MIT 6.7220/15.084 — Nonlinear Optimization (Spring '25) Thu, Mar 13th 2025

Lecture 11
Polarity and oracle equivalence

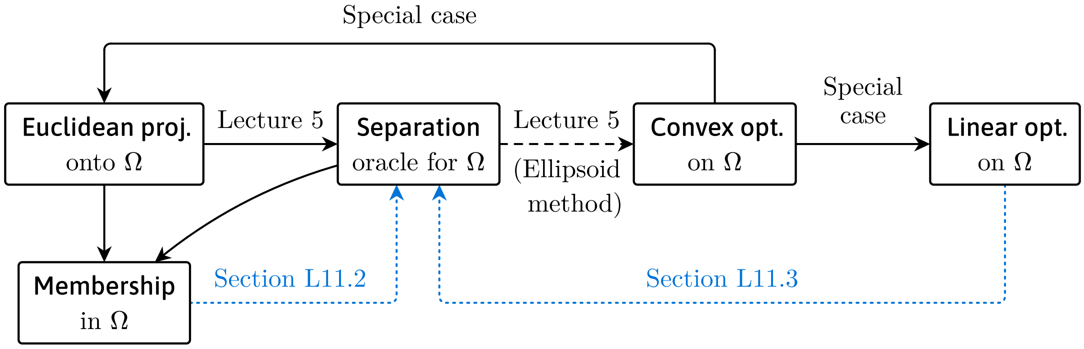
Instructor: Prof. Gabriele Farina (✉ gfarina@mit.edu)*

We conclude this first part of the course by tying one last loose end and giving one last fundamental result regarding the *computational* aspects of convex optimization problems.

Let $\Omega \subseteq \mathbb{R}^n$ be closed, convex, and bounded. In Lecture 5, we saw how *separation*—which we argued captures the essence of duality—can be turned into an algorithmic tool through the notion of *separation oracle*. Specifically, if a separation oracle for Ω can be constructed, then it can be converted into an algorithm for convex optimization on Ω via the ellipsoid algorithm provided that there exist radii $r, R > 0$ and a point $x_0 \in \Omega$ such that $\mathbb{B}_r(x_0) \subseteq \Omega \subseteq \mathbb{B}_R(0)$.

On the other hand, a separation oracle for Ω can always be constructed efficiently from an algorithm for convex optimization on Ω, since the direction connecting y to its projection onto Ω is a separating direction (see Lecture 5).

With a separation oracle, or an algorithm for computing Euclidean projections onto Ω, we can always decide whether $y \in \Omega$ or not—that is, construct a membership oracle. Finally, given an algorithm for convex optimization over Ω, we obtain as a special case an algorithm for *linear* optimization on Ω. We can summarize all these observations pictorially via the black solid arrows in the diagram below, each of which represents an efficient reduction.



The goal of this lecture is to show that, perhaps surprisingly, the reductions indicated by dashed blue arrows in the diagram above hold. Specifically, we will show that:

*These notes are class material that has not undergone formal peer review. The TAs and I are grateful for any reports of typos.

- If one has an oracle for linear optimization on Ω, then one can efficiently "upgrade it" into an oracle for general convex optimization on Ω. This holds even if by "linear optimization oracle" we mean an algorithm that given any $c \in \mathbb{R}^n$ can return the minimum value of $c^\top x$ on Ω, but does not return a minimizer x^* at which this value is achieved.
- A mere *membership* oracle—that is, an algorithm that given any $y \in \mathbb{R}^n$ can decide if $y \in \Omega$ or not—can be efficiently turned into a separation oracle for Ω, and thus into an algorithm for generic convex optimization.

L11.1 Oracles

For this lecture we consider a set $\Omega \subseteq \mathbb{R}^n$ compact and convex with

$$\mathbb{B}_r(x_0) \subseteq \Omega \subseteq \mathbb{B}_R(0) \text{ for some } x_0 \in \Omega.$$

Given Ω, several types of algorithms can be constructed. In particular, we especially focus on these:

- A *membership oracle* takes as input a point $y \in \mathbb{R}^n$ and returns whether $y \in \Omega$ or not.
- A *separation oracle* takes as input a point $y \in \mathbb{R}^n$ and returns a separating direction if $y \notin \Omega$, or the statement "$y \in \Omega$" otherwise.
- A *Euclidean projection oracle* takes as input a point $y \in \mathbb{R}^n$ and returns the projection of y onto Ω.
- A *linear optimization oracle* takes as input a vector $c \in \mathbb{R}^n$ and returns the minimum value of $c^\top x$ over $x \in \Omega$, and optionally a minimizer of this function.

For the purposes of this lecture, by "efficient reduction" between oracles we mean that the reduction can be implemented with a number of calls to the original oracle that is at most polynomial in the dimension n of the space. Furthermore, for simplicity we will not be concerned with approximation errors, nor with representation issues for the numbers involved.

L11.2 From membership to separation

The ability to convert a membership oracle into a separation oracle was first shown by Yudin, D. B., & Nemirovskii, A. S. [YN76] and then crystallized by Grötschel, M., Lovász, L., & Schrijver, A. [GLS93] in their book on the ellipsoid method. A more modern construction was proposed recently by Lee, Y. T., Sidford, A., & Vempala, S. S. [LSV18]. We will follow this latter construction.

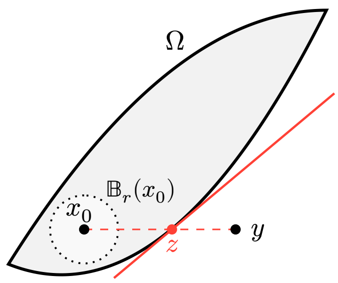
L11.2.1 Intuition

Let's start with the easy case: if $\|y\|_2 > R$, since $\Omega \subseteq \mathbb{B}_R(0)$ we can simply separate y from the ball $\mathbb{B}_R(0)$ by returning the halfspace $\{x \in \mathbb{R}^n : y^\top x \le R\|y\|_2\}$. So, we focus on the case where $\|y\|_2 \le R$.

The main idea hinges on a different construction for a separating hyperplane, which does not rely on the Euclidean projection of the point y onto Ω, unlike what we saw in Lecture 5.

Instead, we will construct a separation oracle as follows:

1. We start from x_0, and move on a straight line towards y until we hit the boundary of the set Ω. Call the boundary point z.
2. We will compute the equation for a plane tangent at the boundary of Ω at z. Such a plane is called a *supporting hyperplane* of Ω at z.[1]



Crucially, figuring out how much we can move from x_0 in the direction of y can be done via binary search on the membership of the points $x_0 + \alpha(y - x_0)$ for $\alpha \in [0, 1]$.

The second step is more delicate, in that we need to perform some kind of "numeric differentiation" along the boundary of Ω to find a supporting hyperplane.

L11.2.2 The depth function

In order to figure out the equation of the supporting hyperplane, we will need a way of constructing points that are close to z and on the boundary of Ω. A natural idea is to construct such points by moving in the direction of $y - x_0$ starting from a point close to x_0, until the boundary is reached, that is, until the membership oracle starts signaling that we have gone too far and left Ω.

In order to make the argument formal, let's introduce the function $d_y : \Omega \to \mathbb{R}$ that informs us of how far we can move in the direction $y - x$ starting from any $x \in \Omega$ until we reach the boundary of Ω:

$$d_y : \Omega \to \mathbb{R}, \qquad d_y(x) := \max\left\{ \alpha \in \mathbb{R}_{\geq 0} : x + \alpha \frac{y - x_0}{\|y - x_0\|_2} \in \Omega \right\}.$$

The maximum is guaranteed to exist because the set on the right is

- nonempty, since the value $\alpha = 0$ always guarantees the condition on the right;
- closed, since Ω is closed; and
- bounded, since $\Omega \subseteq \mathbb{B}_R(0)$. [▷ You should be able to produce a formal proof!]

Furthermore, by convexity of Ω, the set on the right-hand side is guaranteed to be an interval.

We now study the properties of the depth function $d_y(x)$. The proofs of these results are easy exercises, and are left to the appendix at the end of this lecture note [▷ Try to prove these on your own before reading the proofs].

Theorem L11.1. For any given $x \in \Omega$, we have $d_y(x) \in [0, 2R]$. Furthermore, we can compute $d_y(x)$ up to $\epsilon > 0$ error using $O(\log(R/\epsilon))$ calls to a membership oracle for Ω.

Theorem L11.2. The function d_y is concave (*i.e.*, $-d_y$ is convex).

[1] Such a plane need not be unique. [▷ Think of an example.]

Theorem L11.3. The function d_y restricted to the domain $\mathbb{B}_{r/3}(x_0)$ is $(3R/r)$-Lipschitz continuous.

L11.2.3 Constructing the supporting hyperplane

Given the depth function d_y, we can now construct the supporting hyperplane at z by performing numeric differentiation at x_0. A key issue that we will need to overcome is that—while Lipschitz continuous; see Theorem L11.3—the depth function might not differentiable at x_0. We will discuss how to address this issue in a second. Before we do that though, let's first establish the connection between the depth function and the supporting hyperplane under the idealized hypothesis that d_y is differentiable at x_0, just to make sure that the overall direction is aligned with our objectives.

Theorem L11.4. If d_y is differentiable at x_0, the gradient $\nabla d_y(x_0)$ provides a separating direction, that is,

$$\langle \nabla d_y(x_0), y - x \rangle < 0 \qquad \forall x \in \Omega.$$

Proof. By concavity, we know that first-order approximation of the function is a global upper bound,[2] and so

$$d_y(x) \le d_y(x_0) + \langle \nabla d_y(x_0), x - x_0 \rangle \qquad \forall x \in \Omega. \tag{1}$$

In particular, consider the point

$$x' := x_0 - \frac{r}{2R}(y - x_0).$$

This point satisfies $\|x' - x_0\|_2 \le r$ since $\|y - x_0\|_2 \le 2R$. Hence, we have that $x' \in \Omega$. Furthermore,

$$d_y(x') = d_y(x_0) + \frac{r}{2R}\|y - x_0\|_2.$$

Plugging into (1), we then find

$$\begin{aligned} & d_y(x_0) + \frac{r}{2R}\|y - x_0\|_2 \le d_y(x_0) - \frac{r}{2R}\langle \nabla d_y(x_0), y - x_0 \rangle \\ \implies \quad & \|y - x_0\|_2 \le \langle \nabla d_y(x_0), x_0 - y \rangle. \end{aligned}$$

Furthermore, since $y \notin \Omega$, we have $d_y(x_0) < \|y - x_0\|_2$, and so we can write

$$d_y(x_0) < \langle \nabla d_y(x_0), x_0 - y \rangle.$$

Plugging the previous inequality into (1), we find

[2] While Theorem L4.1 was stated for a function differentiable at all points in its domain, it is immediate to see that the proof only required differentiability at the particular point used in the linearization. [▷ You should prove this!]

$$d_y(x) < \langle \nabla d_y(x_0), x - y \rangle \qquad \forall x \in \Omega.$$

Finally, using the fact that $d_y(x) \geq 0$ for all $x \in \Omega$ by Theorem L11.1, the statement follows. □

Theorem L11.4 shows that, as we suspected, the gradient of the depth function at x_0 provides a separating direction. Unfortunately, the depth function might not be differentiable at x_0.

The key to overcoming this issue is that convex functions have very nice properties, and in particular, they are twice differentiable almost everywhere (see Alexandrov's theorem and Rademacher's theorem). So, we can first perturb x_0 slightly to a nearby point $x_{0'}$ where the depth function is differentiable, and then use a numerical gradient of the depth function at $x_{0'}$ as a separating direction. You can take a look at the original paper by Lee, Y. T., Sidford, A., & Vempala, S. S. [LSV18] for the technical details.

L11.3 From linear optimization to separation: Polarity

We now turn our attention to the other connection promised in the introduction: the fact that a linear optimization oracle can be converted into a separation oracle. In particular, this implies that a linear optimization oracle can always be promoted into an oracle for convex optimization. As a special case, this implies that if we know how to optimize linear functions on a closed convex set Ω, we can in particular project onto Ω.

A key notion that will help us establish this connection is that of the *polar set*. The polar set Ω° of a set $\Omega \subseteq \mathbb{R}^n$ that contains the origin is another convex compact set. Crucially, it exhibits some form of "duality" with respect to Ω, in the sense that linear optimization on Ω can be converted into membership in Ω°, and vice versa. We thus reach the following diagram, that completes the picture started in the introduction.

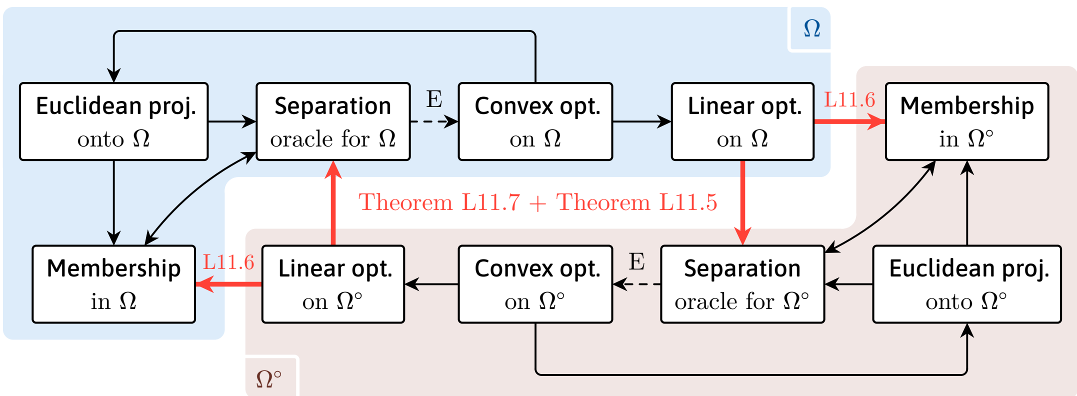


Figure 1: Connections among oracles between a set Ω and its polar Ω°. "E": Ellipsoid method.

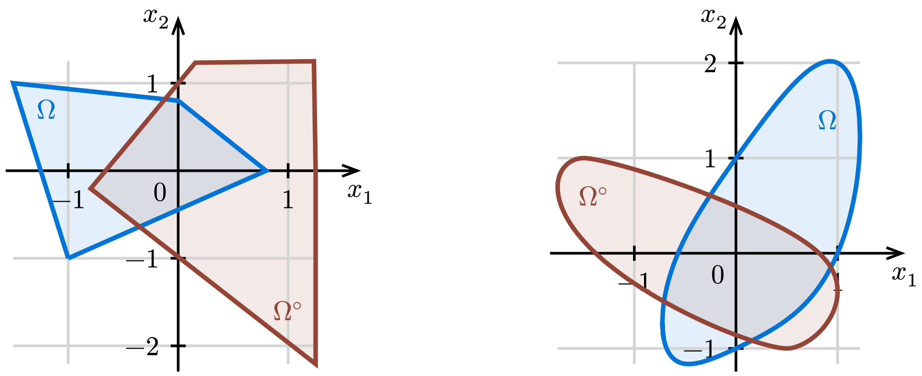
L11.3.1 The polar set Ω°

We start from defining the notion of polar set.

Definition L11.1 (Polar set Ω°)**.** Let Ω be compact and convex, and such that $\mathbb{B}_r(0) \subseteq \Omega \subseteq \mathbb{B}_R(0)$ for some radii $0 < r \le R$. The polar set Ω° to Ω is defined as

$$\Omega^\circ := \{y \in \mathbb{R}^n : \langle x, y\rangle \le 1 \quad \forall x \in \Omega\}.$$

As an example, for each of the two plots below, each set is polar to the other.



The polar set satisfies similar properties to Ω, as we now show. Furthermore, a very important property of polarity is that it is an involution, that is, the polar of the polar is the original set.

Theorem L11.5. Let the set $\Omega \subseteq \mathbb{R}^n$ be convex and compact, and such that $\mathbb{B}_r(0) \subseteq \Omega \subseteq \mathbb{B}_R(0)$ for some $0 < r \le R$. Then,

1. the polar set Ω° is convex and compact, and satisfies $\mathbb{B}_{1/R}(0) \subseteq \Omega^\circ \subseteq \mathbb{B}_{1/r}(0)$; and
2. the bipolar $(\Omega^\circ)^\circ$ is equal to Ω.

Proof. We prove the two points separately.

1. The set Ω° is convex and closed since by definition it is an intersection of halfspaces. Furthermore, since $\Omega \subseteq \mathbb{B}_R(0)$, for all $y \in \mathbb{B}_{1/R}(0)$ and $x \in \Omega$ we have

$$\langle y, x\rangle \le \|y\|_2\|x\|_2 \le (1/R)\cdot R \le 1 \qquad \Longrightarrow \qquad \mathbb{B}_{1/R}(0) \subseteq \Omega^\circ.$$

   We now show that $\Omega^\circ \subseteq \mathbb{B}_{1/r}(0)$. Pick any $y \in \Omega^\circ$, and consider the vector $z := ry/\|y\|_2 \in \mathbb{B}_r(0) \subseteq \Omega$. Being in Ω°, y must then satisfy

$$1 \ge \langle y, z\rangle = \left\langle y, r\frac{y}{\|y\|_2}\right\rangle = r\|y\|_2.$$

   Hence, $\|y\|_2 \le 1/r$. This shows that Ω° is bounded, and since we know it is closed, we conclude that Ω° is compact, concluding the proof.

2. We prove the two directions separately.

- ($\supseteq$) We start from the easy direction, and show that $\Omega \subseteq (\Omega^\circ)^\circ$. Pick any $x \in \Omega$. Showing that $x \in (\Omega^\circ)^\circ$ means showing that $\langle x, y \rangle \le 1$ for all $y \in \Omega^\circ$. But this is direct, since $y \in \Omega^\circ$ implies $\langle x, y \rangle \le 1$ by definition of the polar.
- ($\subseteq$) We show that $(\Omega^\circ)^\circ \subseteq \Omega$. Let $x \in (\Omega^\circ)^\circ$. Assume for contradiction that $x \notin \Omega$. Since Ω is closed and convex, we can then separate x from Ω (Lecture 5). Hence, there exist $u \in \mathbb{R}^n, v \in \mathbb{R}$ such that

$$\langle u, x \rangle > v, \qquad \text{and} \qquad \langle u, y \rangle \le v \quad \text{for all } y \in \Omega.$$

(Note that necessarily $u \neq 0$ or the above inequalities would be inconsistent.) Since $y = ru/\|u\|_2$ belongs to $\mathbb{B}_r(0)$, which is contained in Ω, then in particular $v > 0$. But then, we can divide the inequality on the right by v and find that

$$\left\langle \frac{u}{v}, y \right\rangle \le 1 \qquad \forall y \in \Omega \qquad \implies \qquad \frac{u}{v} \in \Omega^\circ.$$

However, remember that by hypothesis, we assumed that $x \in (\Omega^\circ)^\circ$, and so $\left\langle \frac{u}{v}, x \right\rangle \le 1$. This contradicts the fact that $\langle u, x \rangle > v$. Hence, we conclude that $x \in \Omega$.

□

L11.3.2 Separation over the polar

Assume for now that a point $x_0 \in \Omega$ such that $\mathbb{B}_r(x_0) \subseteq \Omega$ is known. We will show that a linear optimization oracle for Ω can be converted into a membership oracle for the polar set Ω°.

Theorem L11.6. A membership oracle for Ω° can be constructed efficiently starting from a linear optimization oracle for Ω, even if the linear optimization oracle only returns the optimal objective value and not the minimizer.

The construction only requires a single call to the linear optimization oracle.

Proof. Without loss of generality, we assume that $x_0 = 0$. If not, it suffices to shift both Ω and y by the quantity $-x_0$ so that $\mathbb{B}_r(0) \subseteq \Omega$, and continue the argument with the understanding that both Ω and y have been shifted.

Let $y \in \mathbb{R}^n$. Determining if $y \in \Omega^\circ$ means, by definition of the polar, checking whether $\langle y, x \rangle \le 1$ for all $x \in \Omega$. We can do so by optimizing the linear function $\langle y, x \rangle$ over $x \in \Omega$ using the linear optimization oracle. If the optimal value is less than or equal to 1, we return that $y \in \Omega^\circ$. Else, we return that $y \notin \Omega^\circ$. □

Theorem L11.6 provides the top, rightward thick red arrow in Figure 1. By invoking the bipolarity theorem of Theorem L11.5, it follows that Theorem L11.6 also establishes the bottom, leftward thick red arrow in the diagram.

In order to upgrade the linear optimization oracle for Ω into a separation oracle for Ω°, we can use the results discussed in Section L11.2. However, with very little extra work we can also show that the separation oracle for Ω° can be constructed directly starting from a linear optimization oracle for Ω without resorting to the randomized algorithm discussed in Section L11.2, as long as the linear optimization oracle also returns a minimizer.

Theorem L11.7. A separation oracle for Ω° can be constructed efficiently, without use of Section L11.2, starting from a linear optimization oracle for Ω that returns the optimal point.

The construction only requires a single call to the linear optimization oracle.

Proof. Let $y \in \mathbb{R}^n$. As discussed above, if $\max_{x\in\Omega}\langle y, x\rangle \le 1$, then $y \in \Omega^\circ$, and our job is done. Conversely, suppose that $\max_{x\in\Omega}\langle y, x\rangle > 1$, and let $x^\star \in \Omega$ be a maximizer. Then, the direction $d \coloneqq x^\star$ is a separating direction, because $\langle d, y\rangle > 1$, but $\langle d, x\rangle \le 1$ for all $x \in \Omega^\circ$ by definition of polar set. □

Theorem L11.7 provides the right, downward thick red arrow in the diagram of Figure 1. By invoking the bipolarity theorem of Theorem L11.5, it follows that Theorem L11.7 also establishes the left, upward thick red arrow in the diagram.

L11.3.3 Getting rid of the assumption of knowing x_0

Finally, we show that it is possible to convert a linear optimization oracle into a separation oracle, even when the location of the interior point x_0 is unknown. This is in contrast to the conversion from membership to separation, which requires knowledge of x_0 upfront.

▸ **General idea.** The idea for the reduction is to consider an augmented set $\Omega' \subseteq \mathbb{R}^{n+1}$ constructed from Ω, in which we know by construction where an interior x_0' lies. Specifically, consider the convex hull

$$\Omega' \coloneqq \left\{\lambda\begin{pmatrix}x\\0\end{pmatrix} + (1-\lambda)w : x \in \Omega, w \in \mathbb{B}_1(e_{n+1}), \lambda \in [0,1]\right\},$$

where $e_{n+1} \coloneqq (0, ..., 0, 1) \in \mathbb{R}^{n+1}$ is the $(n+1)$-st indicator vector.

It is straightforward to check [▹ You should check!] that Ω' is convex, and it satisfies the inclusions

$$\mathbb{B}_1(e_{n+1}) \subseteq \Omega' \subseteq \mathbb{B}_{R+1}(0).$$

Hence, the point $x_0' \coloneqq e_{n+1}$ is a known interior point, and the ball $\mathbb{B}_1(x_0')$ is fully contained in Ω'.

▸ **Optimization oracle for Ω'.** Before we can invoke the construction with the polar discussed above, we need to ensure that in the process of augmenting Ω into Ω', we have not lost the ability to optimize linear objectives on Ω'.

It is very easy to check [▹ You should check!] that optimization over Ω' can be decomposed greedily according to

$$\max_{x' \in \Omega'} \langle c', x' \rangle = \max\left\{ \max_{x \in \Omega} \left\langle c', \begin{pmatrix} x \\ 0 \end{pmatrix} \right\rangle, \max_{w \in \mathbb{B}_1(x'_0)} \langle c', w \rangle \right\}$$

$$= \max\left\{ \max_{x \in \Omega} \left\langle c', \begin{pmatrix} x \\ 0 \end{pmatrix} \right\rangle, \langle c', x'_0 \rangle + \|c'\|_2 \right\}.$$

The first element in the maximum on the right-hand side can be computed using the original linear optimization oracle for Ω, and so a linear optimization oracle for Ω' can be constructed from one for Ω efficiently (*i.e.*, with only a polynomial-time overhead).

▸ **Separation oracle for Ω.** Since we have a linear optimization oracle for Ω', and we know a point x'_0 in the interior of Ω', we are now in the position of leveraging the result in Section L11.3.1 and obtain a separation oracle for Ω'. To complete the construction, we still need to show that such a separation oracle can be efficiently modified into a separation oracle for Ω.

Suppose we want to separate the point $y \in \mathbb{R}^n$ from Ω. We can construct the point $y' := (y, 0) \in \mathbb{R}^{n+1}$, and query our separation oracle for Ω'.

- Suppose that the separation oracle for Ω' returns that $(y, 0) \in \Omega'$. By inspecting the definition of Ω', it must then be the case that $y \in \Omega$.
- Conversely, suppose that the separation oracle for Ω' returns a separating direction $d' = (d, \gamma) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$\left\langle \begin{pmatrix} d \\ \gamma \end{pmatrix}, \begin{pmatrix} y \\ 0 \end{pmatrix} - \lambda \begin{pmatrix} x \\ 0 \end{pmatrix} - (1-\lambda) w \right\rangle < 0 \qquad \forall x \in \Omega, w \in \mathbb{B}_1(x'_0), \lambda \in [0,1].$$

  Then, the condition holds in particular for $\lambda = 1$, and we find

$$\langle d, y - x \rangle < 0 \qquad \forall x \in \Omega,$$

  implying that d is a direction separating y from Ω.

Bibliography for this lecture

[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`

[LSV18] Lee, Y. T., Sidford, A., & Vempala, S. S. (2018). Efficient convex optimization with membership oracles. *Conference on Learning Theory*, 1292–1294.

L11.A Appendix

L11.A.1 Proof of Theorem L11.1

Theorem L11.1 (Restated)**.** For any given $x \in \Omega$, we have $d_y(x) \in [0, 2R]$. Furthermore, we can compute $d_y(x)$ up to $\epsilon > 0$ error using $O(\log(R/\epsilon))$ calls to a membership oracle for Ω.

Proof. Intuitively, since Ω is contained in a ball of radius R centered in 0, when α becomes too large we must leave the ball and therefore Ω. In particular, for $\alpha > 2R$,

$$\begin{aligned}\left\|\left(x + \alpha\frac{y - x_0}{\|y - x_0\|_2}\right)\right\|_2 &\geq \alpha\left\|\frac{y - x_0}{\|y - x_0\|_2}\right\|_2 - \|x\|_2 \\ &\geq \alpha - R \\ &> R.\end{aligned}$$

Hence, $d_y \in [0, 2R]$. Since the interval has length $2R$, in order to get the desired $\varepsilon > 0$ precision it then suffices to run the binary search for $O(\log(R/\epsilon))$ iterations. □

L11.A.2 Proof of Theorem L11.2

Theorem L11.2 (Restated)**.** The function d_y is concave (*i.e.*, $-d_y$ is convex).

Proof. Pick any $x, x' \in \Omega$, and $t \in [0, 1]$. We need to show that $d_y(tx + (1-t)x') \geq td_y(x) + (1-t)d_y(x')$.

By definition of d_y, the points

$$w := x + d_y(x)\frac{y - x_0}{\|y - x_0\|_2}, \qquad w' := x' + d_y(x')\frac{y - x_0}{\|y - x_0\|_2}$$

belong to Ω. Since Ω is convex, the point $tw + (1-t)w'$ also belongs to Ω. Hence, we have

$$tw + (1-t)w = (tx + (1-t)x') + \big(td_y(x) + (1-t)d_y(x')\big)\frac{y - x_0}{\|y - x_0\|_2} \in \Omega.$$

This implies that $td_y(x) + (1-t)d_y(x') \in \mathbb{R}_{\geq 0}$ belongs to the set

$$\left\{\alpha \in \mathbb{R}_{\geq 0} : (tx + (1-t)x') + \alpha\frac{y - x_0}{\|y - x_0\|_2} \in \Omega\right\},$$

and therefore

$$d_y(tx + (1-t)x') \geq td_y(x) + (1-t)d_y(x')$$

by the definition of d_y. □

L11.A.3 Proof of Theorem L11.3

Theorem L11.3 (Restated). The function d_y restricted to the domain $\mathbb{B}_{r/3}(x_0)$ is $(3R/r)$ -Lipschitz continuous.

Proof. Let $x, x' \in \mathbb{B}_{r/3}(x_0)$. We want to show that $|d_y(x) - d_y(x')| \le (3R/r)\|x - x'\|_2$. If $x = x'$, the result is trivial, so we assume $x \ne x'$. The key insight is that the function d_y is concave, and furthermore $d_y(x'') \ge 0$ for all $x'' \in \mathbb{B}_r(x_0)$ as shown in Theorem L11.1. These two facts combined imply that d_y cannot decrease too rapidly [▷ Make a drawing to convince yourself before jumping into the algebra].

In particular, consider the point

$$x'' := x' + t(x - x'), \quad \text{where } t := \frac{2}{3}\frac{r}{\|x - x'\|_2}.$$

Using the triangle inequality, this point satisfies

$$\|x''\|_2 \le \|x'\|_2 + \frac{2}{3}\frac{r}{\|x - x'\|_2}\|x - x'\|_2 \le \frac{r}{3} + \frac{2r}{3} = r.$$

Hence, $d_y(x'') \ge 0$. Furthermore, x is a convex combination of x' and x'', as

$$x = \left(1 - \frac{1}{t}\right)x' + \frac{1}{t}x''$$

and $t \ge 1$ since $\|x - x'\|_2 \le \|x\|_2 + \|x'\|_2 \le \frac{2}{3}r$. By the concavity of d_y, we have

$$\begin{aligned} d_y(x) &\ge \left(1 - \frac{1}{t}\right)d_y(x') + \frac{1}{t}d_y(x'') \\ &\ge \left(1 - \frac{1}{t}\right)d_y(x') \qquad\qquad \big(\text{since } d_y(x'') \ge 0\big) \\ &= \left(1 - \frac{3}{2}\frac{\|x - x'\|}{r}\right)d_y(x'). \end{aligned}$$

Rearranging, we finally find

$$\begin{aligned} d_y(x') - d_y(x) &\le \left(\frac{3}{2}\frac{d_y(x')}{r}\right)\|x - x'\|_2 \\ &\le \frac{3R}{r}\|x - x'\|_2 \qquad\qquad \big(\text{since } d_y(x') \le 2R;\ \text{see Theorem L11.1}\big). \end{aligned}$$

Finally, repeating the analysis swapping the roles of x and x', we obtain

$$d_y(x) - d_y(x') \le \frac{3R}{r}\|x - x'\|_2.$$

Together, the inequalities imply

$$\big|d_y(x) - d_y(x')\big| \le \frac{3R}{r}\|x - x'\|_2$$

as we wanted to show. □

Changelog

- `Mar 13, 2025: Fixed denominator from R to 2R (spotted in class)`
- `Mar 17, 2025: Fixed typo (thanks Jonathan Huang!)`