I. Convexity

Convex sets

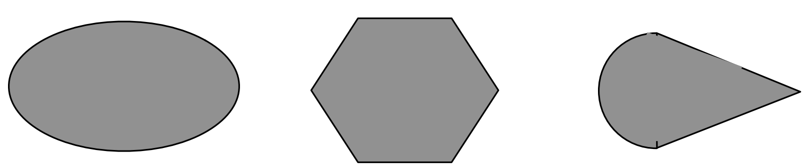
In this section, we will be introduced to some of the mathematical fundamentals of convex sets. In order to motivate some of the definitions, we will look at the *closest point problem* from several different angles. The tools and concepts we develop here, however, have many other applications both in this course and beyond.

A set $\mathcal{C} \subset \mathbb{R}^N$ is **convex** if

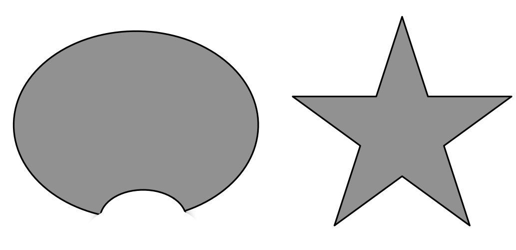
$$\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C} \quad \Rightarrow \quad (1-\theta)\boldsymbol{x} + \theta\boldsymbol{y} \in \mathcal{C} \quad \text{for all } \theta \in [0,1].$$

In English, this means that if we travel on a straight line between any two points in $\mathcal{C}$, then we never leave $\mathcal{C}$.

These sets in $\mathbb{R}^2$ are convex:



These sets are not:



Georgia Tech ECE 6270 Notes by M. Davenport and J. Romberg. Last updated 15:30, October 11, 2022

Examples of convex (and nonconvex) sets

- Subspaces. Recall that if $\mathcal{S}$ is a subspace of $\mathbb{R}^N$, then $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S} \Rightarrow a\boldsymbol{x} + b\boldsymbol{y} \in \mathcal{S}$ for all $a, b \in \mathbb{R}$.
  So $\mathcal{S}$ is clearly convex.
- Affine sets. Affine sets are just subspaces that have been offset by the origin:

$$\{\boldsymbol{x} \in \mathbb{R}^N \ : \ \boldsymbol{x} = \boldsymbol{y} + \boldsymbol{v}, \ \boldsymbol{y} \in \mathcal{T}\}, \quad \mathcal{T} = \text{subspace},$$

  for some fixed vector $\boldsymbol{v}$. An equivalent definition is that $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C} \Rightarrow \theta\boldsymbol{x} + (1-\theta)\boldsymbol{y} \in \mathcal{C}$ for all $\theta \in \mathbb{R}$ — the difference between this definition and that for a subspace is that subspaces must include the origin.
- Bound constraints. Rectangular sets of the form

$$\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^N : \ell_1 \leq x_1 \leq u_1, \ell_2 \leq x_2 \leq u_2, \ldots, \ell_N \leq x_N \leq u_N\}$$

  for some $\ell_1, \ldots, \ell_N, u_1, \ldots, u_N \in \mathbb{R}$ are convex.
- The *"filled in" simplex* in $\mathbb{R}^N$

$$\{\boldsymbol{x} \in \mathbb{R}^N : x_1 + x_2 + \cdots + x_N \leq 1, \ x_1, x_2, \ldots, x_N \geq 0\}$$

  is convex.
- Any subset of $\mathbb{R}^N$ that can be expressed as a set of linear inequality constraints

$$\{\boldsymbol{x} \in \mathbb{R}^N : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$$

  is convex. Notice that both rectangular sets and the simplex fall into this category — for the previous example, take

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \\ 0 & \cdots & & & -1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

In general, when sets like these are bounded, the result is a polyhedron.

- Norm balls. If $\|\cdot\|$ is a valid norm on $\mathbb{R}^N$, then

$$\mathcal{B}_r = \{\boldsymbol{x} \ : \ \|\boldsymbol{x}\| \leq r\},$$

  is a convex set.

- Ellipsoids. An ellipsoid is a set of the form

$$\mathcal{E} = \{\boldsymbol{x} \ : \ (\boldsymbol{x} - \boldsymbol{x}_0)^{\mathrm{T}} \boldsymbol{P}^{-1} (\boldsymbol{x} - \boldsymbol{x}_0) \leq r\},$$

  for a symmetric positive-definite matrix $\boldsymbol{P}$. Geometrically, the ellipsoid is centered at $\boldsymbol{x}_0$, its axes are oriented with the eigenvectors of $\boldsymbol{P}$, and the relative widths along these axes are proportional to the eigenvalues of $\boldsymbol{P}$.

- A single point $\{\boldsymbol{x}\}$ is convex.

- The empty set is convex.

- The set

$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1^2 - 2x_1 - x_2 + 1 \leq 0\}$$

  is convex. (Sketch it!)

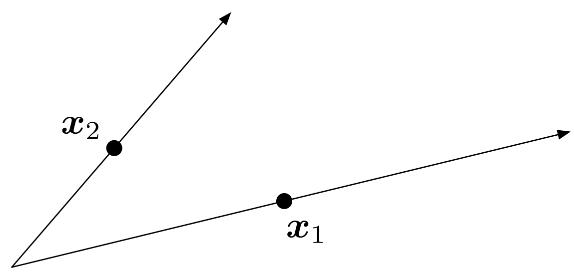
- The set
$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1^2 - 2x_1 - x_2 + 1 \geq 0\}$$
is **not** convex.
- The set
$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1^2 - 2x_1 - x_2 + 1 = 0\}$$
is certainly not convex.
- Sets defined by linear equality constraints where only some of the constraints have to hold are in general not convex. For example
$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1 - x_2 \leq -1 \ \text{ and } \ x_1 + x_2 \leq -1\}$$
is convex, while
$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1 - x_2 \leq -1 \ \text{ or } \ x_1 + x_2 \leq -1\}$$
is not convex.

Cones

A **cone** is a set $\mathcal{C}$ such that
$$\boldsymbol{x} \in \mathcal{C} \quad \Rightarrow \quad \theta\boldsymbol{x} \in \mathcal{C} \ \text{ for all } \theta \geq 0.$$

Convex cones are sets which are both convex and a cone. $\mathcal{C}$ is a convex cone if
$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C} \quad \Rightarrow \quad \theta_1\boldsymbol{x}_1 + \theta_2\boldsymbol{x}_2 \in \mathcal{C} \ \text{ for all } \theta_1, \theta_2 \geq 0.$$

Given an $\boldsymbol{x}_1, \boldsymbol{x}_2$, the set of all linear combinations with positive weights makes a wedge. For practice, sketch the region below that consists of all such combinations of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$:



We will mostly be interested in **proper cones**, which in addition to being convex, are closed, have a non-empty interior[1] ("solid"), and do not contain entire lines ("pointed").

Examples:

Non-negative orthant. The set of vectors whose entries are non-negative,

$$\mathbb{R}^N_+ = \{\boldsymbol{x} \in \mathbb{R}^N \;:\; x_n \geq 0, \text{ for } n = 1, \ldots, N\},$$

is a proper cone.

Positive semi-definite cone. The set of $N \times N$ symmetric matrices with non-negative eigenvalues is a proper cone.

Non-negative polynomials. Vectors of coefficients of non-negative polynomials on $[0, 1]$,

$$\{\boldsymbol{x} \in \mathbb{R}^N \;:\; x_1 + x_2 t + x_3 t^2 + \cdots + x_N t^{N-1} \geq 0 \text{ for all } 0 \leq t \leq 1\},$$

form a proper cone. Notice that it is not necessary that all the $x_n \geq 0$; for example $t - t^2$ $(x_1 = 0, x_2 = 1, x_3 = -1)$ is non-negative on $[0, 1]$.

Norm cones. The subset of $\mathbb{R}^{N+1}$ defined by

$$\{(\boldsymbol{x}, t),\; \boldsymbol{x} \in \mathbb{R}^N,\; t \in \mathbb{R} \;:\; \|\boldsymbol{x}\| \leq t\}$$

[1]See Technical Details for precise definition.

is a proper cone for any valid norm $\|\cdot\|$ and $t > 0$. (We encountered this cone in our discussion about SOCPs in the last set of notes.)

Every proper cone $\mathcal{K}$ defines a **partial ordering** or **generalized inequality**. We write

$$\boldsymbol{x} \preceq_{\mathcal{K}} \boldsymbol{y} \quad \text{when} \quad \boldsymbol{y} - \boldsymbol{x} \in \mathcal{K}.$$

For example, for vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$, we say

$$\boldsymbol{x} \preceq_{\mathbb{R}^N_+} \boldsymbol{y} \quad \text{when} \quad x_n \leq y_n \ \text{ for all } n = 1, \ldots, N.$$

For symmetric matrices $\boldsymbol{X}, \boldsymbol{Y}$, we say

$$\boldsymbol{X} \preceq_{S^N_+} \boldsymbol{Y} \quad \text{when} \quad \boldsymbol{Y} - \boldsymbol{X} \text{ has non-negative eigenvalues.}$$

We will typically just use $\preceq$ when the context makes it clear. In fact, for $\mathbb{R}^N_+$ we will just write $\boldsymbol{x} \leq \boldsymbol{y}$ (as we did above) to mean that the entries in $\boldsymbol{x}$ are component-by-component upper-bounded by the entries in $\boldsymbol{y}$.

Partial orderings obey share of the properties of the standard $\leq$ on the real line. For example:

$$\boldsymbol{x} \preceq \boldsymbol{y}, \;\; \boldsymbol{u} \preceq \boldsymbol{v} \quad \Rightarrow \quad \boldsymbol{x} + \boldsymbol{u} \preceq \boldsymbol{y} + \boldsymbol{v}.$$

But other properties do not hold; for example, it is not necessary that either $\boldsymbol{x} \preceq \boldsymbol{y}$ or $\boldsymbol{y} \preceq \boldsymbol{x}$. For an extensive list of properties of partial orderings (most of which will make perfect sense on sight) can be found in [BV04, Chapter 2.4].

Affine sets

Recall the definition of a linear subspace: a set $\mathcal{T} \subset \mathbb{R}^N$ is a subspace if

$$\boldsymbol{x}, \boldsymbol{y} \in \mathcal{T} \quad \Rightarrow \quad \alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{T}, \;\; \text{for all } \alpha, \beta \in \mathbb{R}.$$

Affine sets (also referred to as affine spaces) are not fundamentally different than subspaces. An affine set $\mathcal{S}$ is simply a subspace that has been offset from the origin:

$$\mathcal{S} = \mathcal{T} + \boldsymbol{v}_0,$$

for some subspace $\mathcal{T}$ and $\boldsymbol{v}_0 \in \mathbb{R}^N$. (It thus make sense to talk about the dimension of $\mathcal{S}$ as being the dimension of this underlying subspace.) We can recast this as a definition similar to the above: a set $\mathcal{S} \subset \mathbb{R}^N$ is affine if

$$\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S} \quad \Rightarrow \quad \lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \mathcal{S}, \;\; \text{for all } \lambda \in \mathbb{R}.$$

Just as we can find the smallest subspace that contains a finite set of vector $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$ by taking their span,

$$\operatorname{Span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}) = \left\{\boldsymbol{x} \in \mathbb{R}^N \; : \; \boldsymbol{x} = \sum_{k=1}^{K} \alpha_k \boldsymbol{v}_k, \;\; \alpha_k \in \mathbb{R}\right\},$$

we can define the **affine hull** (the smallest affine set that contains the vectors) as

$$\operatorname{Aff}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}) = \left\{\boldsymbol{x} \in \mathbb{R}^N \; : \; \boldsymbol{x} = \sum_{k=1}^{K} \lambda_k \boldsymbol{v}_k, \; \lambda_k \in \mathbb{R}, \; \sum_{k=1}^{K} \lambda_k = 1\right\}.$$

Example: Let

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}.$$

Then $\text{Span}(\{\boldsymbol{v}_1, \boldsymbol{v}_2\})$ is all of $\mathbb{R}^2$ while $\text{Aff}(\{\boldsymbol{v}_1, \boldsymbol{v}_2\})$ is the line that connects $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$,

$$\text{Aff}(\{\boldsymbol{v}_1, \boldsymbol{v}_2\}) = \left\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1 + x_2 = 1\right\}.$$

Just as any linear subspace $\mathcal{T}$ of dimension K can be described using a homogeneous set of equations,

$$\boldsymbol{x} \in \mathcal{T} \ \Leftrightarrow \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0},$$

using any $(N-K) \times N$ matrix $\boldsymbol{A}$ whose nullspace is $\mathcal{T}$, any affine set $\mathcal{S}$ of dimension K can be described as the solution to a linear system of equations

$$\boldsymbol{x} \in \mathcal{S} \ \Leftrightarrow \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

for some $(N-K) \times N$ matrix $\boldsymbol{A}$ and $\boldsymbol{b} \in \mathbb{R}^{N-K}$.

It should be clear that every subspace is an affine set, but not every affine set is a subspace. It is easy to show that an affine set is a subspace if and only if it contains the $\boldsymbol{0}$ vector.

Affine sets are of course convex.

Hyperplanes and halfspaces

Hyperplanes and halfspaces are both very simple constructs, but they will be crucial to our understanding to convex sets, functions, and optimization problems.

A **hyperplane** is an affine set of dimension $N-1$; it has the form

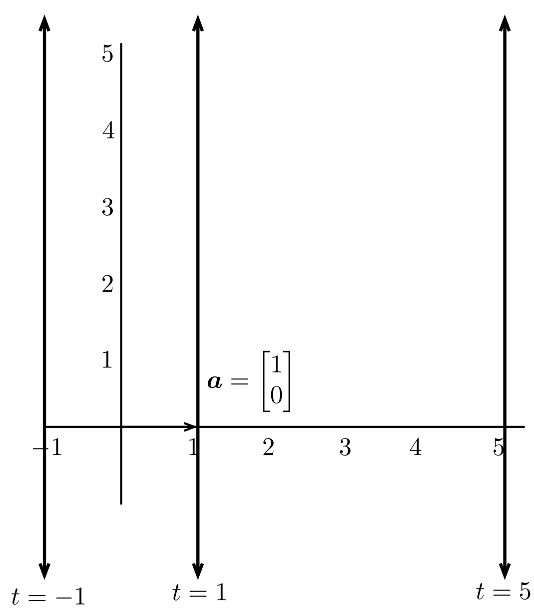
$$\{\boldsymbol{x} \in \mathbb{R}^N \ : \ \langle \boldsymbol{x}, \boldsymbol{a} \rangle = t\}$$

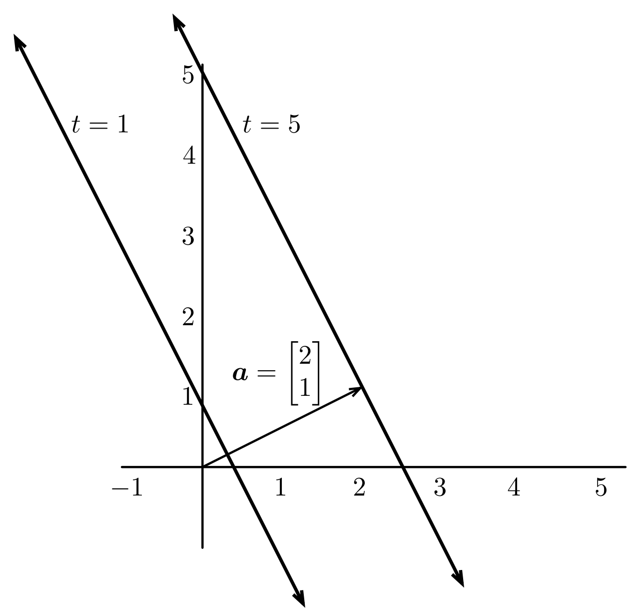
for some fixed vector $\boldsymbol{a} \neq \mathbf{0}$ and scalar t. When $t = 0$, this set is a subspace of dimension $N - 1$, and contains all vectors that are orthogonal to $\boldsymbol{a}$. For $t \neq 0$, this is an affine space consisting of all the vectors orthogonal to $\boldsymbol{a}$ (call this set $\mathcal{A}^\perp$) offset to some $\boldsymbol{x}_0$:

$$\{\boldsymbol{x} \in \mathbb{R}^N \ : \ \langle \boldsymbol{x}, \boldsymbol{a} \rangle = t\} = \{\boldsymbol{x} \in \mathbb{R}^N \ : \ \boldsymbol{x} = \boldsymbol{x}_0 + \mathcal{A}^\perp\},$$

for any $\boldsymbol{x}_0$ with $\langle \boldsymbol{x}_0, \boldsymbol{a} \rangle = t$. We might take $\boldsymbol{x}_0 = t \cdot \boldsymbol{a}/\|\boldsymbol{a}\|_2^2$, for instance. The point is, $\boldsymbol{a}$ is a **normal vector** of the set.

Here are some examples in $\mathbb{R}^2$:

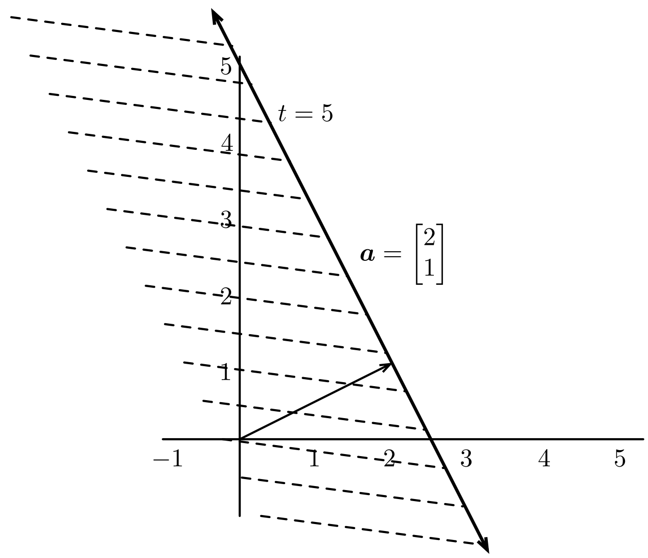




A **halfspace** is a set of the form

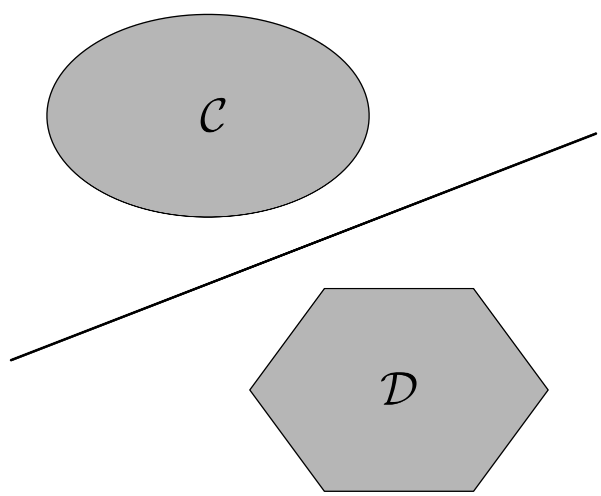
$$\{\boldsymbol{x} \in \mathbb{R}^N \ : \ \langle \boldsymbol{x}, \boldsymbol{a} \rangle \leq t\}$$

for some fixed vector $\boldsymbol{a} \neq \mathbf{0}$ and scalar t. For $t = 0$, the halfspace contains all vectors whose inner product with $\boldsymbol{a}$ is negative (i.e. the angle between $\boldsymbol{x}$ and $\boldsymbol{a}$ is greater than $90°$). Here is a simple example:

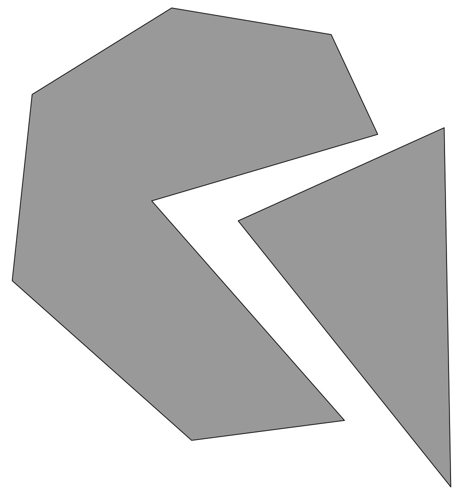


Separating hyperplanes

If two convex sets are disjoint, then there is a hyperplane that separates them. Here is a picture:



This fact is intuitive, and is incredibly useful in understanding the solutions to convex optimization programs (we will see this even in the next section). It is also not true in general if one of the sets is nonconvex; observe:



For sets $\mathcal{C}, \mathcal{D} \subset \mathbb{R}^N$, we say that a hyperplane $\mathcal{H} = \{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{a} \rangle = t\}$

- *separates* $\mathcal{C}$ and $\mathcal{D}$ if for all $\boldsymbol{c} \in \mathcal{C},\ \boldsymbol{d} \in \mathcal{D}$
$$\langle \boldsymbol{c}, \boldsymbol{a} \rangle \leq t \leq \langle \boldsymbol{d}, \boldsymbol{a} \rangle \quad \text{for all} \quad \boldsymbol{c} \in \mathcal{C},\ \boldsymbol{d} \in \mathcal{D}; \tag{1}$$
- *properly separates* $\mathcal{C}$ and $\mathcal{D}$ if (1) holds and both $\mathcal{C}$ and $\mathcal{D}$ are not contained in $\mathcal{H}$ themselves;
- *strictly separates* $\mathcal{C}$ and $\mathcal{D}$ if
$$\langle \boldsymbol{c}, \boldsymbol{a} \rangle < t < \langle \boldsymbol{d}, \boldsymbol{a} \rangle \quad \text{for all} \quad \boldsymbol{c} \in \mathcal{C},\ \boldsymbol{d} \in \mathcal{D};$$
- *strongly separates* $\mathcal{C}$ and $\mathcal{D}$ if there exists $\epsilon > 0$ such that
$$\langle \boldsymbol{c}, \boldsymbol{a} \rangle \leq t - \epsilon \quad \text{and} \quad \langle \boldsymbol{d}, \boldsymbol{a} \rangle \geq t + \epsilon \quad \text{for all} \quad \boldsymbol{c} \in \mathcal{C},\ \boldsymbol{d} \in \mathcal{D}.$$

Note that we can switch the roles of $\mathcal{C}$ and $\mathcal{D}$ above, i.e. we also say $\mathcal{H}$ separates $\mathcal{C}$ and $\mathcal{D}$ if $\langle \boldsymbol{d}, \boldsymbol{a} \rangle \leq t \leq \langle \boldsymbol{c}, \boldsymbol{a} \rangle$ for all $\boldsymbol{c} \in \mathcal{C}, \boldsymbol{d} \in \mathcal{D}$.

Let us start by showing the following:

Strong separating hyperplane theorem
Let $\mathcal{C}$ and $\mathcal{D}$ be disjoint nonempty closed convex sets and let $\mathcal{C}$ be bounded. Then there is a hyperplane that strongly separates $\mathcal{C}$ and $\mathcal{D}$.

To prove this, we show how to explicitly construct a strongly separating hyperplane. Let $d(\boldsymbol{x}, \mathcal{D})$ be the distance of a point $\boldsymbol{x}$ to the set $\mathcal{D}$:

$$d(\boldsymbol{x}, \mathcal{D}) = \inf_{\boldsymbol{y} \in \mathcal{D}} \|\boldsymbol{x} - \boldsymbol{y}\|_2.$$

As we will see below, since $\mathcal{D}$ is closed, there is a unique closest point to $\boldsymbol{x}$ that achieves the infimum on the right. It is also true that $d(\boldsymbol{x}, \mathcal{D})$ is continuous as a function of $\boldsymbol{x}$, so by the Weierstrauss extreme value theorem it achieves its minimum value over the compact set $\mathcal{C}$. That is to say, there exist points $\boldsymbol{c} \in \mathcal{C}$ and $\boldsymbol{d} \in \mathcal{D}$ that achieve the minimum distance

$$\|\boldsymbol{c} - \boldsymbol{d}\|_2 = \inf_{\boldsymbol{x} \in \mathcal{C}, \boldsymbol{y} \in \mathcal{D}} \|\boldsymbol{x} - \boldsymbol{y}\|_2.$$

Since $\mathcal{C}$ and $\mathcal{D}$ are disjoint, we have $\boldsymbol{c} - \boldsymbol{d} \neq 0$.

Define

$$\boldsymbol{a} = \boldsymbol{d} - \boldsymbol{c}, \quad t = \frac{\|\boldsymbol{d}\|_2^2 - \|\boldsymbol{c}\|_2^2}{2}, \quad \epsilon = \frac{\|\boldsymbol{c} - \boldsymbol{d}\|_2^2}{2}.$$

We will show that for these choices,

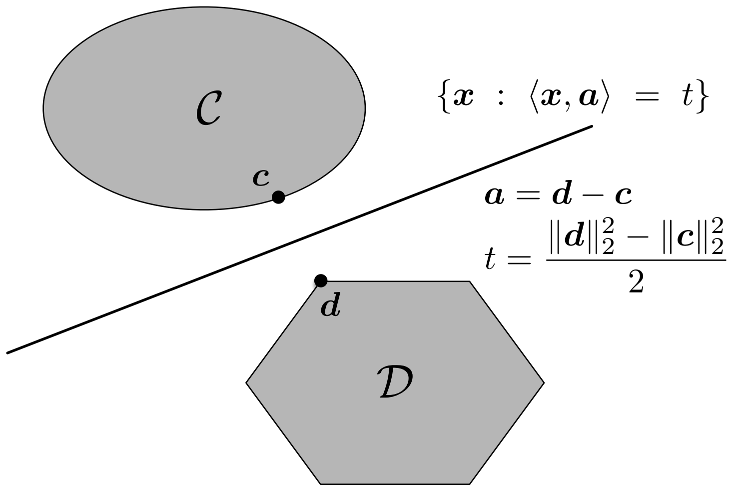
$$\langle \boldsymbol{x}, \boldsymbol{a} \rangle \ \leq t - \epsilon \ \text{ for } \boldsymbol{x} \in \mathcal{C}, \quad \langle \boldsymbol{x}, \boldsymbol{a} \rangle \ \geq t + \epsilon \ \text{ for } \boldsymbol{x} \in \mathcal{D},$$

To see this, we will set

$$f(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{a} \rangle - t,$$

and show that for any point $\boldsymbol{u} \in \mathcal{D}$, we have $f(\boldsymbol{u}) \geq \epsilon$.

Here is a picture to help visualize the proof:



First, we prove the basic geometric fact that for any two vectors $\boldsymbol{x}, \boldsymbol{y}$,

$$\text{if} \quad \|\boldsymbol{x} + \theta \boldsymbol{y}\|_2 \geq \|\boldsymbol{x}\|_2 \quad \text{for all} \quad \theta \in [0, 1] \quad \text{then} \quad \langle \boldsymbol{x}, \boldsymbol{y} \rangle \geq 0. \qquad (2)$$

To establish this, we expand the norm as

$$\|\boldsymbol{x} + \theta \boldsymbol{y}\|_2^2 = \|\boldsymbol{x}\|_2^2 + \theta^2 \|\boldsymbol{y}\|_2^2 + 2\theta \langle \boldsymbol{x}, \boldsymbol{y} \rangle,$$

from which we can immediately deduce that

$$\begin{aligned} \frac{\theta}{2} \|\boldsymbol{y}\|_2^2 + \langle \boldsymbol{x}, \boldsymbol{y} \rangle \geq 0 \quad \text{for all} \quad \theta \in [0, 1] \\ \Rightarrow \langle \boldsymbol{x}, \boldsymbol{y} \rangle \geq 0. \end{aligned}$$

Now let $\boldsymbol{u}$ be an arbitrary point in $\mathcal{D}$. Since $\mathcal{D}$ is convex, we know that $\boldsymbol{d} + \theta(\boldsymbol{u} - \boldsymbol{d}) \in \mathcal{D}$ for all $\theta \in [0, 1]$. Since $\boldsymbol{d}$ is as close to $\boldsymbol{c}$ as any other point in $\mathcal{D}$, we have

$$\|\boldsymbol{d} + \theta(\boldsymbol{u} - \boldsymbol{d}) - \boldsymbol{c}\|_2 \ \geq \ \|\boldsymbol{d} - \boldsymbol{c}\|_2,$$

and so by (2), we know that

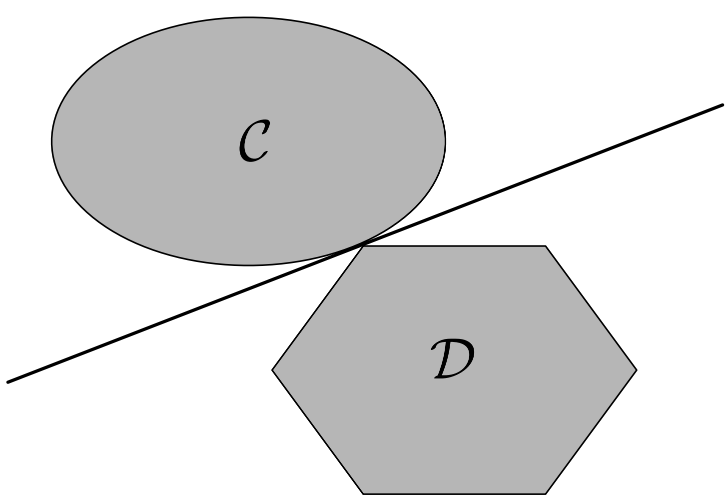
$$\langle \boldsymbol{d} - \boldsymbol{c}, \boldsymbol{u} - \boldsymbol{d} \rangle \ \geq \ 0.$$

This means

$$\begin{aligned} f(\boldsymbol{u}) &= \langle \boldsymbol{u}, \boldsymbol{d} - \boldsymbol{c} \rangle - \frac{\|\boldsymbol{d}\|_2^2 - \|\boldsymbol{c}\|_2^2}{2} \\ &= \langle \boldsymbol{u}, \boldsymbol{d} - \boldsymbol{c} \rangle - \frac{\langle \boldsymbol{d} + \boldsymbol{c}, \boldsymbol{d} - \boldsymbol{c} \rangle}{2} \\ &= \langle \boldsymbol{u} - (\boldsymbol{d} + \boldsymbol{c})/2, \ \boldsymbol{d} - \boldsymbol{c} \rangle \\ &= \langle \boldsymbol{u} - \boldsymbol{d} + \boldsymbol{d}/2 - \boldsymbol{c}/2, \ \boldsymbol{d} - \boldsymbol{c} \rangle \\ &= \langle \boldsymbol{u} - \boldsymbol{d}, \boldsymbol{d} - \boldsymbol{c} \rangle + \frac{\|\boldsymbol{c} - \boldsymbol{d}\|_2^2}{2} \\ &\geq \frac{\|\boldsymbol{c} - \boldsymbol{d}\|_2^2}{2}. \end{aligned}$$

The argument that $f(\boldsymbol{v}) \leq -\epsilon$ for every $\boldsymbol{v} \in \mathcal{C}$ is exactly the same.

We will not prove it here, but there is an even more interesting result that says that the sets $\mathcal{C}$ and $\mathcal{D}$ do not even have to be disjoint — they can intersect at one or more points along their boundaries as shown here:



Separating hyperplane theorem

Nonempty convex sets $\mathcal{C}, \mathcal{D} \subset \mathbb{R}^N$ can be (properly) separated by a hyperplane if and only if their relative interiors are disjoint:

$$\text{relint}(\mathcal{C}) \cap \text{relint}(\mathcal{D}) = \emptyset.$$

See the Technical Details for what exactly is meant by "relative interior" but it is basically everything not on the natural boundary of the set once we account for the fact that it might have dimension smaller than N.