $\mathcal{P}$ if $\beta_i = \beta_j$ when $i, j \in P_t$ for some $1 \le t \le k$; and when $i \in P_t$ and $j \in P_m$ with $t \neq m$, $\beta_i < \beta_j$ if $t < m$ and $\beta_i > \beta_j$ if $t > m$. ♡

Example 5.26. To get an idea of how ordered paritions work, in Lemma 5.24, the linear form ℓ is compatible with the partition

$$\{\{\sigma^{-1}(j)\} : j \in \{1, \dots, d\} \setminus \{i, i+1\}\} \bigcup \{\{\sigma^{-1}(i), \sigma^{-1}(i+1)\}\}$$

and also sorted with respect to the two permutations σ and τ.

Working backwards, we could instead simply pick an ordered partition $\mathcal{P}$ with $d-1$ parts, which would necessarily have the right form, and then read off σ^{-1}, and then σ. Lemma 5.24 would then apply to identity a unique edge of Π_d. ♠

Theorem 5.27. *There is a bijection between ordered paritions of $\{1, \dots, d\}$ and faces of the permutohedron Π_d. Moreover, if $\mathcal{P}$ has k parts, then the associated face has dimension $d - k$.*

Notice that we've already done this with vertices ($\mathcal{P}$ has d parts) and edges ($\mathcal{P}$ has $d-1$ parts). At this point, notice that facets, which have dimension $d-2$ (because Π_d is only $d-1$ dimensional), will be associated with ordered partitions that have two parts.

Proof. Given $\mathcal{P}$, construct a linear form ℓ that is compatible with it. We've already proved that ℓ achieves its maximum value over Π_d at the vertices

$$X = \{x_\sigma : \sigma \in \mathrm{Sym}(d) \text{ with } \ell \text{ sorted w.r.t. } \sigma\}$$

and hence on $F = \operatorname{conv} X$, which is the face we want, using the Main Theorem of polytopes.

For the dimension of F, we need to work out the dimension of the ℓ we can choose. Notice that if ℓ and ℓ' are compatible with $\mathcal{P}$, they both have coordinates that are constant over the parts of $\mathcal{P}$. Hence,

$$\{\ell \in \left(\mathbb{R}^d\right)^d : \ell \text{ is compatible with } \mathcal{P}\}$$

is a convex set. It has dimension k, since we can pick k different values freely, subject to to the inequality between parts, for the β_i and then the rest are forced. We then conclude that $\operatorname{aff} F$ is the intersection of k independent affine hyperplanes, so it has dimension $d - k$. □

Exercises

Exercise 5.4. Show that every vertex of Π_d is in $d-1$ edges. Then show that there are $\frac{1}{2}d!(d-1)$ edges in Π_d. ♣

Exercise 5.5. Show that the permutohedron Π_d has $2^d - 2$ facets. ♣

Exercise 5.6. Let $\alpha_1 \le \alpha_2 \le \cdots \alpha_n$ and $\beta_1 \le \beta_2 \le \cdots \beta_n$ be real numbers. Prove that

$$\alpha_1\beta_n + \alpha_2\beta_{n-1} + \cdots + \alpha_{n-1}\beta_2 + \alpha_n\beta_1 \le \alpha_1\beta_{\sigma(1)} + \cdots + \alpha_n\beta_{\sigma(n)} \le \alpha_1\beta_1 + \alpha_2\beta_2 + \cdots + \alpha_{n-1}\beta_{n-1} + \alpha_n\beta_n$$

for all $\sigma \in \mathrm{Sym}(n)$. ♣

6. Combinatorial convexity

6.1. Helly's Theorem

Suppose that we have a large family $X_1, \dots, X_n$ of convex subsets of $\mathbb{R}^d$. What is a sufficient condition for the intersection $\cap_{i=1}^n X_i$ to be non-empty?

In the simple case where the X_i are affine subspaces, if the intersection is empty, then some subset of at most $d+1$ of the X_i has empty intersection, since the dimension of the intersection drops by at least one as we add to our collection.

We can generalize this a lot. The following result is known as Helly's Theorem.

Theorem 6.1. *Let $X_1, \dots, X_n$ be convex subsets of $\mathbb{R}^d$. Then if the intersection $\cap_{i=1}^n X_i = \varnothing$, there is some $I \subseteq \{1, \dots, n\}$ with $|I| \le d+1$ so that $\cap_{i \in I} X_i = \varnothing$.*

The somewhat clunky formulation is just to take care of edge cases where n is small. Really, we'd like to say that the intersection is non-empty iff all subsets of $d+1$ of the X_i have non-empty intersection, but this doesn't quite work.

The proof uses two lemmas.

Lemma 6.2. *Let $d \ge 2$. If $x_1, \dots, x_{d+2}$ are points in $\mathbb{R}^d$ in general position, then some d of the x_i span a hyperplane that (strictly) separates the other two.*

Proof. By general position, $X = \operatorname{conv}\{x_1, \dots, x_{d+1}\}$ is a simplex.

If $x_{d+2} \notin X$, the (using the Main Theorem) one the planes supporting a facet of X (non-strictly) separates x_{d+2} from the rest of the x_i. This facet doesn't contain all $d+1$ of the other x_i by general position.

If $x_{d+2} \in X$, then the hyperplane spanned by $x_3, \dots, x_{d+2}$ can't support X, so it must separate x_1 and x_2. By general position, this separation is strict. □

Lemma 6.3 (Radon's Theorem). *Let $n \ge d+2$ and $X = \{x_1, \dots, x_n\}$ be a finite set of points in $\mathbb{R}^d$. Then there is a (necessarily non-empty) $I \subseteq \{1, \dots, n\}$ so that* $\operatorname{conv}\{x_i : i \in I\} \cap \{x_i : i \in I^c\}$ *is non-empty.*

Proof. The $n \ge d+2$ statement follows from the $n = d+2$ case, since convex combinations allow zero coefficients. We find the partition of $\{x_1, \dots, x_{d+2}\}$ and then allocate the rest of the points arbitrarily. From now on, we assume that $n = d+2$.

The proof is by induction on the dimension d. For the base $d = 0$, the only possibility is two points that are the same.

Now let $d \ge 2$ be given. If X is not in general position, there is a subset $X' \subseteq X$ of size $d+1$ that lie in a $(d-1)$-dimensional affine hyperplane H. Fix a bijective affine map $T : H \to \mathbb{R}^{d-1}$. The IH applies to $T(X')$ and then, since T^{-1} sends convex combinations to convex combinations, whatever partition of X' we obtain will work inside of H.

From now on, we further assume that X is in general position, so that Lemma 6.2 applies. After renumbering, we may then assume that $H = \operatorname{aff}\{x_3, \dots, x_{d+2}\}$ separates x_1 and x_2, with neither x_1 or x_2 in H.

Let $x_0 = H \cap [x_1, x_2]$. Applying the inductive hypothesis to $\{x_0, x_3, \dots, x_{d+2}\}$ (identifying H with $\mathbb{R}^{d-1}$ as before), we get an $I \subseteq \{0, 3, \dots, d+2\}$ so that $\operatorname{conv}\{x_i : i \in I\}$ and $\operatorname{conv}\{x_i : i \in I^c\}$ intersect; we may assume $0 \in I$. But then, since $x_0 \in \operatorname{conv}\{x_1, x_2\}$, we may close the induction by replacing 0 by $\{1, 2\}$ in I. □

A comment about this proof: x_0 does *not* need to be in $\operatorname{conv}\{x_3, \dots, x_{d+2}\}$. Whether it is or not controls what the size of I we get in the inductive hypothesis.

Proof of Theorem 6.1 We'll prove the contrapositive statement. That is, we assume that all subsets of the X_i of size at most $d+1$ have non-empty intersection and show that all the X_i do. If $n \le d+1$, this is true right away.

The proof is by induction on n. We checked the base cases of $n \le d+1$ above. Now assume that $n \ge d+2$. By the inductive hypothesis, for each each $1 \le i \le n$, there is a

$$y_i \in \bigcap_{j \in \{1,\dots,n\}\setminus\{i\}} X_i$$

Lemma 6.3 gives us a $J \subset \{1, \dots, n\}$ so that there is a

$$z \in \operatorname{conv}\{y_j \,:\, j \in J\} \cap \operatorname{conv}\{y_j \,:\, j \in J^c\}$$

Now let $j \in J$ be given. By construction, y_j is in each of the X_i except maybe X_j. In particular, since $j \in J$, we must have $y_j \in X_i$ for all $i \in J^c$. Since j was arbitrary, for all $j \in J$ and $i \in J^c$, $y_j \in X_i$.

Fix now an $i \in J^c$. Since X_i is convex, we get

$$\operatorname{conv}\{y_j \,:\, j \in J\} \subseteq X_i$$

and, since $i \in J^c$ was arbitrary:

$$\operatorname{conv}\{y_j \,:\, j \in J\} \subseteq \bigcap_{i \in J^c} X_i$$

Switching the roles of J and J^c we now get

$$\operatorname{conv}\{y_j \,:\, j \in J^c\} \subseteq \bigcap_{i \in J} X_i$$

It now follows that z is in all of the X_i, so

$$\varnothing \ne \bigcap_{i=1}^{n} X_i$$

which closes the induction. □□

Exercises

Exercise 6.1. Describe a family of $d+2$ subsets of $\{1, \dots, d+2\}$ so that every $d+1$ of them have non-empty intersection, but the intersection of all $d+2$ is empty. ♣

Exercise 6.2. Suppose that $n \ge d+2$ and that $X_1, \dots, X_n$ is a family of sets in $\mathbb{R}^d$. Let K be a convex subset of $\mathbb{R}^d$.

Show that if for each $I \subset \{1, \dots, n\}$, with $|I| = d+1$ there is a $v_I \in \mathbb{R}^d$ so that

$$K + v_I \supseteq \bigcup_{i \in I} X_i$$

then there is an $x \in \mathbb{R}^d$ so that

$$K + x \supseteq \bigcup_{i=1}^{n} X_i$$

♣

Exercise 6.3. Suppose that $n \geq d + 2$ and that $X_1, \ldots, X_n$ is a family of convex sets in $\mathbb{R}^d$. Let K be a convex subset of $\mathbb{R}^d$.

Show that if for each $I \subset \{1, \ldots, n\}$, with $|I| = d + 1$ there is a $v \in \mathbb{R}^d$ so that

$$K + v \cap X_i \neq \varnothing \qquad (\text{all } i \in I)$$

then there is an $x \in \mathbb{R}^d$ so that

$$K + x \cap X_i \neq \varnothing \qquad (1 \leq i \leq n)$$

♣

Exercise 6.4. Suppose that $n \geq 3$. Show that if $X_1, \ldots, X_n$ are parallel line segments in the plane $\mathbb{R}^2$ and there is a line intersecting every three of the X_i there is a line intersecting all of them. ♣

This one is non-examinable but interesting if you have had topology or analysis.

Exercise 6.5. Show that if $X_1, X_2, \ldots$ is an infinite collection of *compact* convex sets in $\mathbb{R}^d$, then if $\cap_{i \in \mathbb{N}} X_i = \varnothing$ some $d + 1$ of the X_i have empty intersection. ♣

6.2. Centerpoint Theorem

If $x_1 \leq x_2 \leq \cdots, \leq x_n$ are points in $\mathbb{R}^1$, there is a good notion of the most "central" point with respect to the x_i: the median. Unlike the mean, the median splits the x_i into two equally-sized subsets, and doesn't change if we, e.g, send $x_n \to \infty$. One higher-dimensional notion of a median is:

Definition 6.4. Let $X = \{x_1, \ldots, x_n\}$ be a finite point set in $\mathbb{R}^d$. A point $x \in \mathbb{R}^d$ is a *centerpoint* of X if any closed half-space H^+ containing x also contains at least $\frac{n}{d+1}$ of the x_i. ♡

Notice that for $d = 1$ a centerpoint is just the median. We also can't do better than $n/(d + 1)$ by looking at $d + 1$ points in general position in $\mathbb{R}^d$.

It was relatively easy to see that the median exists, and to find it. Even the first statement is non-trivial for higher dimensions. That's our next theorem.

Theorem 6.5. *Let $X = \{x_1, \ldots, x_n\}$ be a finite subset of $\mathbb{R}^d$. Then X has at least one centerpoint.*

It's helpful to reformulate the notion of centerpoint.

Lemma 6.6. *Let $X = \{x_1, \ldots, x_n\}$ be a finite point set in $\mathbb{R}^d$. If x is not a centerpoint of X, then there is an $I \subseteq \{1, \ldots, n\}$ with $|I| > (1 - \frac{1}{d+1})n$ so that $x \notin \operatorname{conv}\{x_i : i \in I\}$.*

Proof. If some closed half-space H^+ containing X has $|X \cap H^+| < \frac{1}{d+1}n$, then the complementary open half-space $\operatorname{int} H^-$ must contain more than $(1 - \frac{1}{d+1})n$ points of X. The hyperplane H already separates $\operatorname{conv}\{x_i : x_i \in \operatorname{int} H^-\}$ from x. □

Proof of Theorem 6.5 Consider the sets

$$X_I = \text{conv}\{x_i \,:\, i \in I\} \qquad (\text{all } I \subseteq \{1, \dots, n\} \text{ with } |I| > \frac{dn}{d+1})$$

The X_I are all convex. Moreover, since each of the X_I contains more than $\frac{dn}{d+1}$ points of X, any $d+1$ of them must intersect (on some x_i, from elementary double counting[11]). Helly's Theorem provides a point

$$x \in \bigcap_I X_I$$

Lemma 6.6 tells us this x must be a centerpoint. □□

Exercises

Exercise 6.6. Let $X_1, \dots, X_k$ be finite subsets of $\mathbb{R}^d$ with $1 \leq k \leq d$. Show that there is a $(k-1)$-dimensional affine subspace A of $\mathbb{R}^d$ so that any closed half-space containing A contains at least $\frac{1}{d+1}|X_i|$ points of each X_i. ♣

[11]Indeed, if we have $d+1$ of these I, say $I_1, \dots, I_{d+1}$ then $\sum_{j=1}^{d+1} |I_j| > (d+1)\frac{dn}{d+1} = dn$. Since $\sum_{j=1}^{d+1} |I_j| = \sum_{i=1}^{n} |\{j \,:\, i \in I_j\}|$, we conclude that there must be an i with $|\{j \,:\, i \in I_j\}| > d$. As $d+1$ is the only possibility, this i is in all the I_j.

7. Hyperplane arrangements

7.1. Arrangements

Definition 7.1. Let $H_1, \dots, H_n$ be affine hyperplanes in $\mathbb{R}^d$. The *arrangement* $\mathcal{A}(\{H_i\}_{i=1}^n)$ is the subdivision of $\mathbb{R}^d$ into polytopes induced by the H_i. We call the d-dimensional polytopes in the subdivision *regions* and the 0-dimensional ones *vertices*; in general, we call the polytops in an arrangement *faces*. ♡

Note that the vertices must be vertices of some region. Similarly the faces are faces of some region.

This defintion makes sense, since we can fix a *sign vector* $\varepsilon \in \{+, -\}^n$ and then consider the H-polytopes

$$\bigcap_{i=1}^n H_i^{\varepsilon_i}$$

These polytopes won't necessarily be compact (or non-empty), but certainly every point $x \in \mathbb{R}^d$ is in at least one of them, since you can obtain a sign vector by parameterising

$$H_i = \{x : \ell_i(x) = \alpha_i\} \qquad (\ell_i \in \left(\mathbb{R}^d\right)^* \text{ and } \alpha_i \in \mathbb{R})$$

and then get a sign vector by the signs of $\ell_i(x) - \alpha_i$.

For point sets we have general position as a non-degeneracy assumption. The analogous notion (which will turn out to be directly related) for arrangements is:

Definition 7.2. An arrangement $\mathcal{A}$ of n hyperplanes $\{H_i\}$ in $\mathbb{R}^d$ is called *simple* if

$$\bigcap_{i \in I} H_i = \{x_I\} \qquad (\text{all } I \subseteq \{1, \dots, n\} \text{ with } |I| = d)$$

for some $x_I \in \mathbb{R}^d$ and

$$\bigcap_{i \in I} H_i = \varnothing \qquad (\text{all } I \subseteq \{1, \dots, n\} \text{ with } |I| = d + 1)$$

♡

Informally, this says that the intersections are all of the expected dimension and all exist. When $d = 2$, the hyperplanes are lines, and this says none of them are parallel and also that no three meet at the same point.

Another kind of arrangement that's useful, since it models linear subspaces is

Definition 7.3. An arrangement $\mathcal{A}$ of n hyperplanes $\{H_i\}$ in $\mathbb{R}^d$ is called *central* if

$$\bigcap_{i=1}^n H_i = \{x\}$$

for some $x \in \mathbb{R}^d$. ♡

We can always translate x to the origin to study the combinatorics of central arrangements using linear hyperplanes.

7.2. Arrangement complexity

The first question we address is the "complexity" of hyperplane arrangements in terms of how many regions the arrangement divides $\mathbb{R}^d$ into. There are 2^n possible sign vectors, but we'll see that this is not a good estimate of the number of regions.

Theorem 7.4. *Let $\mathcal{A}$ be a simple arrangement of n affine hyperplanes $\{H_i\}_{i=1}^n$ in $\mathbb{R}^d$. Then $\mathcal{A}$ divides $\mathbb{R}^d$ into*

$$\binom{n}{d} + \binom{n}{d-1} + \cdots + \binom{n}{2} + \binom{n}{1} + 1$$

d-dimensional H-polytopes.

The regions aren't all compact, so they aren't necessarily V-polytopes. In fact, most of them need not be.

Example 7.5. Let $x_1, \dots, x_{d+1}$ be in general position in $\mathbb{R}^d$. If we define

$$H_i = \operatorname{aff}\{x_j : i \neq j \in \{1, \dots, d+1\}\}$$

we end up with a simple arrangement with

$$2^{d+1} - 1 = \sum_{k=0}^{d} \binom{d+1}{k}$$

non-empty regions, of which only one is bounded, namely the convex hull of the x_i. ♠

We're going to give two different proofs for Theorem 7.4. The first is based on the optimization ideas we have used so far.

Proof I of Theorem 7.4. Denote by $f_d(n)$ the number of regions in a simple arrangement $\mathcal{A}$ of n affine hyperplanes in $\mathbb{R}^d$. We will declare $f_0(n) = 1$ for all n (there is an issue that simple arrangments need $n = 1$ for $d = 0$, so we are setting a convention).

For $d \geq 1$ we will fix $n \geq 1$ and use induction on d. The base case $n = 1$ gives $f_d(n) = n + 1$, since any n distinct points give a simple arrangement and divide the line $\mathbb{R}^1$ into $n + 1$ distinct intervals.

Now consider $d \geq 2$ and let a simple arrangmenet $\mathcal{A}$ of n hyperplanes. We know that $\mathcal{A}$ determines $\binom{n}{d}$ vertices. The main step is to find a covector ℓ so that, at each vertex x of $\mathcal{A}$, ℓ is maximized over exactly one of the regions of $\mathcal{A}$. We leave the proof unexaminable[12].

Once we have selected ℓ, see that there are exactly $\binom{n}{d}$ regions of $\mathcal{A}$ on which ℓ is bounded. To count the rest, we pick an $\alpha \in \mathbb{R}$ strictly greater than the maximum of ℓ over any vertex of $\mathcal{A}$. The intersection of $\mathcal{A}$ with the hyperplane $H = \{x : \ell(x) = \alpha\}$ is a simple arrangement of n hyperplanes in the $d - 1$ dimensional affine space H. By the IH, it has $f_{d-1}(n)$ regions, which correspond bijectively to the regions of $\mathcal{A}$ on which ℓ is unbounded.

[12] To see this, let a vertex x of $\mathcal{A}$ be given. As $\mathcal{A}$ is simple, there are hyperplanes $H_{i_1}, \dots, H_{i_d}$ in $\mathcal{A}$ that contain x. Thinking about the induced central arrangement $\mathcal{A}(H_{i_1}, \dots, H_{i_d})$, it divides $\mathbb{R}^d$ into 2^d cones C_j, and ℓ is maximized over some C_i at x and not maximized over any other C_k iff $-\ell^*$ lies in the interior of C_i.

As the union of the interiors of the C_i are open and dense in $\mathbb{R}^d$, the set of ℓ so that x is maximized at x over exactly one region that contains it is open and dense in $(\mathbb{R}^d)^*$. Intersecting these open dense sets over a finite number of vertices gives an open and dense set of "good" ℓ. We can pick any one of them, so a "randomly chosen" ℓ will have the property we need.

Having counted the bounded and unbounded regions (with respect to ℓ), we get that

$$f_d(n) = \binom{n}{d} + f_{d-1}(n) = \sum_{d=0}^{d} \binom{n}{d}$$

$\square$

When an arrangement isn't simple, the number of regions can only go down.

Corollary 7.6. *Simple arrangements maximize the number of regions.*

Proof. Simple arrangements maximize the number of vertices, because every subset of d hyperplanes in $\mathcal{A}$ gives a different vertex. The number of vertices is the only contribution to the sum as we go up in dimension. $\square$

For the second proof, we will introduce two operations on arrangements.

Definition 7.7. Let $\mathcal{A}$ be an arrangement of n hyperplanes $H_1, \dots, H_n$ in $\mathbb{R}^d$. The *deletion* $\mathcal{A}'$ of $\mathcal{A}$ (by deleting H_i) is the arrangement $\mathcal{A}'$ consisting of $H_1, \dots, H_{i-1}, H_{i+1}, \dots, H_n$. ♡

Informally, we just threw out H_i. This was boring. The next operation is a bit more interesting.

Definition 7.8. Let $\mathcal{A}$ be an arrangement of n affine hyperplanes $H_1, \dots, H_n$ in $\mathbb{R}^d$. The *contraction* of $\mathcal{A}$ around H_i is the arrangement

$$\mathcal{A}' = \mathcal{A}(\{H_i \cap H_j \,:\, j \in \{1, \dots, i-1, i+1, \dots, n\}\})$$

We can identify H_i with $\mathbb{R}^{d-1}$ after an invertible affine map, so $\mathcal{A}'$ is modelled by an arrangement of $n-1$ affine hyperplanes in $\mathbb{R}^{d-1}$. ♡

There are some technical issues with this definition when $\mathcal{A}$ is not simple. For example, there can be repeats among the $H_i \cap H_j$, even if all the H_j are distinct. However, they won't cause us any problems here.

Proof II of Theorem 7.4. We'll establish that

$$f_d(n) = f_d(n-1) + f_{d-1}(n-1)$$

via a "deletion contraction" argument. We start with $\mathcal{A}$ a simple arrangement of the hyperplanes $H_1, \dots, H_n$. Let $\mathcal{A}_0$ by the deletion of $\mathcal{A}$ from deleting H_1; let $\mathcal{A}_1$ by the contraction of $\mathcal{A}$ around H_1. Both of these arrangements have $n-1$ hyperplanes. Each region of $\mathcal{A}_0$ is either missed completely by H_1 or cut into two by it (justifying this is an exercise). Hence, each region of $\mathcal{A}_1$ corresponds (uniquely) to a new region that appears when we add H_1 to $\mathcal{A}_0$ (to get back $\mathcal{A}$). Since $\mathcal{A}_1$ is also a simple arrangment in $\mathbb{R}^{d-1}$, we get the desired recurrence.

Using the base $f_1(n) = n + 1$ and $f_d(0) = 1$, we can solve the recurrence using the Pascal identity:

$$f_d(n) = f_d(n-1) + f_{d-1}(n-1) =$$
$$\left[\sum_{i=0}^{d} \binom{n-1}{i}\right] + \left[\sum_{i=0}^{d-1} \binom{n-1}{i}\right] = \binom{n-1}{0} + \sum_{i=1}^{d-1} \left[\binom{n-1}{i-1} + \binom{n-1}{i}\right] = \sum_{i=0}^{d} \binom{n}{i}$$

$\square$

What's nice about this proof is that we can apply it to arrangements that are *not simple*. We don't get a formula in that case, but we do get a way of computing the number of regions.

Exercises

Exercise 7.1. Count the number of regions a simple central arrangement of n hyperplanes divides $\mathbb{R}^d$ into. ♣

Exercise 7.2. Describe a central hyperplane arrangment that is not simple and doesn't have duplicated hyperplanes. ♣

7.3. Arrangement duality

Given an arrangement $\mathcal{A}$ of hyperplanes $H_1, \dots, H_n$, none of which goes through the origin, we can *dualize* it to a point set by applying the polar duality map to each of the H_i. Now we investigate what happens to different combinatorial features of $\mathcal{A}$ under duality. Let us first make simple observation.

Lemma 7.9. *Let $x \in \mathbb{R}^d \setminus \{0\}$ and let $H = \{x : \ell(x) = 1\} \subseteq \mathbb{R}^d$ be a hyperplane not containing the origin. Then*

$$x \in H^- \qquad \Longleftrightarrow \qquad H^* \in (x^*)^-$$

where H^- is the closed half-space determined by H that contains the origin.

Proof. This follows from the symmetry in the definition of the duality map. Suppose that

$$H = \{z : \langle y, z \rangle = 1\}$$

Then both statements are equivalent to

$$\langle x, y \rangle \leq 1$$

□

By symmetry of H^+ and H^-, we have:

Corollary 7.10. *Let $x \in \mathbb{R}^d \setminus \{0\}$ and let $H \subseteq \mathbb{R}^d$ be a hyperplane not containing the origin. Then $x \in H$ if and only if $x^* \ni H^*$.*

Example 7.11. Suppose that $x, y \in \mathbb{R}^2 \setminus \{0\}$ and H is the line aff$\{x, y\}$. If we dualize everything, we see that H^* lies on the lines x^* and y^*.

To make the example more interesting, let z be a point collinear with x and y. In this case, H^* also lies on z^*, so we see that collinear points dualize to lines all meeting at a point, which was one of the things that we ruled out when we defined simple line arrangements. ♠

Example 7.12. Now consider an arrangement $\mathcal{A}$ of n affine hyperplanes $H_1, \dots, H_n$ in $\mathbb{R}^d$ that do not contain the origin. For convenience, denote by x_i the dual points H_i^*. Let $Y \subseteq \mathbb{R}^d$ be the region of $\mathcal{A}$ that contains the origin. From the convention that $H_i = \{y : \ell_i(x) \leq 1\}$, we have that

$$Y = \bigcap_{i=1}^{n} H_i^-$$

Lemma 7.9 then implies that

$$x_i \in (y^*)^- \qquad \text{for all } y \in Y \text{ and } 1 \leq i \leq n$$

Specializing this to the vertices of Y, Corollary 7.10 gives that the vertices of Y dualize to facets of the convex hull

$$X = \mathrm{conv}\{x_1, \dots, x_n\}$$

of the point set dual to the hyperplanes defining $\mathcal{A}$ and the hyperplanes H_j that support a facet of Y dualize to the vertices of X.

We can go in reverse as well, starting from a point set and getting an arrangement. So we have a correspondence

$$\{\, H_i \text{ supporting a facet of } Y \,\} \qquad \longleftrightarrow \qquad \{\text{vertices of } X\,\}$$

♠

What happened in this example should not be a surprise. If Y is compact, then X is just the polytope dual to Y, and we are back in the situation we looked at when we proved the Main Theorem 4.2. The new part of the analysis is that now we have a description of what happens to dual points the x_i that are not vertices of X: they correspond to planes H_i that don't support a facet of Y.

The preceding discussion, while nice, shows a disadvantage of using polar duality on arrangments: the special role played by the origin. To get around this, we define a different duality map, which is inspired by the one in Exercise 3.2.

Definition 7.13. Let $H \subseteq \mathbb{R}^d$ be a hyperplane

$$H = \{x : \ell(x) = \beta\} \qquad \text{with} \qquad \ell(e_d) = 1$$

We define

$$H^\dagger = \begin{pmatrix} \ell(e_1) \\ \ell(e_2) \\ \vdots \\ \ell(e_{d-1}) \\ -\beta \end{pmatrix}$$

and, for $x = (\alpha_i) \in \mathbb{R}^d$, we define a covector ℓ_x by

$$\ell_x(e_i) = \alpha_i \quad (1 \le i \le d-1) \qquad \text{and} \qquad \ell_x(e_d) = 1$$

and then

$$x^\dagger = \{y \in \mathbb{R}^d : \ell_x(y) = -\alpha_d\}$$

♡

Informally, we normalize $\ell(e_d)$ instead of the rhs of the hyperplane's equation. Under polar duality, linear hyperplanes don't have duals; with the new duality "vertical" hyperplanes don't have duals. An exercise asks you to exaplain more.

This † duality retains an important property.

Lemma 7.14. *Let $x \in \mathbb{R}^d$ and $H = \{x : \ell(x) = \alpha\}$ be such that $\ell(e_d) = 1$. Then*

$$x \in H^- \qquad \Longleftrightarrow \qquad H^\dagger \in (x^\dagger)^-$$

The proof explains the minus signs in Definition 7.13.

Proof. Suppose that $x = (\alpha_i)$. By definition, there are $\beta_1, \dots, \beta_{d-1}$ so that

$$\ell(x) = \left(\sum_{i=1}^{d-1} \alpha_i \beta_i\right) + \alpha_d$$

so, $x \in H^-$ if and only if

$$\left(\sum_{i=1}^{d-1} \alpha_i \beta_i\right) \leq \beta - \alpha_d$$

On the other hand, taking ℓ_x as in Definition 7.13 and the β_i as above, we compute

$$\ell_x(H^\dagger) = \left(\sum_{i=1}^{d-1} \beta_i \alpha_i\right) - \beta$$

and so $H^\dagger \in (x^\dagger)^-$ if and only if

$$\ell_x(H^\dagger) \leq \alpha_d \qquad \Longleftrightarrow \qquad \left(\sum_{i=1}^{d-1} \beta_i \alpha_i\right) - \beta \leq -\alpha_d$$

Rearranging the rhs shows the two statements are the same. □

Exercises

Exercise 7.3. Does the origin $\mathbf{0} \in \mathbb{R}^d$ have a well defined $\mathbf{0}^\dagger$? ♣

Exercise 7.4. Explain why we can't define $H^\dagger$ for a hyperplane $H = \{x : \ell(x) = \alpha\}$ when $\ell(e_d) = 0$. ♣

Exercise 7.5. Describe a set of lines H_1, H_2, H_3 in $\mathbb{R}^2$ so that $H_i^\dagger$ exists and is not in general position. Is the only way to do this to have H_1, H_2 and H_3 all meet at a point. ♣

Exercise 7.6. Suppose that $H_1, \dots, H_n$ are hyperpanes so that the point set $\dagger_1, \dots, H_n^\dagger$ is defined. Describe a condition on the dual point set that implies $\mathcal{A}(H_i)$ is simple. ♣

7.4. Levels in arrangements

Moving back to arrangements, we can define a notion of "depth" of a point.

Definition 7.15. Let $\mathcal{A}$ be an arrangement of n hyperplanes $H_1, \dots, H_n$ in $\mathbb{R}^d$, where $H_i = \{x : \ell_i(x) = 1\}$. The *level* of a point $x \in \mathbb{R}^d$ with respect to $\mathcal{A}$ is the number of H_i so that $x \in \operatorname{int} H_i^+$.

The k-th level of $\mathcal{A}$ is the set of $(d-1)$-dimensional faces of $\mathcal{A}$ with points having level exactly k. ♡

This definition is a little confusing because there is some asymmetry in it. Let us explore it in a little more detail.