Class 2: Farkas' lemma and Carathéodory's theorem

In this class, we consider two well-known from convex analysis: the Farkas' lemma and the Carathéodory's theorem. Let us first define the following set:

$$K_m := \left\{ x \in \mathbb{R}^n \;\middle|\; x = \sum_{i=1}^m \lambda_i a^i,\ \lambda_i \geq 0,\ i = 1, \dots, m \right\} \tag{1}$$

for a given set of vectors $\{a^1, \dots, a^m\} \subset \mathbb{R}^n$. It is easy to see that K_m is a convex cone. The result below shows that for any $x \in \mathbb{R}^n$, there always exists $y \in K_m$ such that it is the nearest point in K_m to the point x. This actually shows that K_m is also a closed set.

Lemma 1. *For all $x \in \mathbb{R}^n$, there exists $y \in K_m$ such that*

$$\|x - y\| \leq \|x - z\| \quad \text{for all } z \in K_m. \tag{2}$$

Proof. If $x \in K_m$, then $y = x$ is already the nearest point in K_m to x. Thus, assume that $x \notin K_m$ and let us prove the result by induction on m. The claim holds trivially when $m = 1$. Then, assume that the result is true when $m = \ell - 1$ for some ℓ. Defining

$$K_m^i := \{x \in \mathbb{R}^n \mid x = \lambda_1 a_1 + \cdots + \lambda_{i-1} a^{i-1} + \lambda_{i+1} a^{i+1} + \cdots + \lambda_m a_m\},$$

we can say, from assumption, that there exists y^i such that

$$\|x - y^i\| \leq \|x - z\| \quad \text{for all } z \in K_m^i. \tag{3}$$

Note that $y^i \in K_m$ also holds because $K_m^i \subset K_m$ for all i. Now, define the following subspace of $\mathbb{R}^n$:

$$L_m := \left\{ x \in \mathbb{R}^n \;\middle|\; x = \sum_{i=1}^m \lambda_i a^i,\ \lambda_i \in \mathbb{R},\ i = 1, \dots, m \right\},$$

and consider the following two cases: (a) $x \in L_m$ and (b) $x \notin L_m$.

(a) Assume that $x \in L_m$ and let $z \in K_m$. We consider y as the nearest point to x among the vectors $y^1, \dots, y^m$, that is,

$$\|x - y\| \leq \|x - y^i\| \quad \text{for all } i = 1, \dots, m. \tag{4}$$

We will show that the inequality in (2) holds in this case. Since $x \in L_m$ and $z \in K_m$, there exist $\alpha_i \in \mathbb{R}$ and $\beta_i \geq 0$ with $i = 1, \dots, m$, such that

$$x = \sum_{i=1}^m \alpha_i a^i \quad \text{and} \quad z = \sum_{i=1}^m \beta_i a^i.$$

Now, define the index set $I(x) := \{i \mid \alpha_i < 0\}$ and the scalar

$$t := \min_{j \in I(x)} \left\{ \frac{\beta_j}{\beta_j - \alpha_j} \right\}.$$

Clearly, t is well-defined because $x \notin K_m$ guarantees that $I(x)$ is nonempty. If $j \in I(x)$, then we have $\beta_j \geq 0$ and $\beta_j - \alpha_j > \beta_j$, and thus $t \in [0, 1)$. Assume that $t = \beta_i/(\beta_i - \alpha_i)$ for some $i \in I(x)$. In this case, we have

$$t\alpha_i + (1-t)\beta_i = 0 \quad \text{and} \quad t\alpha_j + (1-t)\beta_j \geq 0 \quad \text{for all } j \in I(x).$$

If $j \notin I(x)$, then $\alpha_j \geq 0$, and we also obtain $t\alpha_j + (1-t)\beta_j \geq 0$. Therefore, we get

$$tx + (1-t)z = t\left(\sum_{i=1}^{m} \alpha_i a^i\right) + (1-t)\left(\sum_{i=1}^{m} \beta_i a^i\right) = \sum_{i=1}^{m} \big(t\alpha_i + (1-t)\beta_i\big) a^i \in K_m^i.$$

From (3) and (4), we have

$$\|x - y\| \leq \|x - y^i\| \leq \|x - (tx + (1-t)z)\| = (1-t)\|x - z\| \leq \|x - z\|,$$

and the proof is complete for this case.

(b) Assume that $x \notin L_m$. If $e^1, \ldots, e^p$ are the orthonormal basis of L_m, then we can define

$$x' := \langle x, e^1 \rangle e^1 + \cdots + \langle x, e^p \rangle e^p$$

Clearly, $x' \in L_m$ holds. From case (b), there exists $y \in K_m$ such that

$$\|x' - y\| \leq \|x' - z\| \quad \text{for all } z \in K_m. \tag{5}$$

For all $s = 1, \ldots, p$, we have

$$\langle x - x', e^s \rangle = \langle x, e^s \rangle - \langle x', e^s \rangle = \langle x, e^s \rangle - \langle x, e^s \rangle = 0$$

where the second equality holds because $\langle e^i, e^j \rangle = 0$ when $i \neq j$ and $\langle e_i, e_i \rangle = 1$ for all i. So, we obtain

$$\langle x - x', z \rangle = 0 \quad \text{for all } z \in L_m. \tag{6}$$

Moreover, we have

$$\begin{aligned} \|x - x'\|^2 + \|x' - z\|^2 &= \|x\|^2 - 2\langle x', x \rangle + 2\|x'\|^2 - 2\langle x', z \rangle + \|z\|^2 \\ &= \|x\|^2 - 2\langle x', x - x' \rangle - 2\langle x, z \rangle + \|z\|^2 \\ &= \|x - z\|^2 \end{aligned} \tag{7}$$

where the second equality holds from (6), and the third one holds also from (6) and because $x' \in L_m$. Therefore, for all $z \in K_m \subseteq L_m$, we get

$$\|x - y\|^2 \leq \|x - x'\|^2 + \|x' - y\|^2 < \|x - x'\|^2 + \|x' - z\|^2 = \|x - z\|^2,$$

where the first inequality holds from the triangle inequality, the second one from (5) and the third one from (7). This means that y is the closest point of K_m from x. □

Let us now consider the same set of vectors $\{a^1, \ldots, a^m\} \subset \mathbb{R}^n$ used in the definition of K_m, and define the following set:

$$C_m = \{y \in \mathbb{R}^n \mid \langle y, a^i \rangle \le 0, i = 1, \ldots, m\}. \tag{8}$$

Recall also that the polar cone of C_m is given by

$$C_m^\circ = \{x \in \mathbb{R}^n \mid \langle x, y \rangle \le 0 \text{ for all } y \in C_m\}.$$

Lemma 2. (Farkas' Lemma) *Let $\{a^1, \ldots, a^m\} \subset \mathbb{R}^n$ be a set of vectors. If K_m and C_m are defined as* (1) *and* (8)*, respectively, then*

$$K_m = C_m^\circ.$$

Proof. Let $x \in C_m^\circ$ and $y \in K_m$ be the nearest point of K_m to x, which exists from Lemma 1. We will first prove that

$$\langle a^j, x - y \rangle \le 0, \quad j = 1, \ldots, m \tag{9}$$

and

$$\langle -y, x - y \rangle \le 0 \tag{10}$$

hold. Assume otherwise that (9) does not hold for some j. For sufficiently small $t \in (0, 1)$ we obtain

$$\|x - (y + ta^j)\|^2 = \|(x - y) - ta^j\|^2 = \|x - y\|^2 - 2t\langle a^j, x - y \rangle + t^2\|a^j\|^2 < \|x - y\|^2.$$

Observing that $y + ta^j \in K_m$ because K_m is a convex cone, the above inequalty then contradicts the fact that y is the nearest point of K_m to x. Similarly, if we assume that (10) does not hold, then for sufficiently small $t \in (0, 1)$, we get

$$\|x - (y - ty)\|^2 = \|(x - y) + ty\|^2 = \|x - y\|^2 - 2t\langle -y, x - y) + t^2\|y\|^2 < \|x - y\|^2.$$

For such t, we also have $y - ty = (1 - t)y \in K_m$, which is a contradiction.
Now, from (9), we have $x - y \in C_m$. Moreover, the definition of polar cone yields

$$\langle x, x - y \rangle \le 0.$$

This inequality, together with (10), gives

$$0 \ge \langle x, x - y \rangle + \langle -y, x - y \rangle = \|x - y\|^2.$$

Since the norm is always nonnegative, it means that $x = y$. Therefore, $x \in K_m$.
Now, let us assume that $x \in K_m$. Then, there exists $\lambda_i \ge 0$ with $i = 1, \ldots, m$, such that

$$\langle x, y \rangle = \sum_{i=1}^{k} \lambda_i \langle a^i, y \rangle \le 0$$

for all $y \in C_m$. Therefore, we conclude that $x \in C_m^\circ$. □

Definition 1. *A point $x \in \mathbb{R}^n$ is a* convex combination *of $\{x^1, \ldots, x^m\} \subset \mathbb{R}^n$ if it can be written as*

$$x = \sum_{i=1}^{m} \alpha_i x^i \quad \textit{for some } \alpha_i \textit{ with } \alpha_i \geq 0,\ i = 1, \ldots, m,\ \textit{and} \sum_{i=1}^{m} \alpha_i = 1.$$

Definition 2. *The* convex hull *of a set S, denoted by $\operatorname{co} S$, is the smallest convex set that contains S.*

As the name suggests, the convex hull $\operatorname{co} S$ is always convex.

Lemma 3. *Let $x \in \mathbb{R}^n$ be defined as a convex combination of $m \geq n+2$ points in $\mathbb{R}^n$. Then, it is possible to choose $n+1$ points among the m points and write x as as a convex combination of these selected points.*

Proof. Let x be a convex combination of $m \geq n+2$ points $x^1, \ldots, x^m$, that is,

$$x = \sum_{i=1}^{m} \alpha_i x^i, \quad \text{with} \quad \alpha_i > 0, \quad \sum_{i=1}^{m} \alpha_i = 1.$$

Define y^i as $y^i := x^i - x^m$, $i = 1, \ldots, m-1$. Since $m-1 \geq n+1$, we can note that $y^1, \ldots, y^{m-1}$ are not linearly independent. Then, there exists $\beta_1, \ldots, \beta_{m-1}$ with at least one positive β_i such that

$$0 = \sum_{i=1}^{m-1} \beta_i y^i = \sum_{i=1}^{m-1} \beta_i x^i - \left(\sum_{i=1}^{m-1} \beta_i \right) x^m.$$

Defining $\beta_m := -\sum_{i=1}^{m-1} \beta_i$, we have

$$\sum_{i=1}^{m} \beta_i = 0 \quad \text{and} \quad \sum_{i=1}^{m} \beta_i x^i = \sum_{i=1}^{m-1} \beta_i x^i - \left(\sum_{i=1}^{m-1} \beta_i \right) x^m = 0.$$

Then, for all τ,

$$x = \sum_{i=1}^{m} \alpha_i x_i - \tau \sum_{i=1}^{m} \beta_i x^i = \sum_{i=1}^{m} (\alpha_i - \tau \beta_i) x^i \tag{11}$$

and

$$1 = \sum_{i=1}^{m} \alpha_i - \tau \sum_{i=1}^{m} \beta_i = \sum_{i=1}^{m} (\alpha_i - \tau \beta_i)$$

hold. If

$$\bar{\tau} := \min \left\{ \frac{\alpha_i}{\beta_i} \;\middle|\; \beta_i > 0 \right\},$$

then there exists an index j satifying $\beta_j > 0$ and $\alpha_j - \bar{\tau}\beta_j = 0$. Moreover, when $i \neq j$, we have

$$\alpha_i - \bar{\tau}\beta_j \geq 0.$$

Therefore, from (11), x is a convex combination of $x^1, \ldots, x^{j-1}, x^{j+1}, \ldots, x^m$. The result follows by repeating this process $m = n+1$ times. $\square$

Let $S \subset \mathbb{R}^n$ and define

$$S^k := \left\{ x \in \mathbb{R}^n \;\middle|\; x = \sum_{i=1}^{k} \alpha_i x^i, x^i \in S, \alpha_i \geq 0, \sum_{i=1}^{k} \alpha_i = 1 \right\}.$$

as the set of all convex combinations of k elements in S. Clearly, $S^1 \subset S^2 \subset \cdots \subset S^k \cdots$. Moreover, intuitively, we have that the convex hull of S is equivalent to the union of S^k, that is, $\operatorname{co} S = \bigcup_{k=1}^{\infty} S^k$. The following theorem shows that we do not actually need to take k to infinity.

Theorem 1. (Carathéodory's Theorem) *If $S \subset \mathbb{R}^n$, then*

$$\operatorname{co} S = S^{n+1}.$$

Proof. Let us first prove that $S^{n+1} \subseteq \operatorname{co} S$ by induction. Since $S^1 = S$, $S^1 \subseteq \operatorname{co} S$. Now, assume that $S^k \subseteq \operatorname{co} S$ for all $k \geq 1$. Let $x \in S^{k+1}$. Then, there exist $x^i \in S$ and α_i such that

$$x = \sum_{i=1}^{k} \alpha_i x^i + \alpha_{k+1} x^{k+1}, \alpha_i \geq 0, \sum_{i=1}^{k+1} \alpha_i = 1.$$

If $\alpha_{k+1} = 1$, we have $x \in S$ and thus $x \in \operatorname{co} S$. If $\alpha_{k+1} < 1$, we can write

$$x = (1 - \alpha_{k+1}) \left(\sum_{i=1}^{k} \frac{\alpha_i}{1 - \alpha_{k+1}} x^i \right) + \alpha_{k+1} x^{k+1}.$$

Since $\alpha_i / (1 - \alpha_{k+1}) \geq 0$ and

$$\sum_{i=1}^{k} \frac{\alpha_i}{1 - \alpha_{k+1}} = \frac{1 - \alpha_{k+1}}{1 - \alpha_{k+1}} = 1,$$

we obtain

$$\sum_{i=1}^{k} \frac{\alpha_i}{1 - \alpha_{k+1}} x^i \in S^k \subseteq \operatorname{co} S.$$

Because $x^{k+1} \in \operatorname{co} S$ and co$S$ is convex, we conclude that $x \in \operatorname{co} S$. Therefore, $S^{n+1} \subseteq \operatorname{co} S$ holds. Now, assume that $S \subseteq S^{n+1}$ and let us first prove that S^{n+1} is convex. Observe that $S \subseteq S^{n+1}$. If $x, y \in S^{n+1}$, then there exist $x^i, y^i \in S$, and scalars α_i, β_i such that

$$x = \sum_{i=1}^{n+1} \alpha_i x^i, \alpha_i \geq 0, \sum_{i=1}^{n+1} \alpha_i = 1,$$

$$y = \sum_{i=1}^{n+1} \beta_i y^i, \beta_i \geq 0, \sum_{i=1}^{k+1} \beta_i = 1.$$

Letting $\gamma \in [0,1]$, we also obtain

$$\gamma x + (1-\gamma)y = \sum_{i=1}^{n+1} \gamma \alpha_i x^i + \sum_{i=1}^{n+1} (1-\gamma)\beta_i y^i.$$

Moreover, we know that $\gamma\alpha_i \geq 0$, $(1-\gamma)\beta_i \geq 0$ and

$$\sum_{i=1}^{n+1} \gamma\alpha_i + \sum_{i=1}^{n+1} (1-\gamma)\beta_i = 1.$$

Thus, from Lemma 3, $\gamma x + (1-\gamma)y$ can be written as a convex combination of $n+1$ elements of $\{x^i\}$ and $\{y^i\}$. This means that $\gamma x + (1-\gamma)y \in S^{n+1}$, and so S^{n+1} is a convex set. Since $\operatorname{co} S$ is the smallest convex set containing S, we conclude that $\operatorname{co} S = S^{n+1}$. □

The Carathéodory's theorem is used to prove the following properties concerning cones. This result will be used for proving KKT of nonlinear programming problems.

Proposition 1. *Let C and D be cones in $\mathbb{R}^n$. Then, the following statements hold.*

(a) $C \subseteq D \Rightarrow C^\circ \supseteq D^\circ$;

(b) $C^\circ = (\operatorname{co} C)^\circ$.

Proof. (a) If $y \in D^\circ$, from the definition of polar cone, $\langle y, x\rangle \leq 0$ for all $x \in D$. From assumption, $C \subseteq D$ holds, and so $\langle y, x\rangle \leq 0$ holds for all $x \in C$. This means that $y \in C^\circ$, as it was claimed.

(b) Since $C \subseteq \operatorname{co} C$, from (a), we have $C^\circ \supseteq (\operatorname{co} C)^\circ$. Thus, we just need to show that $C^\circ \subseteq (\operatorname{co} C)^\circ$. Let $y \in C^\circ$ and $x \in \operatorname{co} C$ be taken arbitrarily. From Theorem 1, there exist $\alpha_i \geq 0$ and $x^i \in C$, $i = 1, \ldots, n+1$ such that

$$x = \sum_{i=1}^{n+1} \alpha_i x^i.$$

Therefore, we have

$$\langle y, x\rangle = \sum_{i=1}^{n+1} \alpha_i \langle y, x^i\rangle \leq 0$$

This means that $y \in (\operatorname{co} C)^\circ$, and the conclusion follows. □

1st part by Ellen H. Fukuda (e-mail: ellen(at)i.kyoto-u.ac.jp, where (at) = @)