Linear Algebra Background and Convex Sets

Orestis Plevrakis

In this lecture we will cover some background on linear algebra and analytic geometry, along with an introduction to convex sets.

1 Linear Algebra Background

Notation. Let $x, y \in \mathbb{R}^n$. Their inner product is defined as $x \cdot y := \sum_{i=1}^n x_i y_i$. We often view vectors as column matrices, and thus we may write $x^\top y$ instead of $x \cdot y$. The norm of x is $\|x\| := \sqrt{x \cdot x}$. We denote by $S^{n\times n}$ the set of all real $n \times n$ symmetric matrices.

1.1 Spectral Theorem

Theorem 1. *Let $A \in S^{n\times n}$. Then, A is diagonalizable. Furthermore,*

1. *All the eigenvalues of A are real.*
2. *A has an eigenbasis $u_1, u_2, \ldots, u_n \in \mathbb{R}^n$ such that the u_i's are pairwise orthogonal and each has norm one, i.e., A has an orthonormal eigenbasis.*

Remark 2. Let $A \in S^{n\times n}$, and let $u_1, u_2, \ldots, u_n$ an orthonormal eigenbasis and $\lambda_1, \lambda_2, \ldots, \lambda_n$ the corresponding eigenvalues. Let $\Lambda := diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $U := [u_1, u_2, \ldots, u_n]$, i.e., u_i is the i^{th} column of U. Then, $A = U\Lambda U^{-1}$, and since the u_i's are orthonormal, we have that $U^\top U = I$, which gives that $U^{-1} = U^\top$.

Remark 3. A real, square and invertible matrix V for which we have $V^{-1} = V^\top$ is called *orthogonal* matrix. The reason is that the matrix equations $V^\top V = VV^\top = I$ directly imply that both the set of rows and the set of columns of V are orthonormal.

1.2 Positive (Semi)-definite Matrices

Definition 4. Let $A \in S^{n\times n}$. If all the eigenvalues of A are nonnegative, we say that A is *positive semidefinite* (PSD), and we write $A \succcurlyeq 0$. If all the eigenvalues of A are positive, we say that A is *positive definite* (PD), and we write $A \succ 0$.

Theorem 5. *Let $A \in S^{n\times n}$. Then, the following are equivalent:*

1. $A \succcurlyeq 0$
2. *For all $x \in \mathbb{R}^n$, $x^\top Ax \geq 0$.*

Proof. First, we prove that $2 \to 1$. Suppose A is not PSD. From the Spectral Theorem, A has at least one negative eigenvalue λ. Let u be a corresponding eigenvector. Then, $u^\top Au = u^\top(\lambda u) = \lambda\|u\|^2 < 0$, contradiction.

We now prove that $1 \to 2$. Let $u_1, \ldots, u_n$ an orthonormal eigenbasis, and let $\lambda_1, \ldots, \lambda_n \geq 0$ the corresponding eigenvalues. Let $x \in \mathbb{R}^n$. Then, we can write x as a linear combination of the eigenvectors: $x = \sum_{i=1}^n a_i u_i$. Thus, $Ax = \sum_{i=1}^n \lambda_i a_i u_i$, and

$$x^\top Ax = \left(\sum_{i=1}^n a_i u_i\right) \cdot \left(\sum_{i=1}^n \lambda_i a_i u_i\right) = \sum_{i=1}^n \lambda_i a_i^2$$

where the last equality follows from the fact that $u_1, \ldots, u_n$ are orthonormal.

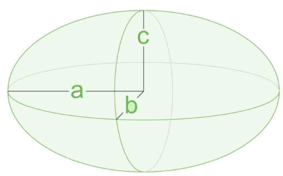
□

Remark 6. The analog of this theorem for PD matrices is that for $A \in S^{n\times n}$, we have $A \succ 0 \leftrightarrow x \in \mathbb{R}^n \setminus \{0\},\ x^\top Ax > 0$. The proof is completely analogous.

Remark 7. For a fixed $A \in S^{n\times n}$, the function of $x \mapsto x^\top Ax$ is called *quadratic form*.

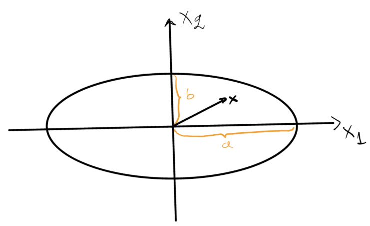
2 Ellipsoids

The three dimensional analog of an ellipse is the *ellipsoid*:



In this section, we will develop the analog of the above shape for $\mathbb{R}^n$, where n can be greater than 3. These high dimensional ellipsoids will be key geometric objects for obtaining a polynomial time algorithm for convex optimization. To develop the analog, we first need to see what are the algebraic expressions that describe the 2d ellipses and 3d ellipsoids.

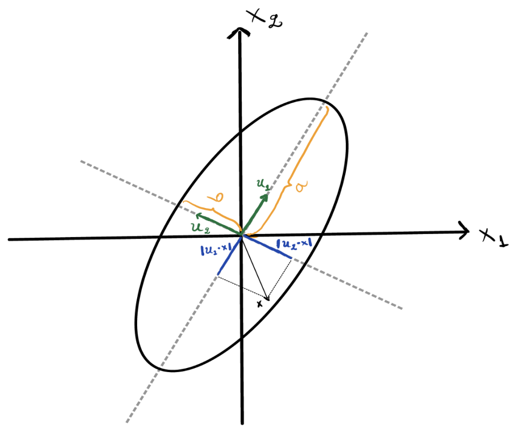
- Axis-aligned ellipse: consider the axis-aligned ellipse below with axis-lengths a, b:



We know that a point x is inside the ellipse iff

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} \leq 1 \tag{1}$$

- Non axis-aligned ellipse: consider the example below, where the axes are along two orthonormal vectors u_1, u_2:

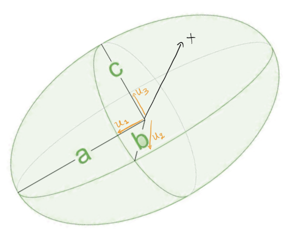


To derive the algebraic expression for this ellipse, note that 1 can be read as follows:

$$\frac{(\text{length of projection on 1st axis})^2}{a^2} + \frac{(\text{length of projection on 2nd axis})^2}{b^2} \leq 1$$

For the ellipse with axes along u_1, u_2, the lengths of the projections of a point x are $|x \cdot u_1|$, $|x \cdot u_2|$, and thus x is in the ellipse iff $\frac{(x \cdot u_1)^2}{a^2} + \frac{(x \cdot u_2)^2}{b^2} \leq 1$

- By analogy, the 3d ellipsoid with axes given by an orthonormal set of vectors u_1, u_2, u_3 and axis-lengths a, b, c is defined by the inequality $\frac{(x \cdot u_1)^2}{a^2} + \frac{(x \cdot u_2)^2}{b^2} + \frac{(x \cdot u_3)^2}{c^2} \leq 1$. This is the resulting shape:



The discussion up to now focused on ellipses and ellipsoids centered at the origin. To get the algebraic expression for an ellipse/ellipsoid centered at some point x_0, we replace x at the above inequalities with $x - x_0$ (why?).

- At this point, the generalization to $\mathbb{R}^n$ is natural: for a given orthonormal basis $u_1, u_2, \ldots, u_n$ (the axes), positive numbers $a_1, a_2, \ldots, a_n$ (axis-lengths), and point x_0 (center), the set

$$E = \left\{ x \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{n} \frac{((x - x_0) \cdot u_i)^2}{a_i^2} \leq 1 \right\} \tag{2}$$

is called an ellipsoid. Now, the sum in 2 can be written compactly using matrices: let $U := [u_1, u_2, \ldots, u_n]$, $\Lambda := diag(1/a_1, 1/a_2, \ldots, 1/a_n)$. Then,

$$\begin{aligned} \sum_{i=1}^{n} \frac{((x - x_0) \cdot u_i)^2}{a_i^2} = \|\Lambda U^\top (x - x_0)\|^2 &= \left(\Lambda U^\top (x - x_0)\right)^\top \left(\Lambda U^\top (x - x_0)\right) \\ &= (x - x_0)^\top U \Lambda^\top \Lambda U^\top (x - x_0) \\ &= (x - x_0)^\top U \Lambda^2 U^\top (x - x_0) \end{aligned}$$

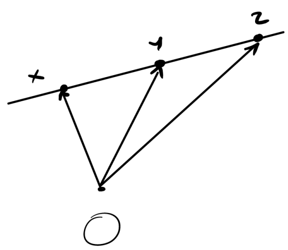
In the middle, we have a symmetric matrix $A := U\Lambda^2 U^\top$, which appears in its eigenvalue decomposition. Since $\Lambda^2 = diag(1/a_1^2, \ldots, 1/a_n^2)$, we have that A is positive definite! Observe that since we started with an arbitrary orthonormal basis $u_1, \ldots, u_n$ and with arbitrary axis-lengths $a_1, \ldots, a_n$, we end with an arbitrary positive definite matrix A. We have essentially proven the proposition below:

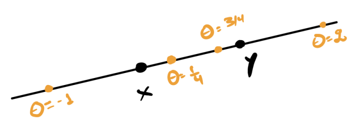
Proposition 8. *A set E is an ellipsoid in $\mathbb{R}^n$ iff $E = \left\{x \in \mathbb{R}^n \mid (x - x_0)^\top A(x - x_0) \leq 1\right\}$, for some positive definite matrix A, and some $x_0 \in \mathbb{R}^n$.*

3 Convex Sets

3.1 Lines and Line Segments

Before we define convexity, we remember the definition of a line in $\mathbb{R}^n$: a line connecting two points $x, y \in \mathbb{R}^n$ is defined to be the set $\{x + \theta(y - x) \mid \theta \in \mathbb{R}\}$. This definition comes from generalizing the algebraic expression for the line on the plane and in space. Here is why: consider a line in space connecting two different points $x, y \in \mathbb{R}^3$. Using vectors, a point z is on the line iff $\overrightarrow{Oz} = \overrightarrow{Ox} + \theta\overrightarrow{xy}$, for some $\theta \in \mathbb{R}$ (see left figure below). Using the corespondance between points and vectors, the last equation is equivalent to $z = x + \theta(y - x)$. Furthermore, from our reasoning follows that as θ moves from 0 to 1, the point z moves from x to y (see right figure below).



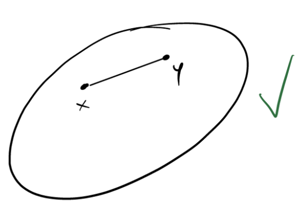


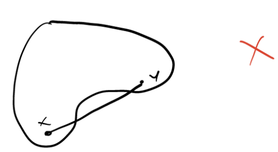
From the discussion above, we see that the natural way to define the line segment connecting two points $x, y \in \mathbb{R}^n$ is $[x, y] := \{x + \theta(y - x) \mid \theta \in [0, 1]\} = \{(1 - \theta)x + \theta y \mid \theta \in [0, 1]\}$

3.2 Convex Sets

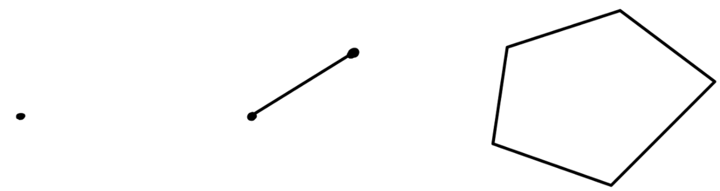
Definition 9. A set K in $\mathbb{R}^n$ is convex if for any $x, y \in K$, we have $[x, y] \subseteq K$.

The sets below are examples of convex and non-convex sets on the plane:





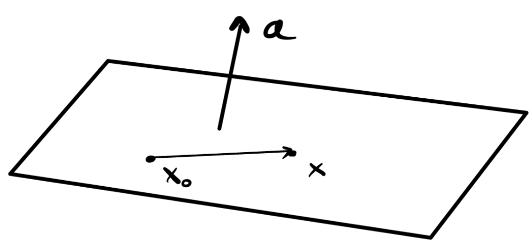
Some more examples of convex sets on the plane are single points, line segments, and convex polygons:



We now move to examples of convex sets in $\mathbb{R}^n$.

3.3 Hyperplanes

A hyperplane is a set $\{x \in \mathbb{R}^n \mid a \cdot x = b\}$, for some $a \in \mathbb{R}^n$, $a \neq 0$ and $b \in \mathbb{R}$. Hyperplanes generalize planes in $\mathbb{R}^3$ and lines in $\mathbb{R}^2$. I remind you why this is the case: consider a plane in $\mathbb{R}^3$ with normal vector a and a point x_0 on it. Then, a point x belongs to the plane iff $a \cdot (x - x_0) = 0 \leftrightarrow a \cdot x = a \cdot x_0$, and $b := a \cdot x_0$. We showed that every plane is a set $\{x \in \mathbb{R}^n \mid a \cdot x = b\}$, for some $a \in \mathbb{R}^n$, $a \neq 0$ and $b \in \mathbb{R}$. We also need to show that every such set is a plane. First of all, such a set is always nonempty (why?). Let x_0 a point in it. Then, from the previous argument, the set is precisely the plane with normal vector a, that contains x_0.



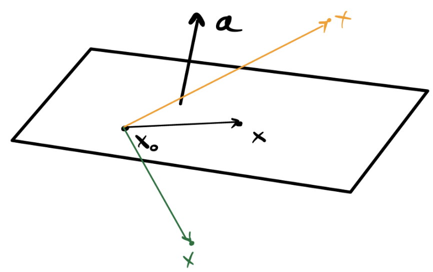
For the case of lines on the plane the same reasoning applies (the vector a will be a normal vector to the line).

Proposition 10. *Hyperplanes are convex.*

Proof. Let x, y points of a hyperplane with parameters a, b. Let $(1-\theta)x + \theta y$ (where $\theta \in [0,1]$) an arbitrary point on $[x, y]$. Then, $a \cdot ((1-\theta)x + \theta y) = (1-\theta)a \cdot x + \theta a \cdot y = b$. □

3.4 Halfspaces

A halfspace is a set $\{x \in \mathbb{R}^n \mid a \cdot x \leq b\}$, where $a \in \mathbb{R}^n, a \neq 0$ and $b \in \mathbb{R}$. It generalizes 3d halfspaces and 2d halfplanes. Here is why: consider a 3d plane; a halspace is everything that is not above the plane:



We see in the picture that a point x is above the plane iff $a \cdot (x - x_0) > 0$. The argument for 2d halfplanes is identical (a 2d halfplane is everything that is not above a line in $\mathbb{R}^2$).

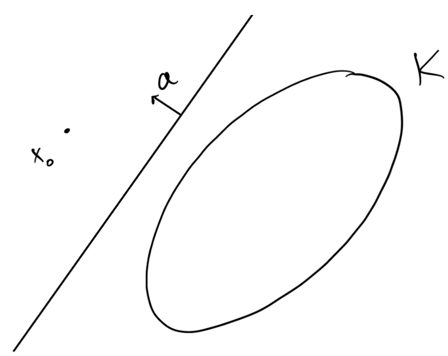
Proposition 11. *Halfspaces are convex.*

The proof is same as before, so I omit it.

In the next class we will see more examples of convex sets in $\mathbb{R}^n$ such as balls, ellipsoids and polyhedra. Now, we move on the most fundamental theorem for convex sets.

3.5 Separating Hyperplane Theorem

The following theorem says that given a closed convex set K, and a point x_0 outside it, there exists a hyperplane that separates K from x_0. Below is an illustration in $\mathbb{R}^2$.



Theorem 12. *Let K be a closed* [1] *convex set in $\mathbb{R}^n$. Let $x_0 \in \mathbb{R}^n$ such that $x_0 \notin K$. Theren, there exist nonzero $a \in \mathbb{R}^n$, and $b \in \mathbb{R}$ such that*

$$\begin{aligned} &a \cdot x_0 > b \\ &a \cdot x < b, \ \forall x \in K \end{aligned}$$

Remark 13. The assumption that K is closed is necessary (why?).

We will prove this theorem in the next class. In the problem posed for next week, you will prove a two-dimensional version of this theorem. The proof for this special case will reveal the key idea for proving the general theorem!

We end this lecture with a reminder from Analysis that will be helpful in subsequent classes:

[1] A set is closed if it contains its boundary.

Theorem 14. *Let $A \subseteq \mathbb{R}^n$ compact* [2]*, and $f : A \to \mathbb{R}$ continuous. Then, f attains its maximum and minimum values inside A.*

[2] Compact means closed and bounded. Bounded means that there exists a ball that contains the set.