Lecture 6: Hyperplane separation theorems

Rajat Mittal *

IIT Kanpur

Last time we looked at generalized inequalities on cones. These inequalities transform to the membership question in a proper cone. It turns out that for solving these convex optimization problems over the cones, we need to understand the membership question. In today's lecture there will be various theorems which will help in this matter.

1 Krein-Milman Theorem

We studied two different characterizations of a bounded polytope last time. One, it was the bounded feasible region of a bunch of inequalities and equalities. Second, it can be expressed as the convex hull of its vertices. The equivalence of these two characterizations is known as *Krein-Milman Theorem*

Theorem 1. *Krein-Milman: Any convex, compact and non-empty set $A \in \mathbb{R}^n$ can be written as the convex hull of its extreme points.*

Notice that the theorem is much more general and talks about any compact, convex set and not just bounded polytopes. In the case of bounded polytopes, the number of extreme points will be finite. We will not cover the proof of this theorem. But lets look at one of the nice implications.

Lemma 1. *Suppose, given a linear function $f : \mathbb{R}^n \to \mathbb{R}, f = a^T x$ and a bounded polytope S. There always exist an extremal point of S which attains the minimum/maximum value of f on S.*

Remark: Note that there can be other points, not extremal, which also attain the maximum or minimum.

Proof. We will show the proof for maximum, same works for minimum value. Suppose the maximum is attained at $x_0 \in \mathbb{R}^n$ and the vertices of S are $x_1, \cdots, x_k$. By Krein-Milman, S is the convex hull of its vertices. Hence,

$$x_0 = \theta_1 x_1 + \cdots + \theta_k x_k,\ \forall i,\ \theta_i \geq 0, \sum_i \theta_i = 1$$

.

Because f is linear,

$$f(x_0) = \theta_1 f(x_1) + \cdots + \theta_k f(x_k)$$

.

This shows that there exists $i \in 1, 2, \cdots, m$, s.t., $f(x_i) \geq f(x_0)$ (why?). Hence the maximum will be attained at one of the vertices. $\square$

Notice the importance of this theorem. Suppose, given a feasible region which is a bounded polytope and a linear function which we need to optimize. This lemma tells us that the optimal value will lie on a vertex and hence cuts the search space by a large amount.

2 Hyperplane separation theorems

There is a general theorem that two disjoint convex sets can be *separated* by a hyperplane. Depending on the convex sets the separation can be different.

* Thanks to books from Boyd and Vandenberghe, Dantzig and Thapa, Papadimitriou and Steiglitz

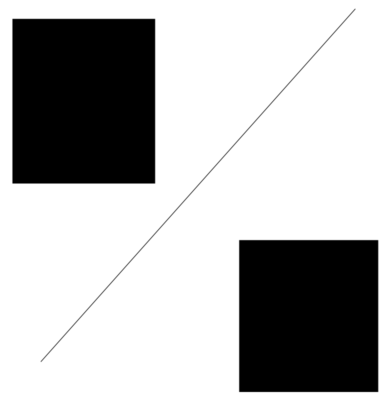

Fig. 1. Convex sets and hyperplanes separating them

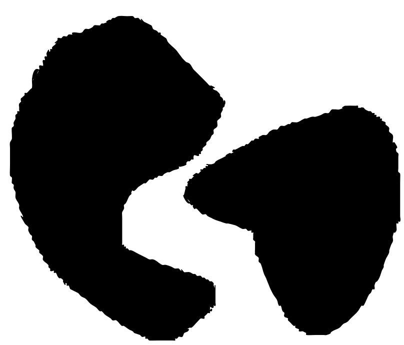

Fig. 2. Non-convex sets

Formally a hyperplane $a^T x - b = 0$ separates sets C and D, if

$$a^T x - b \geq 0, \ \forall x \in C \text{ and } a^T x - b \leq 0, \ \forall x \in D$$

. The same is called a strict separation if strict inequalities hold on both sides.

$$a^T x - b > 0, \ \forall x \in C \text{ and } a^T x - b < 0, \ \forall x \in D$$

.

Exercise 1. Give an example of two sets which cannot be separated.

2.1 Two convex sets

Theorem 2. *Given two convex sets C and D, which are mutually disjoint. There always exists a hyperplane $a^T x - b = 0$ separating them. Hence $a^T x - b \geq 0, \ \forall x \in C$ and $a^T x - b \leq 0, \ \forall x \in D$.*

Notice that the separation here is not strict.

Exercise 2. Construct two convex sets where separation cannot be strict.

Proof. Sketch: We show the proof sketch for two closed sets, such that, one of them is compact. In this case actually strict separation can be achieved. Since one set is compact and other is closed, there exist two points $c \in C$ and $d \in D$ whose distance is minimum among any pair of points, one from C and one from d. We can take the perpendicular bisector between c and d and that will be the separating hyperplane. If it is not the separating hyperplane then there exists a point of C (or D) on the hyperplane. Then we can construct a point closer to d (or c) from the set C (or D). $\square$

2.2 A point and a convex set

Our next example will be a point and a convex set. In this case we get a strict separation by the hyperplane, s.t., point lies on one side of the hyperplane and the set on the other side. Here strict means both the point and the set are disjoint with the hyperplane.

Theorem 3. *Given a closed convex set C and a point p. There always exist a hyperplane $H = a^T x - b$ which strictly separates them. So $a^T p - b > 0$ and $a^T x - b < 0, \ \forall x \in C$.*

The proof follows by constructing a small enough ball around point p which is compact and disjoint with C. How will you get strict separation.

Corollary 1. *Any closed convex set is the intersection of all the half spaces which contain it.*

Proof is left as an exercise.

2.3 Farkas lemma: A point and a cone

A special case of the previous section is the separation between a point and a finitely generated cone. Remember that a finitely generated cone is convex. The proof follows from previous theorems, but it will be instructive to see another proof of this.

Lemma 2. *Farkas: Given a set of vectors $x_1, \cdots, x_k \in \mathbb{R}^n$ (equivalently $C \in \mathbb{R}^{n \times k}$) and a point $b \in \mathbb{R}^n$. Exactly one of the following two conditions are satisfied.*

1. *$\exists \theta \in \mathbb{R}^k_+$ (θ in positive orthant), such that, $b = C\theta$.*
2. *$\exists a \in \mathbb{R}^n$, such that, $a^T b > 0$ and $a^T C \leq 0$ (entrywise).*

Proof. It is clear that both conditions cannot be satisfied simultaneously. Because then,

$$0 \geq a^T C \Rightarrow 0 \geq a^T C\theta \Rightarrow a^T b > 0$$

, which is a contradiction.

Suppose b can't be written as the conic combination of columns of C, i.e., first condition does not hold. Then look at cone generated by columns of C. It is finitely generated cone and hence by Weyl's theorem, it can be expressed as a bunch of inequalities,

$$Cone(x_1, \cdots, x_k) = \{x : Ax \leq 0\}$$

. Because b is not a member of this cone, there exists a row of A whose inner product with b is strictly positive (call it a_i). Then $a_i^T b > 0$ and $\forall i,\ a_i^T x_i \leq 0 \Rightarrow a_i^T C \leq 0$. □

Exercise 3. Interpret Farkas lemma as a hyperplane separation theorem. What do you know about this hyperplane?

This theorem converts the membership question in a cone to finding a separating hyperplane question. The question of membership in the cone is of real importance in convex optimization. It can be shown that optimization over a cone can be done using polynomially many calls to membership/separation algorithm for the cone.