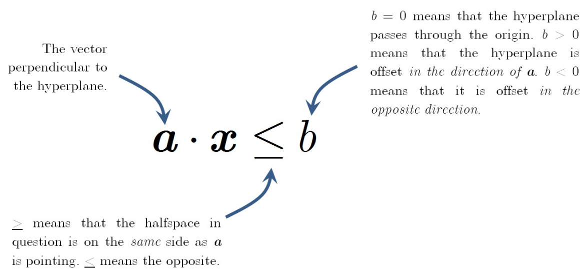
Convex Optimization
Review Session 1

This review session will cover the material on convex sets from lectures. The material is interesting in that the concepts involved aren't the most difficult (mostly the definition of a convex set and the separating hyperplane theorem), but the *application* of these concepts can be fiendishly difficult, depending on the problem.

As such, this first review session will focus mostly on solving various problems on these topics.

That said, there is one quick topic I want to review, and that's the topic of hyperplanes and halfspaces. It is often useful, when solving problem, to be able to visualize the halfspace in question, and I don't know about you, but I've always found it fiendishly difficult to actually visualize a halfspace based on the vector definition. Here's a guide that should make it easier



Having said that, let's get going with the questions!

Question 1 (*Convex Sets)

[1]Which of the following sets are convex?

1. The set of points closer to one set than another

$$\{\mathbf{x} : \mathrm{dist}(\mathbf{x}, \mathscr{S}) \leq \mathrm{dist}(\mathbf{x}, \mathscr{T})\}$$

   where $\mathscr{S}, \mathscr{T} \subseteq \mathbb{R}^n$ and

$$\mathrm{dist}(\mathbf{x}, \mathscr{S}) = \inf_{\mathbf{z} \in \mathscr{S}} \{\|\mathbf{x} - \mathbf{z}\|_2\}$$

2. The set $\{\mathbf{x} : \mathbf{x} + \mathscr{S}_2 \subseteq \mathscr{S}_1\}$, where $\mathscr{S}_1, \mathscr{S}_2 \subseteq \mathbb{R}^n$ and $\mathscr{S}_1$ is convex.

3. The set of point whose distance to $\mathbf{a}$ does not exceed a fixed fraction θ of the distance to $\mathbf{b}$

$$\{\mathbf{x} : \|\mathbf{x} - \mathbf{a}\|_2 \leq \theta \|\mathbf{x} - \mathbf{b}\|_2\}$$

[1]Partly inspired by Exercise 2.12 in Boyd.

You can assume $\mathbf{a} \neq \mathbf{b}$ and $\theta \in [0, 1]$.

4. Let $\mathbf{x} \in \mathbb{R}^m$ be the set of coefficients in a Fourier series $p_{\mathbf{x}}(t) = \sum_{k=1}^{m} x_k \cos(kt)$. The set of interest is the set of coefficients which ensure $|p_{\mathbf{x}}(t)|$ less than or equal 1 for all $|t| \leq \frac{\pi}{3}$

$$\left\{ \mathbf{x} \in \mathbb{R}^m : |p_{\mathbf{x}}(t)| \leq 1 \text{ for } |t| \leq \frac{\pi}{3} \right\}$$

Solution

1. Not necessarily convex. Consider

$$\mathscr{S} = \{-1, 1\} \qquad\qquad \mathscr{T} = \{0\}$$

The points -1 and 1 are closer to $\mathscr{S}$, but their midpoint is closer to $\mathscr{T}$.

2. Intuitively, what we're doing here is placing the set $\mathscr{S}_2$ onto $\mathscr{S}_1$, and finding the set of vectors over which we can move $\mathscr{S}_2$ without rotating it while keeping it within $\mathscr{S}_1$. Since $\mathscr{S}_1$ is convex, it makes sense that this generated set would be too.

Let's prove this. Consider that

$$\begin{aligned} \mathscr{S} &= \{\mathbf{x} : \mathbf{x} + \mathscr{S}_2 \subseteq \mathscr{S}_1\} \\ &= \{\mathbf{x} : \mathbf{x} + \mathbf{y} \in \mathscr{S}_1 \ \forall \mathbf{y} \in \mathscr{S}_2\} \\ &= \bigcap_{\mathbf{y} \in \mathscr{S}_2} \{\mathbf{x} : \mathbf{x} + \mathbf{y} \in \mathscr{S}_1\} \\ &= \bigcap_{\mathbf{y} \in \mathscr{S}_2} (\mathscr{S}_1 - \mathbf{y}) \end{aligned}$$

Each of the sets in the intersection is an affine transformation of $\mathscr{S}_1$, and so since $\mathscr{S}_1$ is convex, so is each set in the intersection. Thus, we have an infinite intersection of convex sets, which is also convex.

3. Let's engage in some algebraic hocus-pocus

$$\begin{aligned} \mathscr{S} &= \{\mathbf{x} : \|\mathbf{x} - \mathbf{a}\|_2 \leq \theta \|\mathbf{x} - \mathbf{b}\|_2\} \\ &= \left\{\mathbf{x} : \|\mathbf{x} - \mathbf{a}\|_2^2 \leq \theta^2 \|\mathbf{x} - \mathbf{b}\|_2^2\right\} \\ &= \left\{\mathbf{x} : (1 - \theta^2)\|\mathbf{x}\|_2^2 + 2(\theta^2 \mathbf{b} - \mathbf{a}) \cdot \mathbf{x} + \|\mathbf{a}\|_2^2 - \theta^2 \|\mathbf{b}\|_2^2 \leq 0\right\} \end{aligned}$$

If $\theta = 1$, the quadratic term in $\mathbf{x}$ disappears, and we're left with a (convex) hyperplane. If not, let $\mathbf{k} = 2(\theta^2 \mathbf{b} - \mathbf{a})$. We can then write the above as

$$\mathscr{S} = \left\{\mathbf{x} : \|\mathbf{x} - \mathbf{k}\|_2^2 \leq \text{Constant}\right\|$$

This is a ball, which is also convex. Thus, $\mathscr{S}$ is convex.

4. Consider that the set in question can be written as

$$\bigcap_{|t|\leq\frac{\pi}{3}} \{\mathbf{x}\in\mathbb{R}^m : |p(t)|\leq 1\}$$

Consider, however, that we can write

$$p(t) = (\cos(t),\ldots,\cos(mt)^\top\mathbf{x}$$

As such, we can write the above as

$$\bigcap_{|t|\leq\frac{\pi}{3}} \left\{\mathbf{x}\in\mathbb{R}^m : -1\leq(\cos(t),\ldots,\cos(mt)^\top\mathbf{x}\leq 1\right\}$$

Each of the sets is the intersection of two hyperplanes (since the cosine vector is constant) and therefore convex. Thus, we have an infinite intersection of convex sets, which is also convex.

■ □ ■

Question 2 (*Midpoint convexity)

[2]A set $\mathscr{C}$ is *midpoint convex* if

$$\mathbf{a},\mathbf{b}\in\mathscr{C}\Rightarrow\frac{\mathbf{a}+\mathbf{b}}{2}\in\mathscr{C}$$

Clearly, all convex sets are midpoint convex. Show that under a mild condition that you should specify, midpoint convexity implies convexity.

Solution

Let's first understand this intuitively. We need to show that given any two points in the set, any point on the line between those two sets is also in the set. Assuming we know the set is midpoint convex, one intuitive way to do this is simply to go halfway along the line. Then halfway again. Then again, etc..., until we get to the desired point in the limit. This last word should cue you in to the condition required – we need the *limit* of a set of point to be in the set – in other words, we need the set to be closed.

Thus, we'll prove that if a set $\mathscr{C}$ is closed and given $\mathbf{a},\mathbf{b}\in\mathscr{C}$ and $\theta\in[0,1]$,

$$\frac{1}{2}(\mathbf{a}+\mathbf{b})\in\mathscr{C}\Rightarrow\theta\mathbf{a}+(1-\theta)\mathbf{b}\in\mathscr{C}$$

Now consider performing m bisections on the line between $\mathbf{a}$ and $\mathbf{b}$, described by the vector $\mathbf{g}\in\{0,1\}^m$, where

$$g_k = \begin{cases} 1 & \text{If bisection number } m-k+1 \text{ is towards } \mathbf{a} \\ 0 & \text{If that bisection is towards } \mathbf{b} \end{cases}$$

[2]Exercise 2.3 in Boyd.

The vector we will reach after these bisections is

$$\mathbf{z}_m^{\mathbf{g}} = \gamma_m^{\mathbf{g}}\mathbf{a} + (1 - \gamma_m^{\mathbf{g}})\mathbf{b} \in \mathscr{C}$$

where

$$\gamma_m^{\mathbf{g}} = \sum_{k=1}^{m} \mathbf{g}_k 2^{-k}$$

and this vector *must* be in $\mathscr{C}$ by midpoint convexity.

Now it is clear that we can find a $\tilde{\mathbf{g}}$ so that

$$\lim_{m\to\infty} \gamma_m^{\tilde{\mathbf{g}}} = \theta$$

We would then have

$$\lim_{m\to\infty} \mathbf{z}_m^{\tilde{\mathbf{g}}} = \theta\mathbf{a} + (1-\theta)\mathbf{b}$$

We have therefore found a sequence that tends to $\theta\mathbf{a} + (1-\theta)\mathbf{b}$ in which each element is in the set. Assuming the set is closed, the limit must also be in the set. Thus, the set is convex.

■ □ ■

Question 3 (*Convex Hulls)

[3] Show that the convex hull of a set $\mathscr{S}$ (denoted $\text{conv}(\mathscr{S})$) is the intersection of all convex sets that contain $\mathscr{S}$.

Solution

Let $\mathbb{C}$ be the set of convex sets containing $\mathscr{S}$

$$\mathbb{C} = \{\mathscr{C} : \mathscr{C} \text{ is convex}, \mathscr{S} \subseteq \mathscr{C}\}$$

Clearly, $\text{conv}(\mathscr{S}) \in \mathbb{C}$, and so

$$\bigcap_{\mathscr{C}\in\mathbb{C}} \mathscr{C} \subseteq \text{conv}(\mathscr{S})$$

To prove the other direction, we need to show that every point in the convex hull $\text{conv}(\mathscr{S})$ must also appear in *every* convex set in $\mathbb{C}$ (and therefore in their intersections). This is trivial, however, because

- Every point in $\text{conv}(\mathscr{S})$ is a convex combination of points in $\mathscr{S}$.
- Every set in $\mathbb{C}$ must contain every point in $\mathscr{S}$.

[3] Exercise 2.4 in Boyd.

- Every set in $\mathbb{C}$ must be convex and therefore contain every convex combination of points in $\mathscr{S}$.

Thus, every point in conv($\mathscr{S}$) is also in every set in $\mathbb{C}$. Thus,

$$\text{conv}(\mathscr{S}) \subseteq \bigcap_{\mathscr{C}\in\mathbb{C}} \mathscr{C}$$

Thus,

$$\text{conv}(\mathscr{S}) = \bigcap_{\mathscr{C}\in\mathbb{C}} \mathscr{C}$$

■ □ ■

Question 4 (**Using Helly's Theorem)

[4]Let $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \subset \mathbb{R}^2$, $k > 2$. Using Helly's Theorem, devise a method to compute the smallest covering radius for X efficiently (ie: with complexity that grows polynomially with k).

Recall that the smallest covering radius of X (denoted r_X^*) is the radius of the smallest circle that contains all points in X.

Solution

First note that

$$\begin{aligned} r_X^* &= \min\left\{r : \bigcap_{i=1}^{k} B_r(\mathbf{x}_i) \neq \varphi\right\} \\ &= \sup\left\{r : \bigcap_{i=1}^{k} B_r(\mathbf{x}_i) = \varphi\right\} \end{aligned}$$

(Where $B_r(\mathbf{x}) = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leq r\}$).

Consider the second formulation above, and note that

$$\begin{aligned} r_X^* > \rho &\Leftrightarrow \bigcap_{i=1}^{k} B_\rho(\mathbf{x}_i) = \varphi \\ &\Leftrightarrow \exists\{i_1, i_2, i_3\} \text{ s.t. } B_\rho(\mathbf{x}_{i_1}) \cap B_\rho(\mathbf{x}_{i_2}) \cap B_\rho(\mathbf{x}_{i_3}) = \varphi \\ &\Leftrightarrow \exists\{i_1, i_2, i_3\} \text{ s.t. } r^*_{\{\mathbf{x}_{i_1},\mathbf{x}_{i_2},\mathbf{x}_{i_3}\}} > \rho \end{aligned}$$

(The second step follows by Helly's Theorem).

Thus, we have that

$$r_X^* = \max\left\{r^*_{\{\mathbf{x}_{i_1},\mathbf{x}_{i_2},\mathbf{x}_{i_3}\}} : \{i_1, i_2, i_3\} \in \{1, 2, \cdots, k\}, i_1 \neq i_2 \neq i_3\right\}$$

[4]Taken from a Homework from Garud Iyengar's Convex Optimization course, Spring 2011

The number of sets of three such points to check is

$$\frac{k(k-1)(k-2)}{3!} = \tfrac{1}{6}k^3 - 3k^2 + 2k$$

This is clearly polynomial in k. Furthermore, for a triple $\mathscr{T} = \{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \mathbf{x}_{i_3}\}$, the smallest covering radius is simple to calculate and given by

$$\begin{aligned} r^*_{\mathscr{T}} &= \max\left\{\tfrac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|, \tfrac{1}{2}\|\mathbf{x}_2 - \mathbf{x}_3\|, \tfrac{1}{2}\|\mathbf{x}_3 - \mathbf{x}_1\| \right. \\ &\qquad \left. \tfrac{1}{2}\|\mathbf{x}_1 - \tfrac{1}{3}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3)\right\} \end{aligned}$$

That is, the maximum of

- Half the distance between any two of the three points.
- The distance between each point and the center of the triangle formed by the three points.

This is illustrated in figure 1.

■ □ ■

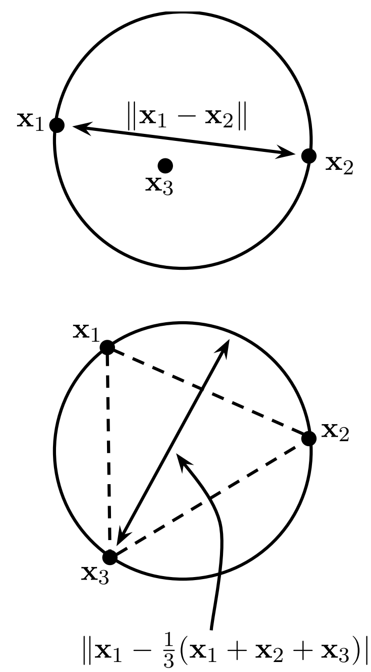


Figure 1: Depending on the configuration of the three points, the minimum covering radius of the three points can take one of four different values.

Question 5 (**Farkas' Lemma)

Prove Farkas' Lemma without using linear programming duality. Recall that Farkas' Lemma states that, given $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, exactly one of the following two statements is true

1. There exists an $\mathbf{x} \in \mathbb{R}^n$ with $\mathbf{x} \geq \mathbf{0}$ such that $A\mathbf{x} = \mathbf{b}$.
2. There exists a $\mathbf{y} \in \mathbb{R}^m$ with $\mathbf{b} \cdot \mathbf{y} < 0$ such that $A^\top \mathbf{y} \geq 0$.

Solution

To answer this question, we need to reformulate the two statements above geometrically. First, define the following cone

$$\mathscr{C} = \{\mathbf{z} : \mathbf{z} = A\mathbf{x}, \mathbf{x} \geq \mathbf{0}\}$$

It is the cone generated by the columns of A. Now, note that the two statements above can be re-written as follows:

1. The first statement states that $\mathbf{b}$ consists of a positive linear combination of the columns of A – in other words, that $\mathbf{b} \in \mathscr{C}$.
2. The second statement states that there is a vector $\mathbf{y}$ that makes an acute angle with with every column of A (the dot product of every column of A with $\mathbf{y}$ is positive), but that makes an obtuse angle with $\mathbf{b}$.

All we need to do is show that the second statement is equivalent to the statement $\mathbf{b} \notin \mathscr{C}$, and we're home, because $\mathbf{b}$ is *in* $\mathscr{C}$ or out of it, so only one of the two statements above can be true.

- Statement 2 $\Rightarrow$ Outside $\mathscr{C}$ Suppose the vector $\mathbf{b}$ satisfies statement 2, and consider an arbitrary point $\mathbf{c} \in \mathscr{C}$, with $\mathbf{c} = A\mathbf{x}$. We then have

  $$\mathbf{c} \cdot \mathbf{y} = \mathbf{c}^\top \mathbf{y} = \mathbf{x}^\top A^\top \mathbf{y} = \mathbf{x} \cdot A^\top \mathbf{y} \geq 0$$

  (The last step follows because $A^\top \mathbf{y} \geq \mathbf{0}$ by statement 2, and so we have the dot product of two positive vectors).

  However, since $\mathbf{b}$ satisfies the second statement above, we know that $\mathbf{b} \cdot \mathbf{y} < 0$. As such

  $$\mathbf{b} \cdot \mathbf{y} < \mathbf{c} \cdot \mathbf{y} \qquad \forall \mathbf{c} \in \mathscr{C}$$

  Thus, $\mathbf{y}$ is a hyperplane that strictly separates $\mathbf{b}$ from $\mathscr{C}$. Clearly, therefore, $\mathbf{b} \notin \mathscr{C}$.

- Outside $\mathscr{C}$ $\Rightarrow$ Statement 2 This direction is harder, and will require the Separating Hyperplane Theorem.

  Suppose $\mathbf{b} \notin \mathscr{C}$. Since $\mathscr{C}$ is convex[1] and closed, there is a strictly separating hyperplane $\mathbf{y}$ such that

  $$\mathbf{b} \cdot \mathbf{y} < \mathbf{c} \cdot \mathbf{y} \qquad \forall \mathbf{c} \in \mathscr{C} \tag{1}$$

  Now, consider

  - We have $\mathbf{0} \in \mathscr{C}$. Thus, putting $\mathbf{c} = \mathbf{0}$ in equation 1, we find that

    $$\mathbf{b} \cdot \mathbf{y} < 0$$

  - Consider the RHS of equation 1, and imagine that we had some $\mathbf{c} \in \mathscr{C}$ such that $\mathbf{c} \cdot \mathbf{y} < 0$. By the definition of a cone, we can multiply $\mathbf{c}$ by any arbitrary positive constant and still keep it in the cone. By doing this with an arbitrarily large constant, it becomes clear that the RHS of equation 1 would become unbounded below. This is clearly impossible, since the LHS is finite. Thus, we must have

    $$\mathbf{c} \cdot \mathbf{y} \geq 0 \qquad \forall \mathbf{c} \in \mathscr{C} \tag{2}$$

    Now, each column of A is, in its own right, a member of the cone $\mathscr{C}$. Thus, by successively setting $\mathbf{c}$ in equation 2 to each column of A, we find that the inner product of each column with $\mathbf{y}$ is positive. Thus,

    $$A^\top y \geq 0$$

  We have therefore shown that $\mathbf{b}$ satisfies both the conditions in statement 2.

[1] If $\mathbf{z}_1 = A\mathbf{x}_1$ and $\mathbf{z}_2 = A\mathbf{x}_2$ are both in $\mathscr{C}$, then for $\lambda \in [0,1]$, $\mathbf{z}_3 = \lambda\mathbf{z}_1 + (1-\lambda)\mathbf{z}_2$ must also be in $\mathscr{C}$, because $\mathbf{z}_3 = A\left(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2\right)$, and since $\mathbf{x}_1$ and $\mathbf{x}_2$ are both ≥ 0, so is $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$

■ □ ■

Question 6 (*Dual Cones)

A cone $\mathscr{K}$ is said to be *proper* if

- $\mathscr{K}$ is convex.
- $\mathscr{K}$ is closed.
- $\mathscr{K}$ has a non-empty interior.
- $\mathscr{K}$ is pointed (ie: $\mathbf{x} \in \mathscr{K}, -\mathbf{x} \in \mathscr{K} \Rightarrow \mathbf{x} = \mathbf{0}$).

Consider a proper cone $\mathscr{K}$. The dual cone of $setK$, denoted $\mathscr{K}^*$, is given by

$$\mathscr{K}^* = \{\mathbf{y} : \mathbf{x} \cdot \mathbf{y} \geq 0 \text{ for all } \mathbf{x} \in \mathscr{K}\}$$

We also define a *generalized inequality* with respect to the cone $\mathscr{K}$ as follows

$$\mathbf{x} \succeq_{\mathscr{K}} \mathbf{y} \Leftrightarrow \mathbf{y} - \mathbf{x} \in \mathscr{K}$$

$$\mathbf{x} \succ_{\mathscr{K}} \mathbf{y} \Leftrightarrow \mathbf{y} - \mathbf{x} \in \text{int}(\mathscr{K})$$

Given these definitions

- Show that $\mathscr{K}^{**} = \mathscr{K}$.
- Consider a set $\mathscr{C}$ (not necessarily convex). The element $\mathbf{x}$ is the *minimum element* of $\mathscr{C}$ with respect to $\succeq_{\mathscr{K}}$ if and only if $\mathbf{y} \succeq_{\mathscr{K}} \mathbf{x}$ for every $\mathbf{y} \in \mathscr{C}$.

  Show that $\mathbf{x}$ is the minimum element of $\mathscr{C}$ if and only if for all $\mathbf{g} \succ_{\mathscr{K}^*} \mathbf{0}$, $\mathbf{x}$ is the unique minimizer of $\mathbf{g} \cdot \mathbf{z}$ over $\mathbf{z} \in \mathscr{C}$.

Solution

Let's first crunch through the proof, and then using the insights gained there we'll go back and understand this intuitively.

First, we prove that $\mathscr{K}^{**} = \mathscr{K}$, in two parts.

- $\boxed{\mathscr{K} \subseteq \mathscr{K}^{**}}$ Take any vector $\mathbf{k} \in \mathscr{K}$. Consider that for any $\mathbf{x} \in \mathscr{K}^*$, $\mathbf{k} \cdot \mathbf{x} \geq 0$, by definition of $\mathscr{K}^*$. Since $\mathbf{x}$ was chosen arbitrarily, we have that

  $$\mathbf{k} \cdot \mathbf{x} \geq 0 \quad \forall \mathbf{x} \in \mathscr{K}^*$$

  Thus, $\mathbf{k} \in \mathscr{K}^{**}$.

- $\boxed{\mathscr{K}^{**} \subseteq \mathscr{K}}$ We will show that if $\mathbf{x} \notin \mathscr{K}$, then $\mathbf{x} \notin \mathscr{K}^{**}$.

  If $\mathbf{x} \notin \mathscr{K}$, then there is a strictly separating[2]hyperplane $\mathbf{g}$

[2]Note that the strict inequality (corresponding to strict separation) is only possible if $\mathscr{K}$ if closed. If not, the implications in this question only apply to $\text{cl}(\mathscr{K})$. Thus, in a more general case in which $\mathscr{K}$ is not closed, the true statement is $\mathscr{K}^{**} = \text{cl}(\mathscr{K})$.

such that

$$\mathbf{x} \cdot \mathbf{g} < \mathbf{k} \cdot \mathbf{g} \text{ for all } \mathbf{k} \in \mathscr{K} \tag{3}$$

Now, note that

- Since $\mathscr{K}$ is a cone, $\mathbf{0} \in \mathscr{K}$, and so putting $\mathbf{k} = \mathbf{0}$ into the RHS, we obtain

$$\mathbf{x} \cdot \mathbf{g} < 0$$

- It must be the case that $\mathbf{k} \cdot \mathbf{g} \geq 0$ for all $\mathbf{k} \in \mathscr{K}$. If that was not the case, multiplying $\mathbf{k}$ by an arbitrarily large constant α would result in an $\alpha\mathbf{k} \in \mathscr{K}$ with an arbitrarily large negative RHS in equation 3. This is impossible, since the LHS of equation 3 is finite.

  Thus, $\mathbf{g} \in \mathscr{K}^*$.[3]

Clearly, therefore, there is at least one vector in $\mathscr{K}^*$ whose dot product with $\mathbf{x}$ is not greater or equal to 0. Thus, $\mathbf{x} \notin \mathscr{K}^{**}$.

The main insight to pull out from the result above is that for each point $\notin \mathscr{K}$, the set $\mathscr{K}^*$ contains a hyperplane that separates that point from $\mathscr{K}$. Any such hyperplane *must* correspond to a point in $\mathscr{K}$. This might sound like a trivial result, but it's actually quite profound, and it's only true because $\mathscr{K}$ is a cone. If $\mathscr{K}$ were *not* a cone, this would not be true, because set $\mathscr{K}^*$ only contains hyperplanes of the form $\mathbf{x} \cdot \mathbf{y} \geq 0$ – in other words, only hyperplanes that pass through the origin. In the case of a cone, this is enough – any point not in the cone is separated from the cone by a hyperplane of that form. This is not the case with a general convex set. Figure 2 illustrates this concept.

Having said that, it is important to realize that though it might be tempting to take $\mathscr{K}$ and $\mathscr{K}^*$ and draw them on the same diagram, it'd be nonsense if we choose the interpret $\mathscr{K}$ as described in the previous paragraph. Indeed, even though both reside in $\mathbb{R}^n$, their substance is very different. The elements of $\mathscr{K}$ are vectors. The elements of $\mathscr{K}^*$ are *hyperplanes* – functions on vectors. It is unfortunate (or fortunate, depending on your point of view!) that in $\mathbb{R}^n$, both vectors and hyperplanes can be represented by a vector in $\mathbb{R}^n$.[4] Later in this course, we will encounter other spaces in which things aren't so simple.

The result that $\mathscr{K}^{**} = \mathscr{K}$ should then not be surprising given the interpretation above. The set $\mathscr{K}^{**}$ can be interpreted as 'those vectors that are separated from all other vectors by the hyperplanes in $\mathscr{K}^*$'. Since we know that a convex set is defined by its supporting hyperplanes, we do indeed get $\mathscr{K}$ back. This would certainly not be the case for a general convex set – indeed, the set

[3] Note that what we've just done is shown that for *any* point $\notin \mathscr{K}$, the set $\mathscr{K}^*$ contains a hyperplane that separates the point from $\mathscr{K}$. We'll be using this fact later...

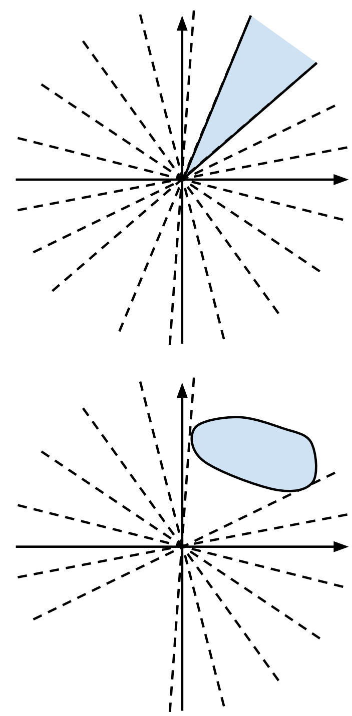


Figure 2: The top figure illustrates the fact that separating hyperplanes that pass through 0 only are sufficient in describing a convex cone. The bottom figure illustrates that for a general convex set, specifying the set of hyperplanes that separate the set and pass through 0 isn't enough – there are many more hyperplanes that separate the set and which are not in $\mathscr{K}$ because they do not pass through 0.

[4] It turns out that this is because $\mathbb{R}^n$ is a Hilbert space, equipped with an inner product. More on this in the last part of the course!

of hyperplanes that pass through 0 and separate such a set would not properly define the set, as illustrated in figure 2. Instead, they define the smallest convex cone that contains the set – thus, for a general convex set $\mathscr{C}$, $\mathscr{C}^{**}$ is in fact the smallest convex cone that contains $\mathscr{C}$.

Let's now proceed to the second part of the question. We will, once again, complete our proof in two parts

- $\boxed{\mathbf{x} \text{ minimum element } \Rightarrow \text{ condition}}$ Suppose $\mathbf{x}$ is the minimum element of $\mathscr{C}$ with respect to $\mathscr{K}$. Then $\mathbf{z} - \mathbf{x} \succeq_{\mathscr{K}} \mathbf{0}$ (in other words, $\mathbf{z} - \mathbf{x} \in \mathscr{K}$). Thus, given any $\mathbf{g} \succ_{\mathscr{K}^*} \mathbf{0}$ (ie: $\mathbf{g} \in \text{int}(\mathscr{K}^*)$), we have $\mathbf{g} \cdot (\mathbf{z} - \mathbf{x}) > 0$.[5] As such

  $$\mathbf{g} \cdot \mathbf{z} > \mathbf{g} \cdot \mathbf{x} \qquad \forall \mathbf{z} \in \mathscr{C}$$

  In other words, $\mathbf{g} \cdot \mathbf{x}$ is the unique minimizer of $\mathbf{g} \cdot \mathbf{z}$ over $\mathbf{z} \in \mathscr{C}$.

  This concept is illustrated for the case $\mathscr{K} = \mathbb{R}^2_+$ in Figure 3.

- $\boxed{\text{Condition } \Rightarrow \mathbf{x} \text{ minimum element}}$ Suppose $\mathbf{g}\cdot\mathbf{x}$ is the unique minimizer of $\mathbf{g} \cdot \mathbf{z}$ over $\mathbf{z} \in \mathscr{C}$ for all $\mathbf{g} \succ_{\mathscr{K}^*} \mathbf{0}$, but that $\mathbf{x}$ is not the minimum element. This means there exists another $\mathbf{z} \in \mathscr{C}$ such that $\mathbf{z} \not\succeq_{\mathscr{K}} \mathbf{x}$. Since $\mathbf{z} - \mathbf{x} \not\succeq_{\mathscr{K}}$ (ie: $\mathbf{z} - \mathbf{x} \notin \mathscr{K}$), there exists a strictly separating hyperplane which, by the result above, must correspond to a vector $\tilde{\mathbf{g}} \in \mathscr{K}^*$ (ie: $\tilde{\mathbf{g}} \succeq_{\mathscr{K}^*} \mathbf{0}$) with $\tilde{\mathbf{g}} \cdot (\mathbf{z} - \mathbf{x}) < 0$. Since the inequality is not strict, this is also true in a neighborhood of $\tilde{\mathbf{g}}$. This constradicts the assumption $\mathbf{x}$ is the unique minimizer of $\tilde{\mathbf{g}} \cdot \mathbf{z}$ over $\mathbf{z} \in \mathscr{C}$.

[5]Note the importance of the strict inequality in $\mathbf{g} \succ_{\mathscr{K}^*} \mathbf{0}$. Without it, the hyperplane would not be strictly separating and $\mathbf{g}\cdot\mathbf{x}$ would not be the *unique* minimizer.

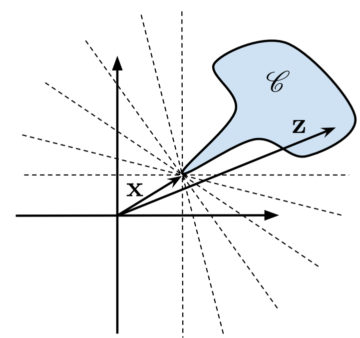


Figure 3: This figure illustrates the problem of finding the minimum element of set $\mathscr{C}$ with respect to the cone $\mathscr{K}$ corresponding to $\mathbb{R}_+$, the closed positive halfspace. The dotted lines indicate all the hyperplanes in $\mathscr{K}^*$ (shifted to meet the point $\mathbf{x}$ under consideration). Clearly, if the point sec x is to be the minimum element of $\mathscr{C}$, $\mathbf{g} \cdot \mathbf{x}$ must be less than $\mathbf{g} \cdot \mathbf{z}$ for every point $\mathbf{z} \in \mathscr{C}$, and every one of the dotted hyperplanes.

■ □ ■

Question 7 (**Polar sets)

The *polar* of a set $\mathscr{C} \subseteq \mathbb{R}^n$ is defined as the set

$$\mathscr{C}^\circ = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} \cdot \mathbf{x} \leq 1 \text{ for all } \mathbf{x} \in \mathscr{C}\}$$

Answer the following questions

- Show that $\mathscr{C}^\circ$ is convex, even if $\mathscr{C}$ is not.
- Show that if $\mathscr{C}$ is closed and convex with $\mathbf{0} \in \text{int}\mathscr{C}$, then $\mathscr{C}^{\circ\circ} = \mathscr{C}$.
- Explain intuitively why the condition $\mathbf{0} \in \text{int}\mathscr{C}$ is needed in the previous part. What is $\mathscr{C}^{\circ\circ}$ for a set that does not contain $\mathbf{0}$?

Solution

Let's first show that $\mathscr{C}$ is convex

$$\begin{aligned}\mathscr{C}^\circ &= \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} \cdot \mathbf{x} \leq 1 \text{ for all } \mathbf{x} \in \mathscr{C}\} \\ &= \bigcap_{\mathbf{x} \in \mathscr{C}} \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} \cdot \mathbf{x} \leq 1\}\end{aligned}$$

$\mathscr{C}^\circ$ is therefore an infinite intersection of (convex) halfspaces, and is therefore convex.

We show that $\mathscr{C} = \mathscr{C}^{\circ\circ}$ in two steps

- $\boxed{\mathscr{C} \subseteq \mathscr{C}^{\circ\circ}}$ Consider that for any $\mathbf{x} \in \mathscr{C}$, we have that

  $$\mathbf{x} \cdot \mathbf{y} \leq 1 \text{ for all } \mathbf{y} \in \mathscr{C}^\circ$$

  By definition, this means that $\mathbf{x} \in \mathscr{C}^{\circ\circ}$.

- $\boxed{(\mathscr{C}^\circ)^\circ \subseteq \mathscr{C}}$ We will show that $\mathbf{c} \notin \mathscr{C}$ implies that $\mathbf{c} \notin \mathscr{C}^{\circ\circ}$.

  Consider that if $\mathbf{c} \notin \mathscr{C}$, since $\mathscr{C}$ is closed and convex, there is a strictly separating hyperplane $\mathbf{g}$ such that

  $$\mathbf{c} \cdot \mathbf{g} > \mathbf{x} \cdot \mathbf{g} \text{ for all } \mathbf{x} \in \mathscr{C}$$

  Now, let's try to make this inequality as tight as possible by finding the largest value the RHS can take, which we'll denote

  $$\alpha = \sup_{\mathbf{x} \in \mathscr{C}} \{\mathbf{x} \cdot \mathbf{g}\}$$

  Note that

  - $\alpha \geq 0$, because $\mathbf{0} \in \mathscr{C}$.
  - Since $\mathbf{0} \in \text{int}\mathscr{C}$, an entire neighborhood of $\mathbf{0}$ is also contained in $\mathscr{C}$. By feeding various elements of that neighborhood into the RHS above, we find that the RHS therefore takes on both positive and negative values. Thus, $\alpha > 0$.
  - The RHS above is bounded above by the LHS, which is finite. Thus, $\alpha < \infty$.

  The three facts above mean that we can re-scale $\mathbf{g}$ to $\tilde{\mathbf{g}}$ to set $\alpha = 1$. This gives[6]

  $$\mathbf{c} \cdot \tilde{\mathbf{g}} > 1 = \sup_{\mathbf{x} \in \mathscr{C}} \{\mathbf{x} \cdot \tilde{\mathbf{g}}\}$$

  Now consider the two parts of this statement

  - The fact that $\sup_{\mathbf{x} \in \mathscr{C}} \{\mathbf{x} \cdot \mathbf{g}\} = 1$ tells us that $\tilde{\mathbf{g}} \in \mathscr{C}^\circ$.
  - The fact that $\mathbf{c} \cdot \tilde{\mathbf{g}} > 1$ tells us that there is at least one point in $\mathscr{C}^\circ$ whose dot product with $\mathbf{c}$ is strictly greater than 1. Thus, $\mathbf{c}$ cannot be in $\mathscr{C}^{\circ\circ}$.

[6] Just like in the previous question, note the importance of what we've just done here. We've shown that *any* hyperplane that separates our set $\mathscr{C}$ from any point has a corresponding element in the set $\mathscr{C}^\circ$. In the previous question, we found a way to do this for a cone. In this question, we've found a way to do this for *any* convex set, but we've had to add the proviso that the set contain $\mathbf{0}$ in its interior (more on that later).

The question that begs to be asked, of course, is why the requirement that $\mathbf{0} \in \text{int}(\mathscr{C})$ is so important. What happens if it isn't? The answer lies in the kind of hyperplanes that polars can encode. We saw in the previous question that dual cones only include hyperplanes that pass through the origin. Similarly, every hyperplane in the polar set must take the form $\mathbf{g} \cdot \mathbf{x} \leq$ a strictly positive number. This means that polar sets can only encode halfspaces that *do* include the origin. Figure 4 illustrates this idea.

Because of this, if $\mathscr{C}$ does not contain the origin, $\mathscr{C}^\circ$ is *not* able to encode every hyperplane that separates the set from an arbitrary point. This is illustrated in figure 5.

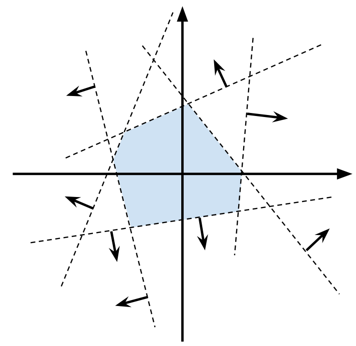


Figure 4: A sample of the kinds of halfspaces polar sets can encode. Note that every halfspace here contains the origin. Polar sets are not able to encode halfspaces that do not include the origin.

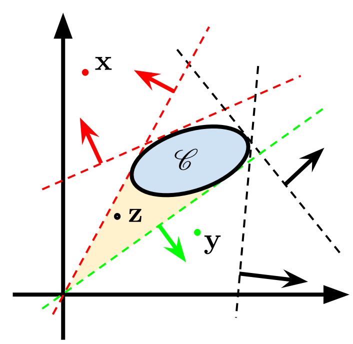


Figure 5: The polar set only includes halfpsaces that include the origin. So for example, the hyperplanes that contain $\mathscr{C}$ and separate it from $\mathbf{x}$ and $\mathbf{y}$ are in $\mathscr{C}^\circ$, whereas the hyperplanes that contain $\mathscr{C}$ and separate it from $\mathbf{z}$ (or indeed, any point in the yellow area above) are not.

■ □ ■

Question 8 (***Blackwell's Theorem)

[5]Consider the following model of a decision problem

- Let $S = \{s_1, \ldots, s_{n_s}\}$ be the finite set of states of nature and $\mathbf{p} \in \mathbb{R}^{n_s}$ be a vector of probabilities associated with each state of nature (later, we will let $\pi = \text{diag}(\mathbf{p})$, the matrix containing the components of $\mathbf{p}$ on its diagonal).
- The states are not observed directly – instead, a set of signals $Y = \{y_1, \ldots, y_{n_y}\}$ is observed. An *information structure* $Q \in \mathbb{R}^{n_s \times n_y}$ is a matrix of probabilities such that

$$Q_{ij} = \mathbb{P}(y_j \text{ displayed}|s_i \text{ happens})$$

- Let $A = \{a_1, \ldots, a_{n_a}\}$ be the set of actions available to the decision maker, and let the payoff matrix $U \in \mathbb{R}^{n_a \times n_s}$ be such that

$$U_{ij} = \text{ Payoff when action } a_i \text{ is taken in state } s_j$$

The decision maker observes signals and must decide which action to take. We encode these decisions in a matrix $D \in \mathbb{R}^{n_s \times n_a}$, which satisfies

$$D_{ij} = \mathbb{P}(\text{Taking action } a_j|\text{Observing signal } y_i)$$

For two information structures P and Q, Q is said to be *more informative* than P (denoted $P \subseteq Q$) if it allows a higher optimal expected payoff for all values of U and $\mathbf{p}$. Prove Blackwell's Theorem, which states that Q is more informative than P if and

[5]This question is based on *An elementary proof of Blackwell's theorem*, M. Leshno and Y. Spector, Mathematical Social Sciences 25(1992):95-98

only if there exists a Markov matrix M (ie: a matrix with entries between 0 and 1 and whose rows sum to 1) such that $P = QM$.

In the economics literature, M is called a *garbling matrix*, presumably because it 'garbles' some of the information in Q to produce a less informative matrix P.

Solution

We first need to find an expression for the optimal expected payoff for a given information matrix Q, payoff matrix U and probability vector $\mathbf{p}$. It turns out the expression is given by

$$F(Q, U, \mathbf{p}) = \max_{D \in \mathscr{M}} \operatorname{tr}(QDU\pi)$$

where $\mathscr{M}$ is the set of all Markov matrices. The maximization part is obvious – clearly, we need to maximize over our possible decisions D. We need, however, to understand the expression we are maximizing – why is the expected payoff of a given decision D given by $\operatorname{tr}(QDU\pi)$? Consider:

- DU outputs a matrix in which the rows correspond to signals and the columns correspond to states. For each signal and state, it provides the expected payoff when a certain signal is observed in a given state of nature.
- Each row of Q corresponds to a state. When we multiply it by the column corresponding to the *same* state in DU, the diagonal element that results is the expected payoff for a given state, because it weights each signal by the probability the signal will appear in that state.
- Finally, we multiply each of these diagonal elements by the probability of the state occurring, and sum them using the trace.

Thus, expressed more compactly, Blackwell's Theorem states that

$$F(Q, U, \mathbf{p}) \geq F(P, U, \mathbf{p}) \;\; \forall U, \mathbf{p} \Leftrightarrow \exists M \in \mathscr{M} \text{ s.t. } P = QM$$

Before we begin, let's introduce some notation. In proving this theorem, we will need to make use of the matrix inner product defined by

$$A \cdot B = \sum_i \sum_j A_{ij} B_{ij}$$

In other words, we take each component of one matrix, multiply it by the corresponding component in the other and add them. In a way, you can think of this as simply stacking the columns of A on top of each other to obtain a large vector – the matrix inner product is then the simple inner product on that augmented vector. The following three facts are proved in the margin

1. $AB \cdot C = A \cdot CB^\top$
2. $\text{tr}(PDU\pi) = PDU \cdot \pi$

Let's now prove both directions. Let M be a Markov matrix.

- $\boxed{P = QM \Rightarrow Q \text{ more informative than } P}$ This side is trivial. Consider that

$$
\begin{aligned}
F(P, U, \mathbf{p}) &= \max_{D \in \mathscr{M}} \text{tr}(PDU\pi) \\
&= \max_{D \in \mathscr{M}} \text{tr}(QMDU\pi) \qquad (4) \\
&\leq^7 \max_{D' \in \mathscr{M}} \text{tr}(QD'U\pi) \qquad (5) \\
&= F(Q, U, \mathbf{p})
\end{aligned}
$$

 This is clearly true for all U and $\mathbf{p}$, and so Q is indeed more informative than P.

- $\boxed{Q \text{ more informative than } P \Rightarrow \exists M \text{ s.t. } P = QM}$ We prove this by contradiction. Define a set

$$\mathscr{C}_Q = \{A : \exists M \in \mathscr{M} \text{ s.t. } A = QM\}$$

Suppose Q is more informative than P, but that for every Markov matrix M, we have $P \neq QM$. In other words, $P \notin \mathscr{C}_Q$.

Note, however, that the set $\mathscr{C}_Q$ is convex[8]and closed.

We now use the Separating Hyperplane Theorem [9] – since $\mathscr{C}_Q$ is convex and $P \notin \mathscr{C}_Q$, there is a Hyperplane $\hat{U}$ that separates them. In other words,

$$A \cdot \hat{U} < P \cdot \hat{U} \qquad \forall A \in \mathscr{C}_Q$$

Or alternatively,

$$QD \cdot \hat{U} < P \cdot \hat{U} \qquad \forall D \in \mathscr{M}$$

Now, construct a matrix $U^\top = \pi^{-1}\hat{U}$ and use it as a payoff matrix. This equation then becomes

$$QD \cdot \pi U^\top < P \cdot \pi U^\top \qquad \forall D \in \mathscr{M}$$

Applying property 1 of inner products above, this becomes

$$QDU \cdot \pi < PU \cdot \pi \qquad \forall D \in \mathscr{M}$$

Applying property 2 of inner products above, this becomes

$$\text{tr}(QDU\pi) < \text{tr}(PU\pi) \qquad \forall D \in \mathscr{M}$$

Maximizing both sides over all $D \in \mathscr{M}$, we get

$$\max_{D \in \mathscr{M}} \text{tr}(QDU\pi) < \text{tr}(PIU\pi)$$

Property 1: We will use 'Einstein notation', in which any repeated indices are assumed to be summed over. So for example, $A_{ij}B_{jk} = \sum_j A_{ij}B_{jk} = AB$. Note that $AB \cdot C = (AB)_{ij}C_{ij} = A_{ik}B_{kj}C_{ij} = A_{ik}C_{ij}B^\top_{jk} = A_{ik}(CB^\top)_{ik}$.

Property 2: Note that since *pi* is diagonal, $\text{tr}(PDU\pi) = (PDU)_{ii}\pi_{ii} = (PDU)_{ij}\pi_{ij} = PDU \cdot \pi$

[7] This step follows because since M and D are both Markov matrices, MD must also be a Markov matrix. This means that by setting $D' = MD$, we see that $QMDU\pi$ in line 4 is just one possible value of the quantity being maximized in line 5. If we don't restrict ourselves to that value of D' but instead maximize over *all* such values, we clearly end up with something larger.

[8] To see why, consider $A_1 = QM_1$ and $A_2 = QM_2$, both in $\mathscr{C}_Q$. Now consider some $\lambda \in [0, 1]$ and let $\tilde{\lambda} = 1 - \lambda$. Note that

$$\lambda A_1 + \tilde{\lambda} A_2 = Q\left(\lambda M_1 + \tilde{\lambda} M_2\right)$$

Furthermore, note that $\lambda M_1 + \tilde{\lambda} M_2$ is also a Markov matrix (this is trivial to check). Thus, the convex combination of the A's is of the form QD' with $D' \in \mathscr{M}$, and is therefore also in $\mathscr{C}_Q$.

[9] The matrix inner product can easily be shown to be a bona-fide inner product, so all our results for $\mathbb{R}^n$ extend here, including the separating hyperplane theorem. But if you're bothered with using a separating hyperplane argument on a set of matrices, just think of it as a vector consisting of the columns of the matrix, stacked on top of each other (as discussed above when we introduced the inner product). The matrix inner product is then the simple inner product on $\mathbb{R}^(n_a n_s)$, and all our results follow.

Clearly, $I \in \mathscr{M}$, and so replacing this I by D and maximizing over all $D \in \mathscr{M}$ can only increase the RHS. Thus,

$$\max_{D \in \mathscr{M}} \operatorname{tr}(QDU\pi) \leq \max_{D \in \mathscr{M}} \operatorname{tr}(PDU\pi)$$

Or in other words,

$$F(Q, U, \mathbf{p}) \leq F(P, U, \mathbf{p})$$

As such, we have found at least one payoff matrix U for which P provides a greater maximum expected gain than Q. This contradicts our assumption that Q is more informative than P.

■ □ ■