Lecture 5
The driver of duality: Separation

Instructor: Prof. Gabriele Farina (✉ gfarina@mit.edu)⋆

In Lecture 3 we guessed the expression for the normal cone to the intersection of halfspaces. We then saw that our guess, natural in its graphical intuition, quite surprisingly implied immediately strong linear programming duality. In light of this, whatever proof techniques confirms our guess for the expression of the normal cone correct, rightfully deserves our attention, as it must encode the grain of duality.

As it turns our, the key idea behind the proof is the concept of *separation*.

L5.1 Separating a point from a closed convex set

An important property of any convex set Ω is that whenever a point y is not in Ω, then we can *separate y* from Ω using a *hyperplane*. In other words, *flat* separating surfaces are enough for certifying that a point $y \notin \Omega$.

> **Theorem L5.1.** Let $\Omega \subseteq \mathbb{R}^n$ be a nonempty, closed, and convex set, and let $y \in \mathbb{R}^n$ be a point. If $y \notin \Omega$, then there exist $u \in \mathbb{R}^n, v \in \mathbb{R}$ such that
>
> $$\langle u, y \rangle < v, \quad \text{and} \quad \langle u, x \rangle \geq v \quad \forall x \in \Omega.$$

Proof. The proof of the result rests on a very simple idea: the direction of the halfspace will be made orthogonal to the line that connects y to its projection x^* onto Ω, and the halfspace boundary will be set so that it passes through x^*. We now make the argument formal.

First, since Ω is nonempty and closed, a Euclidean projection x^* of y onto Ω exists,[1] as we discussed in Lecture 1. In other words, the nonlinear optimization problem

$$\min_x \ \tfrac{1}{2}\|x - y\|_2^2$$
$$\text{s.t.} \ \ x \in \Omega$$
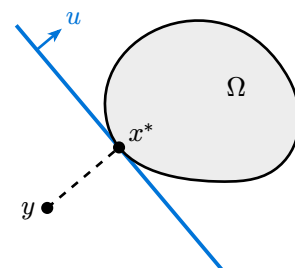
⋆These notes are class material that has not undergone formal peer review. The TAs and I are grateful for any reports of typos.

[1]In fact, it is easy to prove that the projection is unique (see strict convexity, Lecture 4). However, we do not need uniqueness for the argument that follows.

must have at least a solution $x^* \in \Omega$. Furthermore, since the objective function is differentiable and Ω is convex, from the first-order optimality conditions (see Lecture 2) we know that

$$\langle x^* - y, x - x^* \rangle \geq 0 \quad \forall x \in \Omega. \tag{1}$$

Let now

$$u := x^* - y, \qquad [\triangleright \text{ this is the direction that connects } y \text{ to } x^*]$$

$$\text{and } v := \langle u, x^* \rangle. \qquad [\triangleright \text{ so that the halfspace boundary passes through } x^*]$$

Note that $u \neq 0$, since $x^* \in \Omega$ but $y \notin \Omega$. So, $\|u\| > 0$ and therefore

$$\langle u, y \rangle = \langle u, x^* - u \rangle = v - \|u\|_2^2 < v.$$

Thus, to complete the proof, we now need to show that $\langle u, x \rangle \geq v$ for all $x \in \Omega$. But this is exactly what (1) guarantees, since $u = x^* - y$ and $v = \langle u, x^* \rangle$ by definition. $\square$

The result above might not seem like much. After all, the proof is pretty straightforward, and the geometric intuition strong enough that one might be tempted to just take it for granted. Instead, the consequences of the result are deep, far-reaching, and intimately tied to some of the most significant breakthroughs in mathematical optimization theory.

> **Remark L5.1.** The result of Theorem L5.1 holds even if we insist on only having strict inequalities, that is $\langle u, y \rangle < v$ and $\langle u, x \rangle > v$ for all $x \in \Omega$. We can see this in two ways:
> - Graphically, in the proof we could have chosen v so that the halfspace would pass through the midpoint of the line connecting y and x^*.
> - Algebraically, let u, v be as in Theorem L5.1. We will show that we can always perturb v to make both inequalities hold strictly. The key is the observation that
>
> $$\langle u, y \rangle = \frac{1}{2}\langle u, y \rangle + \frac{1}{2}\langle u, y \rangle < \frac{1}{2}(v + \langle u, y \rangle)$$
>
> $$\langle u, x \rangle \geq v = \frac{1}{2}(v + v) > \frac{1}{2}(v + \langle u, y \rangle) \qquad \forall x \in \Omega.$$
>
> Hence, in both cases we have that the scalar $v' := \frac{1}{2}(v + \langle u, y \rangle)$ satisfies $\langle u, y \rangle < v'$, $\langle u, x \rangle > v'$ for all $x \in \Omega$.

L5.1.1 Separating a point from a convex cone

Before we prove the expression for the normal cone at the intersection of halfspaces, we will find it helpful to use the following corollary of separation for *convex cones*. A *cone* is a set with the property that the ray $\{\lambda \cdot x : \lambda \geq 0\}$ generated by any point x in the set is fully contained in the set.

> **Definition L5.1** (Cone). A set S is a *cone* if, for any $x \in S$ and $\lambda \in \mathbb{R}_{\geq 0}$, the point $\lambda \cdot x \in S$.

Convex cones are among the simplest convex sets, and they appear all the time in optimization theory.[2] In particular, in the next theorem we show that separation of a point from a nonempty closed convex cone can always be achieved using a hyperplane passing through the origin.

> **Theorem L5.2.** Let $S \subseteq \mathbb{R}^n$ be a nonempty closed convex cone, and $y \notin S$ be a point in $\mathbb{R}^n$. Then, there exists a hyperplane *passing through the origin* that separates y from S; formally, there exists $u \in \mathbb{R}^n$ such that
>
> $$\langle u, y \rangle < 0 \quad \text{and} \quad \langle u, x \rangle \geq 0 \quad \forall x \in S.$$

Proof. We already know from Theorem L5.1 that there exist $u \in \mathbb{R}^n, v \in \mathbb{R}$ such that

$$\langle u, y \rangle < v \quad \text{and} \quad \langle u, x \rangle \geq v \quad \forall x \in S. \tag{2}$$

Consider any point $a \in S$. By definition of cone, $\lambda \cdot a \in S$ for all $\lambda \geq 0$. Thus, the separation condition on the right in (2) implies that $v \leq \lambda \cdot \langle u, a \rangle$ for all $\lambda \geq 0$. In particular, by plugging $\lambda = 0$, we find that $v \leq 0$, yielding $\langle u, y \rangle < 0$. Furthermore, dividing by λ we find that

$$\langle u, a \rangle \geq \frac{v}{\lambda} \quad \forall \lambda \geq 0 \quad \implies \quad \langle u, a \rangle \geq \sup_{\lambda \to \infty} \frac{v}{\lambda} = 0.$$

Since $a \in S$ was arbitrary, the statement follows. $\qquad \square$

L5.2   A second look at the normal cone of linear constraints

In Lecture 3, we considered normal cones at the intersection of halfspaces. On that occasion, we drew a picture and were convinced that the normal cone at a point at the intersection of halfspaces was given by the conic hull of the directions orthogonal to those halfspaces (see the figure on the right).
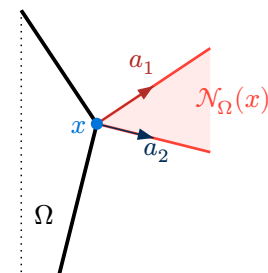
This led to the following guess, which was left unproven.

> **Theorem L5.3.** Let $\Omega \subseteq \mathbb{R}^n$ be defined as the intersection of m linear inequalities
>
> $$\Omega := \{x \in \mathbb{R}^n : Ax \leq b\}, \quad \text{where} \quad A = \begin{pmatrix} - & a_1^\top & - \\ & \vdots & \\ - & a_m^\top & - \end{pmatrix} \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m.$$
>
> Given a point $x \in \Omega$, define the index set of the "active" constraints
>
> $$I(x) := \left\{ j \in \{1, ..., m\} : a_j^\top x = b_j \right\}.$$

---

[2]Hiriart-Urruty, J.-B., & Lemaréchal, C. [HL01], referring to convex cones, write: *"they are important in convex analysis (the "unilateral" realm of inequalities), just as subspaces are important in linear analysis (the "bilateral" realm of equalities)".*

Then, the normal cone at any $x \in \Omega$ is given by

$$\mathcal{N}_\Omega(x) = \left\{ \sum_{j \in I(x)} \lambda_j a_j : \lambda_j \geq 0 \right\} = \left\{ A^\top \lambda : \lambda^\top (b - Ax) = 0, \lambda \in \mathbb{R}^m_{\geq 0} \right\},$$

where the second equality rewrites the condition $j \in I(x)$ via the *complementary slackness* (see Lecture 3).

We now give the proof of the above result. We will do that by invoking the machinery of separation to argue that a direction outside of the normal cone must form an acute angle with at least one direction that remains in the feasible set Ω.

Proof of Theorem L5.3. Fix any $x \in \Omega$ and let

$$\mathcal{C}(x) := \left\{ \sum_{j \in I(x)} \lambda_j a_j : \lambda_j \geq 0 \right\}.$$

We will show that $\mathcal{N}_\Omega(x) = \mathcal{C}(x)$ by proving the two directions of inclusion separately.

- We start by showing that any $d \in \mathcal{C}(x)$ belongs to $\mathcal{N}_\Omega(x)$, that is,

$$\langle d, y - x \rangle \leq 0 \text{ for all } y \in \Omega.$$

Let d be expressed as $\sum_{j \in I(x)} \lambda_j a_j$ with $\lambda_j \geq 0$. Then, for any $y \in \Omega$,

$$\left\langle \sum_{j \in I(x)} \lambda_j a_j, y - x \right\rangle = \sum_{j \in I(x)} \lambda_j \langle a_j, y - x \rangle$$

$$= \sum_{j \in I(x)} \lambda_j (\langle a_j, y \rangle - b_j) \qquad \text{(by definition of } I(x), \langle a_j, x \rangle = b_j)$$

$$\leq \sum_{j \in I(x)} \lambda_j (b_j - b_j) = 0. \qquad \text{(since } y \in \Omega \text{ and } \lambda_j \geq 0)$$

This shows that $d \in \mathcal{N}_\Omega(x)$ and concludes the proof of this direction of the inclusion.

- We now look at the other direction. Take any $d \notin \mathcal{C}(x)$. Since $\mathcal{C}$ is a nonempty closed convex cone [▷ you should verify this claim], by the conic separation result of Theorem L5.2, there must exist $u \in \mathbb{R}^n$ such that

$$\langle u, d \rangle < 0, \quad \text{and} \qquad \langle u, a \rangle \geq 0 \quad \forall a \in \mathcal{C}(x). \tag{3}$$

We argue that for $\delta > 0$ small enough, the point $y := x - \delta \cdot u$ belongs to Ω. We do so by showing that it satisfies all the inequalities $a_j^\top x \leq b_j$ that define Ω:
  - if $j \in I(x)$, then $\langle a_j, x - \delta \cdot u \rangle = b_j - \delta \cdot \langle a_j, u \rangle \leq b_j$ since $\langle a_j, u \rangle \geq 0$ by (3).
  - if $j \notin I(x)$, then $b_j - \langle a_j, x \rangle > 0$. By continuity, small enough perturbations of x, in any direction, will not affect the strict inequality.

Thus, the direction $\delta \cdot u$ remains inside of Ω starting from x. We now argue that it forms a strictly positive inner product with d. Indeed, note that from (3)

$$\langle d, y - x \rangle = \langle d, -\delta \cdot u \rangle = -\delta \cdot \langle d, u \rangle > 0.$$

This shows that $d \notin \mathcal{C}(x) \implies d \notin \mathcal{N}_\Omega(x)$, completing the proof. □

L5.3 Separation oracles

The result established in Theorem L5.1 justifies the following definition.

Definition L5.2 ((Strong) separation oracle). Let $\Omega \subseteq \mathbb{R}^n$ be convex and closed. A *strong separation oracle* for Ω is an algorithm that, given any point $y \in \mathbb{R}^n$, correctly outputs one of the following:

- "$y \in \Omega$", or

- "$(y \notin \Omega, u)$", where the vector $u \in \mathbb{R}^n$ is such that

$$\langle u, y \rangle < \langle u, x \rangle \qquad \forall x \in \Omega.$$

L5.3.1 Finding separating hyperplanes in practice

Theorem L5.1 guarantees the *existence* of a separating hyperplane. In many problems of interest, *constructing* a separation oracle is simple.

Example L5.1 (Separation oracle for a convex polytope). Let Ω be a convex polytope, that is, the convex set defined by the intersection of a finite number of halfspaces (linear inequalities)

$$\Omega := \{x \in \mathbb{R}^n : Ax \le b\}, \quad \text{where} \quad A = \begin{pmatrix} - a_1^\top - \\ \vdots \\ - a_m^\top - \end{pmatrix} \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m.$$

Then, given a point $y \in \mathbb{R}^m$, we can implement a separation oracle as follows:

- if $Ay \le b$, return "$y \in \Omega$";

- else, at least one of the inequalities $a_j^\top y \le b_j$, $j \in \{1, ..., m\}$ is violated. In other words, there exists j such that $a_j^\top y > b_j$, while by definition of Ω, $a_j^\top x \le b_j$ for all x. This shows that the response "$(y \notin \Omega, -a_j)$" is a valid response.

Remark L5.2. Example L5.1 shows that whenever we have a finite number m of inequalities, a separation oracle for the polytope defined by those inequalities can be implemented in time that depends linearly on m and the dimension of the embedding space. This result establishes a *blanket* guarantee, but in some cases, one can do better: depending on the structure of the inequalities, sometimes one can get away with sublinear complexity in m. In some cases, one might be able to construct an efficient separation oracle even for polytopes that have an infinite number of inequalities!

We proceed with another classic example of a feasible set that admits a simple separation oracle.

> **Example L5.2** (Separation oracle for the semidefinite cone). Let $\Omega = \{M \in \mathbb{R}^{n \times n} : M \succeq 0\}$ be the set of semidefinite matrices, that is, all symmetric matrices such that $v^\top M v \geq 0$ for all $v \in \mathbb{R}^n$—or, equivalently, such that all of M's eigenvalues are nonnegative.
>
> Then, given a point $Y \in \mathbb{R}^{n \times n}$, we can implement a separation oracle as follows:
>
> - if Y is *not* symmetric, then there exist $i, j \in \{1, ..., n\}$ such that $Y_{ij} < Y_{ji}$; return "$(Y \notin \Omega, E_{ij} - E_{ji})$", where E_{ij} is the matrix of all zeros, except in position i, j where it has a 1.
>
> - else, if Y is symmetric, we can compute all of its eigenvalues and eigenvectors. If one eigenvalue is negative, then the corresponding eigenvector w must be such that $w^\top Y w = \langle Y, ww^\top \rangle < 0$. Hence, return "$(Y \notin \Omega, ww^\top)$".
>
> - otherwise, return "$Y \in \Omega$".

As we show next, a fundamental result in optimization theory reveals that under mild hypotheses, if the feasible set admits an efficient separation oracle and the objective function is convex, then the solution can be computed efficiently.

L5.4 Optimization via separation

In a major breakthrough in mathematical optimization, Khachiyan, L. G. [Kha80] proposed a polynomial-time algorithm for using separation oracles to find the minimum of a linear function. The algorithm, which goes under the name of *ellipsoid method* is actually more general, and applies to general convex objectives on sets for which separation oracles are available. The result builds on top of previous work by Šor, N. Z. [Šor77] and Yudin, D. B., & Nemirovskii, A. S. [YN76].

In particular, Khachiyan's result was the first to show that linear programming problems can be solved in polynomial time. This was an unexpected result at the time, and in fact, the complexity of linear programming solvers was conjectured to be *not* polynomial (more on this in the next section). The result of Khachiyan stirred so much enthusiasm in the research community that the New York Times even advertised it on its first page.

Despite the enthusiasm, the ellipsoid method turned out to be very impractical. Still, it is a great theoretical idea, and its consequences are pervasive.

L5.4.1 The intuition behind the ellipsoid method

Formalizing the details of the ellipsoid method is rather complex. A major source of difficulty is the fact that the algorithm needs to approximate square roots using fractions to be implementable on a finite-precision machine, and that causes all sorts of tricky analyses that the approximation error can indeed be kept under control. These details are certainly important, but are notoriously tedious, and fundamentally they are just that, *details*. If you are curious to read a formal account, I recommend the authoritative book by Grötschel, M., Lovász, L., & Schrijver, A. [GLS93]. For this lecture, we just focus on the *idea* behind the ellipsoid method.

Figure 1: https://timesmachine.nytimes.com/timesmachine/1979/11/07/issue.html

The idea behind the ellipsoid method is rather elegant. At its core, it is a generalization of *binary search* from one dimension to multiple dimensions. At every iteration of the algorithm, the space is "cut" by using a separating hyperplane.

■ **Feasibility.** To build intuition, ignore for now the objective function, and consider the following problem: given a separation oracle for Ω (closed and convex), either find $x \in \Omega$, or determine that Ω is empty. You are given two radiuses:

- the radius $R > 0$ guarantees that if Ω is not empty, then $\Omega \cap \mathbb{B}_R(0) \neq \emptyset$;
- the radius $r > 0$ guarantees that if Ω is not empty, then it contains a ball of radius r in its interior.

If this problem were one-dimensional, then Ω would be either empty or an interval, and a separation oracle would be an algorithm that, given any $y \in \mathbb{R}$, would return whether $y \in \Omega$, or one of the statements "y is too small" / "y is too large". Solving the problem now appears easy: start from the interval $[-R, R]$, and perform a binary search using the separation oracle to guide the search. Once the size of the search interval drops below r, we know that Ω is empty.

The ellipsoid method generalizes this idea to multiple dimensions. At every iteration, it keeps track of a "search space" (the generalization of the search interval above). Then, it queries the separation oracle for the center c_t of this search space. If the point does not belong to Ω, and the separation oracle returns the separating direction $u \in \mathbb{R}^n$, then the search space is cut by considering now only the subset of the search space that intersects $\{x \in \mathbb{R}^n : \langle u, x - c_t \rangle \geq 0\}$. The process continues until the volume of the search space becomes smaller than the radius r. The reason why this method is called the "ellipsoid method" is that the search space in the multi-dimensional case is not kept in the form of an interval, but rather as an ellipsoid. This is mostly for computational reasons, since we need to have an internal way of representing the search domain that is convenient to use.

■ **Incorporating the objective.** The above idea can be extended to incorporate an objective function $f(x)$. To do that, we will need to start cutting not only the search ellipsoid, but also the feasible set to make sure we end up at the optimum. In other words, you can think of this extended ellipsoid method as having "two modes": while it has not found a feasible point in Ω, it cuts the search ellipsoid; then, once feasible points are found, it cuts the feasible set to exclude all values above the current value.

- Initialize at time $t = 1$ with the starting point $y_1 := 0 \in \mathbb{R}^n$, starting ellipsoid $\mathcal{E}_1 := \mathbb{B}_R(0)$, and starting feasible set $\Omega_1 := \Omega$.

- At each time t, we ask a separation oracle for Ω_t whether the center $c_t \in \mathbb{R}^n$ of the search ellipsoid $\mathcal{E}_t$ belongs to Ω_t or not.[3] There are only two cases:

  ‣ If the center c_t is *not* feasible, then set $\Omega_{t+1} := \Omega_t$, and cut the search space by setting $\mathcal{E}_{t+1}$ to an ellipsoid that contains the intersection between $\mathcal{E}_t$ and the halfspace containing Ω_t returned by the separation oracle.

---

[3]There is a caveat here: technically, we are assuming as given a separation oracle for Ω, *not* Ω_t. Yet, because Ω_t is obtained from Ω by intersecting with halfspaces, it is easy to see that one separation oracle for Ω_t can be constructed efficiently starting from that for Ω and the description of the intersected hyperplanes. Try working out the details!

- If the center c_t is feasible, then we know for sure that all points $x \in \Omega_t$ such that $\langle \nabla f(c_t), x - c_t \rangle \geq 0$ are such that $f(x) \geq f(c_t)$. This follows trivially from the linear lower bound property of convex functions (Theorem L4.1 of Lecture 4):

$$\langle \nabla f(c_t), x - c_t \rangle \geq 0 \quad \implies \quad f(x) \geq f(c_t) + \langle \nabla f(c_t), x - c_t \rangle \geq f(c_t).$$

  Hence, we can cut *both* the search ellipsoid $\mathcal{E}_t$ and the feasible set Ω_t by considering their intersection with the halfspace $H_t := \{x \in \mathbb{R}^n : \langle \nabla f(c_t), x - c_t \rangle \leq 0\}$. In particular, we set $\Omega_{t+1} := \Omega_t \cap H_t$, and set $\mathcal{E}_{t+1}$ to a smaller ellipsoid that contains $\mathcal{E}_t \cap H_t$.

- Finally, after the volume of the search ellipsoid has gotten sufficiently small (this happens after $T = O(n^2) \log(R/r)$ iterations), we output the following:
  - If we never encountered a center c_t that was feasible, then we report that Ω was infeasible.
  - Else, we output the c_t that minimizes f, out of those that were feasible.

Assuming that we can ignore all sorts of tedious rounding issues, the following guarantee can be shown [Gup20].

Theorem L5.4. Let R and r be as above, and let the range of the function f on Ω be bounded by $[-B, B]$. Then, the ellipsoid method described above run for $T \geq 2n^2 \log(R/r)$ steps either correctly reports that $\Omega = \emptyset$, or produces a point x^* such that

$$f(x^*) \leq f(x) + \frac{2BR}{r} \exp\left(-\frac{T}{2n(n+1)}\right) \qquad \forall x \in \Omega.$$

L5.4.2 Takeaway message: Separation implies optimization

If you squint your eyes, what the ellipsoid method proves constructively is the following: if we know how to construct a separation oracle for a set Ω, then we can optimize over Ω. Of course, this is a bit of a simplification (and there are all sorts of little conditions here and there as we have seen above), but nonetheless it is a good first approximation of the general message.

In a later lecture, we will discuss how the opposite direction is also known to be true, even when by "optimization" we simply mean optimization of linear objective functions.

Further readings and bibliography

If you want to read more about the ellipsoid method, the book by Grötschel, M., Lovász, L., & Schrijver, A. [GLS93] is a standard and accessible reference on the topic. The bound on the approximation error incurred by the ellipsoid method was taken from Gupta, A. [Gup20].

[HL01]   Hiriart-Urruty, J.-B., & Lemaréchal, C. (2001). *Fundamentals of Convex Analysis.* Springer. `https://link.springer.com/book/10.1007/978-3-642-56468-0`

[Kha80] Khachiyan, L. G. (1980). Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, *20*(1), 53–72.

[Šor77] Šor, N. Z. (1977). Cut-off method with space extension in convex programming problems. *Cybernetics*, *13*(1), 94–96.

[YN76] Yudin, D. B., & Nemirovskii, A. S. (1976). Informational complexity and efficient methods for the solution of convex extremal problems. *Matekon*, *13*(2), 22–45.

[GLS93] Grötschel, M., Lovász, L., & Schrijver, A. (1993). *Geometric Algorithms and Combinatorial Optimization*. Springer. `https://link.springer.com/book/10.1007/978-3-642-78240-4`

[Gup20] Gupta, A. (2020). *The Centroid and Ellipsoid Algorithms*. `https://www.cs.cmu.edu/~15850/notes/lec21.pdf`

Changelog
- Feb 20, 2025: Added Remark L5.1 and edited Definition L5.2.
- Feb 24, 2025: Changed compact → closed in Definition L5.2. (Thanks `https://piazza.com/class/m6lg9aspoutda/post/m7jzmkp7jyv6xd`)