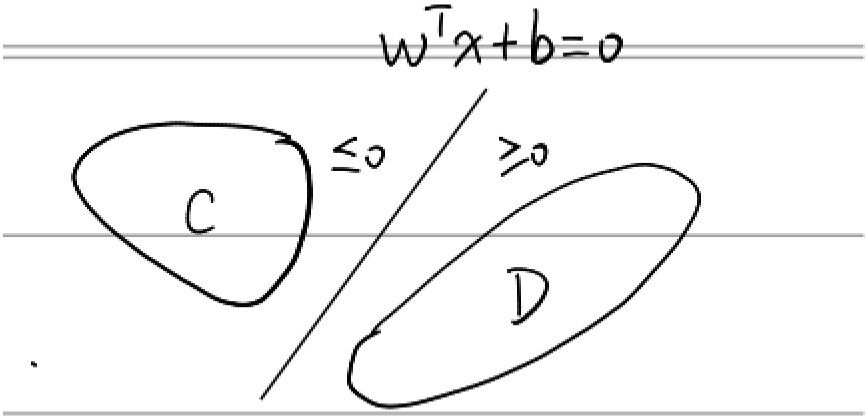
Lecture 4. Supporting and Separating Hyperplane Theorem

We would like to present a fundamental property of convex sets. Roughly speaking, we would like to show that every convex set $C \subseteq \mathbb{R}^d$ can be characterized by its 'supporting hyperplanes', and every two convex sets can be separated by a hyperplane.



4.1 Projection to convex sets

Given a set $C \subseteq \mathbb{R}^n$, the distance between a point $\boldsymbol{x}$ and C is defined by

$$\operatorname{dist}(\boldsymbol{x}, C) = \inf_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\| .$$

We define the *(metric) projection* of $\boldsymbol{x}$ onto C as the closest points in C to $\boldsymbol{x}$.

Definition (*Projection*)

Let $C \subseteq \mathbb{R}^n$ be a nonempty, closed and convex set. Then for any $\boldsymbol{x} \in \mathbb{R}^n$, the *projection* of $\boldsymbol{x}$ onto C is defined as

$$\mathcal{P}_C(\boldsymbol{x}) \triangleq \arg\min_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\| .$$

That is, $\|\boldsymbol{x} - \mathcal{P}_C(\boldsymbol{x})\| = \operatorname{dist}(\boldsymbol{x}, C)$.

Question

Is this well-defined?

Clearly, if $\boldsymbol{x} \in C$ then $\mathcal{P}_C(\boldsymbol{x}) = \boldsymbol{x}$. Now we assume that $\boldsymbol{x} \notin C$.
We first show that the minimizer exists. Since $C \neq \emptyset$, select any $\boldsymbol{z} \in C$ and let $r = \|\boldsymbol{x} - \boldsymbol{z}\|$. Then $\mathcal{B}(\boldsymbol{x}, r) \cap C \neq \emptyset$. Since C is closed, $\mathcal{B}(\boldsymbol{x}, r) \cap C$ is bounded and closed, and thus compact. Note that

$$\inf_{y \in C} \|\boldsymbol{x} - \boldsymbol{y}\| = \inf_{y \in \mathcal{B}(\boldsymbol{x}, r) \cap C} \|\boldsymbol{x} - \boldsymbol{y}\| .$$

By the extreme value theorem, the infinium can be achieved by some $\boldsymbol{y} \in C$.
Next, we show that the minimizer is unique. Suppose there are two points $\boldsymbol{y}_1 \neq \boldsymbol{y}_2 \in C$ such that $\operatorname{dist}(\boldsymbol{x}, C) = \|\boldsymbol{x} - \boldsymbol{y}_1\| = \|\boldsymbol{x} - \boldsymbol{y}_2\|$. Let $\boldsymbol{y}_c = \frac{1}{2}(\boldsymbol{y}_1 + \boldsymbol{y}_2)$.
Since C is convex, $\boldsymbol{y}_c \in C$ and thus $\|\boldsymbol{x} - \boldsymbol{y}_c\| \geq \operatorname{dist}(\boldsymbol{x}, C)$. However, we have

$$\begin{aligned}
\|\boldsymbol{x} - \boldsymbol{y}_c\|^2 = \left\|\boldsymbol{x} - \frac{\boldsymbol{y}_1 + \boldsymbol{y}_2}{2}\right\|^2 &= \left\|\frac{1}{2}(\boldsymbol{x} - \boldsymbol{y}_1) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{y}_2)\right\|^2 \\
&= \frac{1}{4}\left\|(\boldsymbol{x} - \boldsymbol{y}_1) + (\boldsymbol{x} - \boldsymbol{y}_2)\right\|^2 \\
&= \frac{1}{4}\langle(\boldsymbol{x} - \boldsymbol{y}_1) + (\boldsymbol{x} - \boldsymbol{y}_2), (\boldsymbol{x} - \boldsymbol{y}_1) + (\boldsymbol{x} - \boldsymbol{y}_2)\rangle \\
&= \frac{1}{4}\left(\langle \boldsymbol{x} - \boldsymbol{y}_1, \boldsymbol{x} - \boldsymbol{y}_1\rangle + \langle \boldsymbol{x} - \boldsymbol{y}_2, \boldsymbol{x} - \boldsymbol{y}_2\rangle + 2\langle \boldsymbol{x} - \boldsymbol{y}_1, \boldsymbol{x} - \boldsymbol{y}_2\rangle\right) \\
&= \frac{1}{4}\left(\|\boldsymbol{x} - \boldsymbol{y}_1\|^2 + \|\boldsymbol{x} - \boldsymbol{y}_2\|^2 + 2\langle \boldsymbol{x} - \boldsymbol{y}_1, \boldsymbol{x} - \boldsymbol{y}_2\rangle\right),
\end{aligned}$$

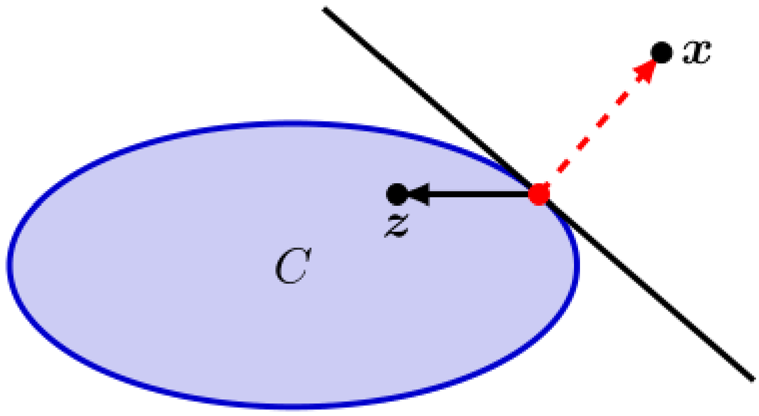
which yields that

$$\begin{aligned}
0 \leq \|\boldsymbol{x} - \boldsymbol{y}_c\|^2 - \operatorname{dist}(\boldsymbol{x}, C)^2 &= \frac{1}{4}\left(-\|\boldsymbol{x} - \boldsymbol{y}_1\|^2 - \|\boldsymbol{x} - \boldsymbol{y}_2\|^2 + 2\langle \boldsymbol{x} - \boldsymbol{y}_1, \boldsymbol{x} - \boldsymbol{y}_2\rangle\right) \\
&= -\frac{1}{4}\langle(\boldsymbol{x} - \boldsymbol{y}_1) - (\boldsymbol{x} - \boldsymbol{y}_2), (\boldsymbol{x} - \boldsymbol{y}_1) - (\boldsymbol{x} - \boldsymbol{y}_2)\rangle \\
&= -\frac{1}{4}\|\boldsymbol{y}_2 - \boldsymbol{y}_1\|^2 < 0 .
\end{aligned}$$

Thus the minimizer is unique, and $\mathcal{P}_C(\boldsymbol{x})$ is well-defined.

Lemma

Let C be a nonempty, closed and convex set. Given $\boldsymbol{x}$ and $\boldsymbol{y} = \mathcal{P}_C(\boldsymbol{x})$, for any $\boldsymbol{z} \in C$, it holds that $\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{z} - \boldsymbol{y}\rangle \leq 0$.



Proof

Note that for all $t \in (0,1)$, $\boldsymbol{y} + t(\boldsymbol{z} - \boldsymbol{y}) \in C$. So

$$\|\boldsymbol{x} - \boldsymbol{y} - t(\boldsymbol{z} - \boldsymbol{y})\|^2 \geq \|\boldsymbol{x} - \boldsymbol{y}\|^2 .$$

Thus, $-2t\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{z} - \boldsymbol{y} \rangle + t^2\|\boldsymbol{z} - \boldsymbol{y}\|^2 \geq 0$ for all $t \in (0,1)$, which concludes that $\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{z} - \boldsymbol{y} \rangle \leq 0$.

Corollary

Let $C \subseteq \mathbb{R}^n$ be a nonempty, closed and convex set. For any $\boldsymbol{x} \notin C$, there exists $\boldsymbol{w} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ such that

$$\sup_{\boldsymbol{z} \in C} \langle \boldsymbol{w}, \boldsymbol{z} \rangle < \langle \boldsymbol{w}, \boldsymbol{x} \rangle .$$

Geometrically, this means C and $\boldsymbol{x} \notin C$ can be strictly separated by a hyperplane. This is a special case of the separating hyperplane theorem we will discuss shortly.

Proof

Let $\boldsymbol{y} = \mathcal{P}_C(\boldsymbol{x})$, and $\boldsymbol{w} = \boldsymbol{x} - \boldsymbol{y}$. Since $\boldsymbol{x} \notin C$, $\boldsymbol{w} \neq \boldsymbol{0}$. Then we have for any $\boldsymbol{z} \in C$,

$$\langle \boldsymbol{w}, \boldsymbol{z} - \boldsymbol{y} \rangle \leq 0 ,$$

which is equivalent to

$$\langle \boldsymbol{w}, \boldsymbol{z} \rangle \leq \langle \boldsymbol{w}, \boldsymbol{y} \rangle = \langle \boldsymbol{w}, \boldsymbol{x} \rangle - \langle \boldsymbol{w}, \boldsymbol{w} \rangle .$$

Taking the supremum over C, it gives that

$$\sup_{\boldsymbol{z}\in C}\,\langle \boldsymbol{w}, \boldsymbol{z}\rangle \le \langle \boldsymbol{w}, \boldsymbol{x}\rangle - \langle \boldsymbol{w}, \boldsymbol{w}\rangle < \langle \boldsymbol{w}, \boldsymbol{x}\rangle\,.$$

In fact, the hyperplane orthogonal to $\boldsymbol{x} - \mathcal{P}_C(\boldsymbol{x})$ separates $\boldsymbol{x}$ and C. We can also generalize this lemma to two convex sets.

Theorem (*Strictly separating hyperplane theorem*)

Let $C, D \subseteq \mathbb{R}^n$ be two disjoint closed convex sets, and at least one of them is bounded. Then there exists $\boldsymbol{w} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ such that

$$\sup_{\boldsymbol{x}\in C}\,\langle \boldsymbol{w}, \boldsymbol{x}\rangle < \inf_{\boldsymbol{y}\in C}\,\langle \boldsymbol{w}, \boldsymbol{y}\rangle\,.$$

Namely, there exists $\boldsymbol{w} \neq \boldsymbol{0}$ and b such that

$$\forall\,\boldsymbol{x} \in C,\quad \boldsymbol{w}^\mathsf{T}\boldsymbol{x} + b < 0 \quad \text{and} \quad \forall\,\boldsymbol{y} \in D,\quad \boldsymbol{w}^\mathsf{T}\boldsymbol{y} + b > 0\,.$$

The idea is to find $\boldsymbol{x} \in C, \boldsymbol{y} \in D$ such that $\|\boldsymbol{x} - \boldsymbol{y}\| = \mathrm{dist}(C, D)$, and show that $\boldsymbol{w} = \boldsymbol{y} - \boldsymbol{x}$ (the hyperplane orthogonal to $\boldsymbol{y} - \boldsymbol{x}$) is a desired one.

4.2 Supporting hyperplane theorem

Definition

- The *interior* of a set C is defined as:

$$\mathrm{int}(C) \triangleq \{x \in C \mid \exists\,\epsilon > 0, \mathcal{B}(x, \epsilon) \subseteq C\}\,.$$

- The *closure* of a set C is defined as

$$\mathrm{cl}(C) \triangleq \{x \in \mathbb{R}^n \mid \exists\, x_1, \cdots, x_n, \cdots \in C,\ \lim_{n\to\infty} x_n = x\}\,.$$

- The *boundary* of a set C is defined as

$$\mathrm{bd}(C) \text{ or } \partial C \triangleq \mathrm{cl}(C) \setminus \mathrm{int}(C)$$

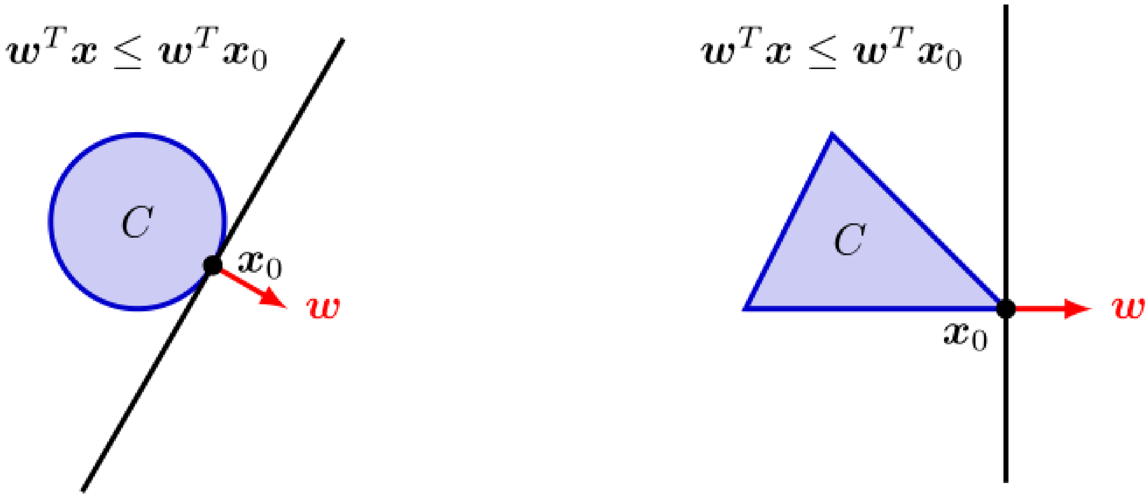
  or equivalently,

$$\partial C \triangleq \{x \in \mathbb{R}^n \mid \forall \epsilon > 0, \mathcal{B}(x,\epsilon) \cap C \neq \emptyset \wedge \mathcal{B}(x,\epsilon) \nsubseteq C\}.$$

Theorem (*Supporting hyperplane theorem*)

Given a nonempty convex set $C \subseteq \mathbb{R}^n$, and a point $\boldsymbol{x}_0 \in \partial C$, there exists $\boldsymbol{w} \neq \boldsymbol{0} \in \mathbb{R}^n$ such that $P = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{w}^\mathsf{T}\boldsymbol{x} = \boldsymbol{w}^\mathsf{T}\boldsymbol{x}_0\}$ is a *supporting hyperplane* of C at $\boldsymbol{x}_0$, namely,

$$\forall \boldsymbol{x} \in C, \qquad \langle \boldsymbol{w}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{w}, \boldsymbol{x}_0 \rangle.$$



Proof

- If $\text{int}(C) = \emptyset$, then C lies in an affine set of dimension less than n. Otherwise, there exists $n+1$ affinely independent points in C, which implies that C contains a n-simplex. However, the interior of the simplex is nonempty, which contradicts $\text{int}(C) = \emptyset$. Now choose any hyperplane that the affine set lies on and we are done.
- If $\text{int}(C) \neq \emptyset$, let $C_\varepsilon \triangleq \{x \mid \mathcal{B}(x,\varepsilon) \subseteq \text{cl}(C)\}$. Note that $x_0 \notin C_\varepsilon$ and C_ε is closed. By the corollary in Section 4.1, for all $\varepsilon < 0$, there exists a hyperplane strictly separates C_ε and x_0, namely, $\exists\, w_\varepsilon \neq 0$ such that $w_\varepsilon^\mathsf{T} x < w_\varepsilon^\mathsf{T} x_0, \forall\, x \in C_\varepsilon$. We normalize w_ε such that $\|w_\varepsilon\| = 1$.
  Next we consider a series of points $\varepsilon_k = \frac{1}{k}, k = 1, 2, \ldots$ For each k, ε_k corresponds to a w_{ε_k}, and $\|w_{\varepsilon_k}\| = 1$. Hence, by the Bolzano–Weierstrass theorem, there exists a convergent subsequence of $\{w_{\varepsilon_k}\}$. Denote by w its limit. Then we show that this w is the coefficient of the desired

hyperplane.
For any $x \in \text{int}(C)$, there exists $N > 0$ such that

$$\forall\, k > N, \qquad w_{\varepsilon_k}^\top x < w_{\varepsilon_k}^\top x_0\,.$$

Thus, $w^\top x \leq w^\top x_0$ by taking the limit on both sides.
For any $y \in \partial C$, there exists a sequence $\{y_k \in \text{int}\, C\}_{k\in\mathbb{N}} \to y$ by convexity. (Why?) Since $w^\top y_k \leq w^\top x_0$ for each k, we can conclude that $w^\top y \leq w^\top x_0$.

Proposition

Let $C \subset \mathbb{R}^n$ be a convex set with nonempty interior, $\boldsymbol{x} \in \partial C$ be a boundary point. Then there exists a sequence $\{\boldsymbol{x}_k \in \text{int}\, C\}_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}$.

Proof

By definition, there exists $\{\boldsymbol{y}_k \in C\}_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} \boldsymbol{y}_k = \boldsymbol{x}$. Since $\text{int}(C) \neq \emptyset$, choose any point $\boldsymbol{z} \in \text{int}(C)$, thus there exists $r > 0$ such that $\mathcal{B}(\boldsymbol{z}, r) \subseteq C$. That is, for any $\boldsymbol{w} \in \mathbb{R}^n$ with $\|\boldsymbol{w}\| \leq r$, $\boldsymbol{z} + \boldsymbol{w} \in C$. So by convexity, for any $k \in \mathbb{N}$ and $t \in (0, 1)$,

$$\boldsymbol{y}_k + t(\boldsymbol{z} + \boldsymbol{w} - \boldsymbol{y}_k) = \boldsymbol{y}_k + t(\boldsymbol{z} - \boldsymbol{y}_k) + t\boldsymbol{w} \in C\,,$$

which implies that $\mathcal{B}(\boldsymbol{y}_k + t(\boldsymbol{z} - \boldsymbol{y}_k), rt) \subseteq C$. Thus, $\boldsymbol{y}_k + t(\boldsymbol{z} - \boldsymbol{y}_k) \in \text{int}(C)$.
Let $\boldsymbol{x}_k = \boldsymbol{y}_k + \frac{1}{k}(\boldsymbol{z} - \boldsymbol{y}_k)$. We have $\boldsymbol{x}_k \in \text{int}(C)$ for all $k \in \mathbb{N}$, and $\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}$.

Corollary of the supporting hyperplane theorem

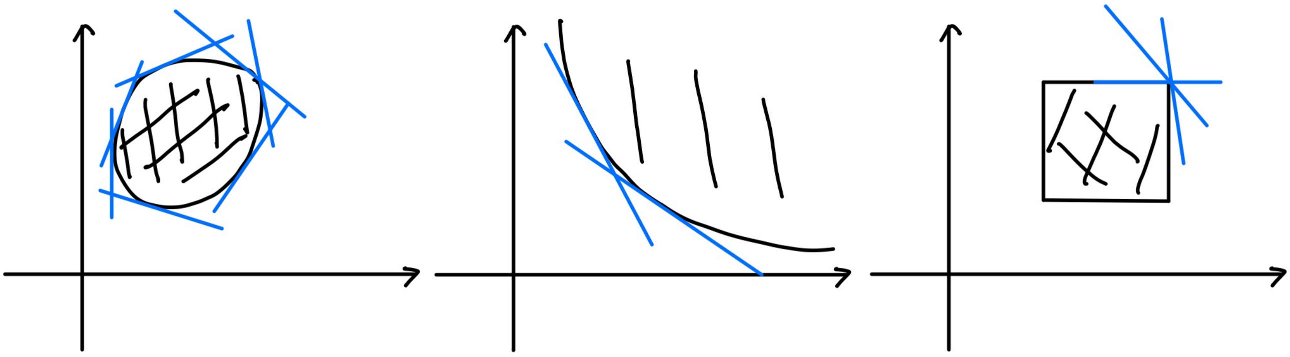
Any nonempty closed convex is the intersection of some halfspaces.

In fact, for any closed convex $C \neq \emptyset$,

$$C = \cap\{H \mid H \text{ is a closed subspace containing } C\} = \cap\{H : H \text{ is a supporting halfspace}\}\,.$$

We will not give the formal proof of this proposition in our lecture. However, let us try to understand this proposition intuitively.
For a 2-dimension convex set. We can find a tangent line at each boundary point,

and the set only lies in a single side of the line. For all of these boundary points, we can get a lot of tangent lines, and an area bounded by these lines. Hence, the proposition tells us this area is just the original convex set.



When considering high dimensional spaces, we can just use the supporting hyperplane theorem. For each boundary point $x_0 \in \partial C$, let

$$P = \{x \mid w^\top x = w^\top x_0\},$$

and make C lie in the halfspace of $w^\top x \leq w^\top x_0$. Thus, this proposition tells us that

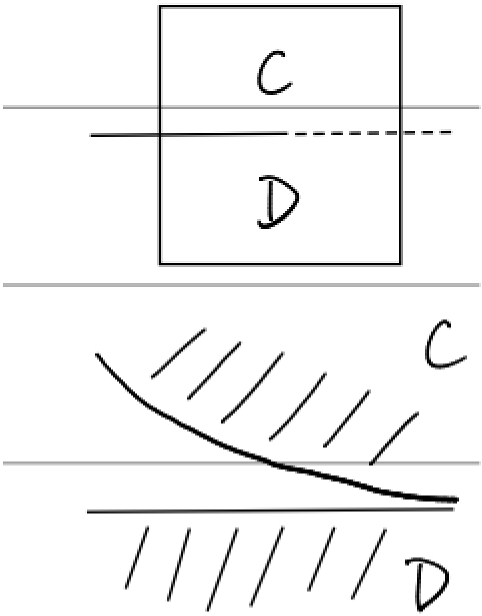
$$C = \bigcap_{x_0 \in \partial C} \{x \mid w^\top x \leq w^\top x_0\}.$$

Remark

Note that the number of those halfspaces may be infinite and even uncountable.

4.3 Separating hyperplane theorem

We would like to show that any two disjoint (not necessarily bounded or closed!) convex sets can be separated by a hyperplane. Note that the hyperplane may not separate these two sets strictly.



Theorem (*Separating hyperplane theorem*)

Let C and D be two disjoint convex sets. Then there exists a hyperplane $\{\boldsymbol{x} \mid \boldsymbol{w}^\top \boldsymbol{x} + b = 0, \boldsymbol{w} \neq \boldsymbol{0}\}$ separating C and D, namely, for all $\boldsymbol{x} \in C$, $\boldsymbol{w}^\top \boldsymbol{x} + b \leq 0$ and for all $\boldsymbol{x} \in D$, $\boldsymbol{w}^\top \boldsymbol{x} + b \geq 0$.

Proof

Consider the set

$$C - D \triangleq \{u - v \mid u \in C, v \in D\}.$$

It suffices to separates $C - D$ and $\{0\}$. This is because, if there exists a hyperplane $w \neq 0$ such that $\forall x \in C - D$, $w^\top x \leq 0$. Then for all $u \in C$ and $v \in D$, we have $w^\top u \leq w^\top v$. Finally, let $b = -\sup_{u \in C} w^\top u$.

- **Case 1:** $0 \notin \partial(C - D)$. By the corollary in Section 4.1, there is a hyperplane separating $\{0\}$ and $\operatorname{cl}(C - D)$.
- **Case 2:** $0 \in \partial(C - D)$. Applying the supporting hyperplane theorem, we can find a supporting hyperplane for $C - D$ at 0, which separates $\{0\}$ and $C - D$.

4.4 Farkas' lemma

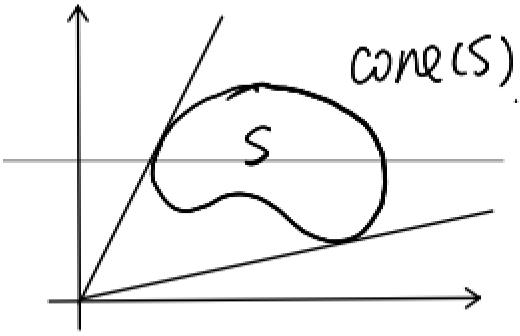
We now present an application of the separating hyperplane theorem. This lemma will help us prove the strong duality in Lecture 8.

Theorem (*Farkas' lemma*)

Let $\boldsymbol{A} \in \mathbb{R}^{n \times m}, \boldsymbol{b} \in \mathbb{R}^n$. Then exactly one of the following sets is empty:

1. $\{\boldsymbol{x} \in \mathbb{R}^m \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}\}$;
2. $\{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{A}^\top \boldsymbol{y} \leq \boldsymbol{0}, \boldsymbol{b}^\top \boldsymbol{y} > 0\}$.

Recall the conic combination and the cone hull.



The two sets can also be understood in the following ways:

1. The first set is non-empty means that $\boldsymbol{b}$ locates in a cone hull of $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$: Let $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)$. If there exists $\boldsymbol{x} \geq \boldsymbol{0}$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, then $\boldsymbol{b} = x_1\boldsymbol{a}_1 + \cdots + x_m\boldsymbol{a}_m \in \text{cone}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)$.
2. The second set is non-empty means that there exists $\boldsymbol{y} \in \mathbb{R}^n$ such that the hyperplane $\{\boldsymbol{x} \mid \boldsymbol{y}^\top \boldsymbol{x} = 0\}$ separates $\boldsymbol{b}$ and the column vectors of $\boldsymbol{A}$. The Farkas' lemma tells us there exists a separating hyperplane passing through $\boldsymbol{0}$ unless $\boldsymbol{b} \in \text{cone}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)$.

Proof

First, we prove that if the first set is nonempty, the second one must be empty. Otherwise, there exist $\boldsymbol{x}$ and $\boldsymbol{y}$ such that:

$$0 < \boldsymbol{b}^\top \boldsymbol{y} = (\boldsymbol{A}\boldsymbol{x})^\top \boldsymbol{y} = \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{y} \leq 0\,.$$

Next, we prove that if the first set is empty, the second one must be nonempty. It is easy to find a hyperplane to separate $\text{cone}(\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\})$ and $\boldsymbol{b}$ by the strictly separting hyperplane theorem in Section 4.1. (Why?) Hence, there exists $\boldsymbol{y}$ and t such that

$$\forall\, \boldsymbol{z} \in \text{cone}(\boldsymbol{A}), \quad \boldsymbol{z}^\top \boldsymbol{y} + t < 0 \quad \text{and} \quad \boldsymbol{b}^\top \boldsymbol{y} + t > 0\,.$$

The key problem is how to make the separating hyperplane pass through the original point. Actually, we can show that the hyperplane $\{\boldsymbol{z} \mid \boldsymbol{y}^\top \boldsymbol{z} = 0\}$ is also a separating hyperplane.

1. For all $\boldsymbol{a}_i$ and $\lambda_i > 0$, $\lambda_i \boldsymbol{a}_i \in \mathrm{cone}(A)$. Then $\lambda_i \boldsymbol{a}_i^\top \boldsymbol{y} + t < 0$, which is equivalent to $\boldsymbol{a}_i^\top \boldsymbol{y} + t/\lambda_i < 0$. Taking the limit as $\lambda_i \to \infty$, it gives $\boldsymbol{a}_i^\top \boldsymbol{y} \leq 0$.
2. In addition, $\mathbf{0} \in \mathrm{cone}(\boldsymbol{A})$ implies that $t < 0$. Thus $\boldsymbol{b}^\top \boldsymbol{y} > 0$.

Therefore, the hypeplane $\{\boldsymbol{z} \mid \boldsymbol{y}^\top \boldsymbol{z} = 0\}$ is a desired hyperplane, and the second set is nonempty in this case.
Overall, exactly one of the two sets must be nonempty whenever the first one is empty or nonempty.

Question

Why is $\mathrm{cone}(\boldsymbol{A})$ closed?