# Lecture 17: Duality and Conic Programming

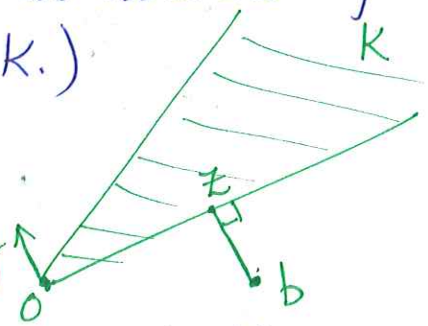
**Lemma.** Let $K \subseteq V$ be a closed convex cone. Then $(K^*)^* = K$.

**Proof.** It is easy to see that $K \subseteq (K^*)^*$. For the other direction, we need the following separation theorem.

**Theorem.** Let $K \subseteq V$ be a closed convex cone and let $b \in V \setminus K$. Then $\exists y \in V$ such that

$$(y, x) \geq 0 \;\; \forall x \in K \text{ and } (y, b) < 0.$$

(Thus the hyperplane with $y$ as its normal separates the point $b$ from the cone $K$.)




Observe that the theorem implies the above lemma. Suppose $b \notin K$. Then $\exists y \in V$ s.t. $\underbrace{(y, x) \geq 0 \;\forall x \in K}_{\text{so } y \in K^*}$ and $(y, b) < 0$. Thus $b \notin (K^*)^*$.

So $(K^*)^* \subseteq K$.

---

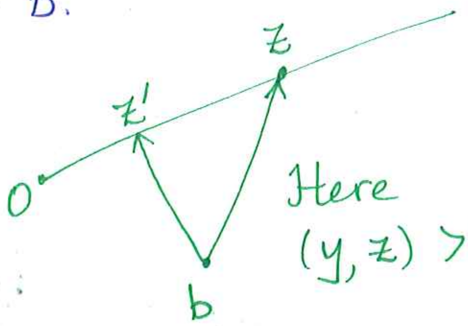
Let us prove the above separation theorem now. Let $z$ be the point of $K$ closest to $b$ (the $z-b$ distance $\|z-b\| = \sqrt{(z-b, z-b)}$). We will show that $y = z - b$ is the desired vector.

Note that the existence of a "nearest point" $z$ follows from the general theory: if $C$ is a nonempty closed and convex set in a finite-dimensional vector space $V$ and $b \in V$ is arbitrary, then there exists a unique $z \in C$ nearest to $b$ among all points of $C$.

So $z \in K$ is the nearest point of $K$ to $b$.

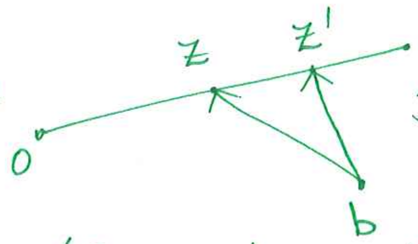
Let $y = z - b$. Let us show that $\underbrace{(y, z) = 0}_{\text{this is so if } z = 0}$.

So let us assume $z \neq 0$. Then by slightly moving along the ray $\{tz : t \geq 0\}$ either towards $0$ (if $(y, z) > 0$) or away from $0$ (if $(y, z) < 0$), we will get a point $z' \in K$ that is closer to $b$ than $z$, contradicting $z$ being the nearest point of $K$ to $b$.




Here $(y, z) > 0$.

or




Here $(y, z) < 0$.

[Proof by picture - please prove this formally.]

Thus $(y, z) = 0$. Let us now show that $(y, b) < 0$.

Observe that $y \neq 0$ (since $b \notin K$) and $b = z - y$.

Thus $(y, b) = (y, z - y) = \underbrace{(y, z)}_{=0} - \underbrace{(y, y)}_{>0} < 0$.

What is left to show is that $(y, x) \geq 0$ for all $x \in K$.
The angle $\angle bzx$ has to be at least $90^\circ$ – otherwise points on the segment $zx$ sufficiently close to $z$ would be nearer to $b$ than $z$ (an analogous argument to why $(y, z) = 0$), a contradiction. Thus $(b - z, x - z) \leq 0$

So $0 \geq (b - z, x - z) = (-y, x - z) = -(y, x) + \underbrace{(y, z)}_{=0}$
$= -(y, +x)$.

Thus $(y, x) \geq 0$. ☒

## The Farkas Lemma

Let us recall the Farkas lemma from linear programming.

**Lemma.** Let $A \in \mathbb{R}^{m\times n}$ be an $m \times n$ matrix and let $b \in \mathbb{R}^m$. Then

- Either the system $Ax = b$, $x \geq 0$ has a solution $x \in \mathbb{R}^n$
- Or the system $A^T y \geq 0$, $b^T y < 0$ has a solution $y \in \mathbb{R}^m$

but not both.

Interestingly, the Farkas lemma is a special case of the separation theorem that we just proved. To see this, let $V = \mathbb{R}^m$ and $K = \{Ax : x \in \mathbb{R}^n_+\} \subseteq V$.

Thus $K$ consists of all nonnegative linear combinations of columns of $A$. This is a finitely generated cone. Such cones are closed and convex, hence the separation theorem applies.

The system $Ax = b$, $x \geq 0$ having no solution means that $b \notin K$. By the separation theorem, there exists $y \in V = \mathbb{R}^m$ such that $\underbrace{y^T A x \geq 0 \text{ for all } x \in \mathbb{R}^n_+}_{\text{This means that } A^T y \in (\mathbb{R}^n_+)^* = \mathbb{R}^n_+.}$ and $y^T b < 0$.

So there is indeed a solution to the second system $A^T y \geq 0$, $b^T y < 0$. What is left to show is that both the systems are not simultaneously satisfiable. This is easy to check: if there exists $x \in \mathbb{R}^n_+$ s.t. $Ax = b$ and there exists $y \in \mathbb{R}^m$ s.t. $A^T y \geq 0$ and $b^T y < 0$ then $x^T A^T y \geq 0$, i.e., $b^T y \geq 0$, a contradiction to $b^T y < 0$.

Now we want to generalize the Farkas lemma to deal with systems of the form $A(x) = b$, $x \in K$, where $K \subseteq V$ is some closed convex cone, and $A$ is a linear operator from $V$ to $W$.

The "standard" Farkas lemma deals with the case $K = \mathbb{R}^n_+$, $V = \mathbb{R}^n$, and $W = \mathbb{R}^m$. Here a linear operator is represented by a matrix. In SDP, we need to consider the case $K = PSD_n \subseteq V = SYM_n$ and $W = \mathbb{R}^m$.

What $A^T$ is for a matrix needs to be generalized for a general linear operator $A$. When $A$ is a matrix,

$(y, Ax) = y^T A x = (A^T y)^T x = (A^T y, x)$.

So let us define $A^T$ for a linear operator $A$ exactly as given above.

Definition. Let $A: V \to W$ be a linear operator. A linear operator $A^T: W \to V$ is called an adjoint of $A$ if $(y, Ax) = (A^T y, x)$ $\forall x \in V$ and $y \in W$.

In SDP, we have $V = SYM_n$ and $W = \mathbb{R}^m$. The adjoint is easy to determine here.

Claim. Let $V = SYM_n$ and $W = \mathbb{R}^m$. Let $A: V \to W$ be defined as $A(X) = (A_1 \circ X, A_2 \circ X, \ldots, A_m \circ X)$.

Then $A^T(y) = \sum_{i=1}^{m} y_i A_i$.

---

Reference

1. Approximation Algorithms and SDP (Chapter 4)
(B. Gärtner and J. Matoušek)