Optimization Methods and Applications

Lecture 2. Separation Theorems, and Optimization Problems

Dr. Huiqi Guan

Fudan University

October 14, 2025

Roadmap

Last lecture

- syllabus
- introduction to optimization problems
- convexity
  - convex sets
  - convex functions
  - convex programming

This lecture

- separation theorems
  - Farkas' Lemma
  - Gordan's Lemma
- unconstrained optimization problems

Separation Theorem

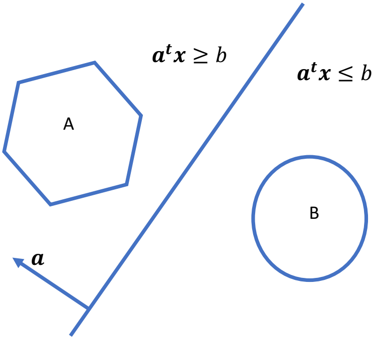

Figure 1: The hyperplane $\mathbf{a}^{\mathbf{t}}\mathbf{x} = b$ separates the disjoint convex sets A and B.

Compact Set

Theorem 2.1

In $\mathbb{R}^n$ a set X is compact if and only if every sequence $\{\mathbf{x}_k\}$ of points in X has a subsequence $\{\mathbf{x}_{k_j}\}$ that converges to a point in X.

Theorem 2.2

In $\mathbb{R}^n$ a set X is compact if and only if X is closed and bounded.

Lemma 2.3

If X is a compact set in $\mathbb{R}^n$ and $f : X \to \mathbb{R}$ is a continuous function, then f attains the minimum value on X, i.e there is a $\mathbf{x}^* \in X$ such that $f(\mathbf{x}^*) = \min_{\mathbf{x}\in X} f(\mathbf{x})$.

Proof.

Denote $f_0 = \inf_{\mathbf{x}\in X} f(\mathbf{x})$. Then there exists a sequence $\{\mathbf{x}_k\}$ of points in X such that $\lim_{k\to\infty} f(\mathbf{x}_k) = f_0$. Since X is a compact set, then the sequence $\{\mathbf{x}_k\}$ has a subsequence $\{\mathbf{x}_{k_j}\}$ that converges to a point $\mathbf{x}^* \in X$. Note that f is continuous on X. Then $f(\mathbf{x}^*) = \lim_{k_j\to\infty} f(\mathbf{x}_{k_j}) = f_0$. Hence, $f(\mathbf{x}^*) = \min_{\mathbf{x}\in X} f(\mathbf{x})$. $\square$

Distance Function

Let $d(\mathbf{y}, \mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|$ for any two vectors $\mathbf{y}, \mathbf{x}$. If X is a set and $\mathbf{y}$ is a vector, then we define the distance $\mathbf{y}$ to X denoted by $d(\mathbf{y}, X)$, to be

$$d(\mathbf{y}, X) = \inf\{\|\mathbf{y} - \mathbf{x}\| : \mathbf{x} \in X\} = \inf_{\mathbf{x} \in X} d(\mathbf{y}, \mathbf{x}).$$

Lemma 2.4

- $d(\mathbf{y}, \mathbf{x}) : \mathbf{x} \in X \mapsto \|\mathbf{y} - \mathbf{x}\|$ *is a continuous function on* X.
- *If* X *is a nonempty closed set, then there exists a point* $\mathbf{x}^* \in X$ *such that* $d(\mathbf{y}, X) = \|\mathbf{y} - \mathbf{x}^*\|$, *which indicates that point* $\mathbf{x}^*$ *in* X *is closest to* $\mathbf{y}$.

Proof.

- $|d(\mathbf{y}, \mathbf{x}_1) - d(\mathbf{y}, \mathbf{x}_2)| = |\|\mathbf{y} - \mathbf{x}_1\| - \|\mathbf{y} - \mathbf{x}_2\|| \leq \|\mathbf{x}_1 - \mathbf{x}_2\|$.
- Let $\mathbf{x}_0 \in X$ and let $r > \|\mathbf{y} - \mathbf{x}_0\|$. Then $X_1 = \overline{B(\mathbf{y}, r)} \cap X$ is nonempty, closed, and bounded and hence is compact. Since $d(\mathbf{y}, \mathbf{x})$ is continuous on X_1, it attains its minimum at some point $\mathbf{x}^*$ in X_1. Thus, for all $\mathbf{x} \in X_1, d(\mathbf{y}, \mathbf{x}) \geq d(\mathbf{y}, \mathbf{x}^*)$. For $\mathbf{x} \in X$ and $\mathbf{x} \notin X_1$, we have $\|\mathbf{y} - \mathbf{x}\| > r > \|\mathbf{y} - \mathbf{x}_0\| \geq \|\mathbf{y} - \mathbf{x}^*\|$, since $\mathbf{x}_0 \in \overline{B(\mathbf{x}, r)}$. Therefore, $\mathbf{x}^*$ is the point in X with minimum distance from $\mathbf{y}$, i.e. $\|\mathbf{y} - \mathbf{x}^*\| = d(\mathbf{y}, X) = \inf\{\|\mathbf{y} - \mathbf{x}\| : \mathbf{x} \in X\}$.

□

Separation Theorem: One Point with a Closed Convex Set

Theorem 2.5

Let C be a nonempty closed convex set in $\mathbb{R}^n$ and $\mathbf{x} \notin C$. Then there exists a nonzero vector $\mathbf{p}$ and a scalar α such that $\mathbf{p}^{\mathbf{t}}\mathbf{x} > \alpha$ and $\mathbf{p}^{\mathbf{t}}\mathbf{y} \leq \alpha$ for each $\mathbf{y} \in C$.

In order to prove the above theorem, we will need the following lemma.

Lemma 2.6

Let C be a nonempty closed convex set in $\mathbb{R}^n$ and $\mathbf{x} \notin C$. Then there exists a **unique** *point $\bar{\mathbf{y}} \in C$ with* **minimum** *distance from $\mathbf{x}$. Furthermore, $\bar{\mathbf{y}}$ is the minimizing point if and only if $(\mathbf{x} - \bar{\mathbf{y}})^{\mathbf{t}}(\mathbf{y} - \bar{\mathbf{y}}) \leq 0$ for all $\mathbf{y} \in C$.*

Note: $(\mathbf{x} - \bar{\mathbf{y}})^{\mathbf{t}}(\mathbf{y} - \bar{\mathbf{y}}) \leq 0 \Rightarrow (\mathbf{x} - \bar{\mathbf{y}})^{\mathbf{t}}\mathbf{y} \leq (\mathbf{x} - \bar{\mathbf{y}})^{\mathbf{t}}\bar{\mathbf{y}} < (\mathbf{x} - \bar{\mathbf{y}})^{\mathbf{t}}\mathbf{x}(\mathbf{x} \neq \bar{\mathbf{y}})$ and let $\mathbf{p} = \mathbf{x} - \bar{\mathbf{y}}$ and $\alpha = (\mathbf{x} - \bar{\mathbf{y}})^{\mathbf{t}}\bar{\mathbf{y}}$.

Separation Theorem: One Point with a Closed Convex Set continued

Proof of Lemma 2.6.

By Lemma 2.4, there exists $\bar{\mathbf{y}} \in C$ with minimum distance from $\mathbf{x}$, i.e. $\|\bar{\mathbf{y}} - \mathbf{x}\| = d(\mathbf{x}, C) = \inf\{\|\mathbf{y} - \mathbf{x}\| : \mathbf{y} \in C\}$.
For $\mathbf{y} \in C$, since C is convex, then $\bar{\mathbf{y}} + t(\mathbf{y} - \bar{\mathbf{y}}) \in C$ with $t \in [0, 1]$. By the definition of $\bar{\mathbf{y}}$, $\|\bar{\mathbf{y}} + t(\mathbf{y} - \bar{\mathbf{y}}) - \mathbf{x}\| \geq \|\bar{\mathbf{y}} - \mathbf{x}\|$. Therefore, $t^2\|\mathbf{y} - \bar{\mathbf{y}}\|^2 + 2t\langle \bar{\mathbf{y}} - \mathbf{x}, \mathbf{y} - \bar{\mathbf{y}} \rangle \geq 0$ for any $t \in [0, 1]$. Hence, $\langle \bar{\mathbf{y}} - \mathbf{x}, \mathbf{y} - \bar{\mathbf{y}} \rangle \geq 0$.
If $\langle \bar{\mathbf{y}} - \mathbf{x}, \mathbf{y} - \bar{\mathbf{y}} \rangle \geq 0$ for all $\mathbf{y}$ in C, then $t^2\|\mathbf{y} - \bar{\mathbf{y}}\|^2 + 2t\langle \bar{\mathbf{y}} - \mathbf{x}, \mathbf{y} - \bar{\mathbf{y}} \rangle \geq 0$ for any $t \in [0, 1]$. Hence, $\|\bar{\mathbf{y}} + t(\mathbf{y} - \bar{\mathbf{y}}) - \mathbf{x}\| \geq \|\bar{\mathbf{y}} - \mathbf{x}\|$. Choosing $t = 1$, we get $\|\mathbf{y} - \mathbf{x}\| \geq \|\bar{\mathbf{y}} - \mathbf{x}\|$. (or
$\|\mathbf{y} - \mathbf{x}\|^2 = \langle \mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x} \rangle = \langle \mathbf{y} - \bar{\mathbf{y}} + \bar{\mathbf{y}} - \mathbf{x}, \mathbf{y} - \bar{\mathbf{y}} + \bar{\mathbf{y}} - \mathbf{x} \rangle =$
$\langle \mathbf{y} - \bar{\mathbf{y}}, \mathbf{y} - \bar{\mathbf{y}} \rangle + 2\langle \bar{\mathbf{y}} - \mathbf{x}, \mathbf{y} - \bar{\mathbf{y}} \rangle + \langle \bar{\mathbf{y}} - \mathbf{x}, \bar{\mathbf{y}} - \mathbf{x} \rangle \geq \langle \bar{\mathbf{y}} - \mathbf{x}, \bar{\mathbf{y}} - \mathbf{x} \rangle = \|\bar{\mathbf{y}} - \mathbf{x}\|^2$)
If $\hat{\mathbf{y}}$ is also the point with minimum distance from $\mathbf{x}$, then $\langle \hat{\mathbf{y}} - \mathbf{x}, \bar{\mathbf{y}} - \hat{\mathbf{y}} \rangle \geq 0$.
Note $\langle \bar{\mathbf{y}} - \mathbf{x}, \hat{\mathbf{y}} - \bar{\mathbf{y}} \rangle \geq 0$. Add them together, and we get
$-\|\bar{\mathbf{y}} - \hat{\mathbf{y}}\|^2 = \langle \hat{\mathbf{y}} - \bar{\mathbf{y}}, \bar{\mathbf{y}} - \hat{\mathbf{y}} \rangle \geq 0$, which yields $\hat{\mathbf{y}} = \bar{\mathbf{y}}$. □

Separation Theorem: One Point with a Closed Convex Set continued

Corollary 2.7

Let C be a nonempty closed convex set in $\mathbb{R}^n$ and $C \neq \mathbb{R}^n$. Then C is the intersection of all half spaces containing C.

Proof.

Let $\tilde{C}$ be the intersection of all half spaces containing C. Then $\tilde{C}$ contains C. Therefore, $\tilde{C}$ is a nonempty closed convex set in $\mathbb{R}^n$. If $\tilde{C} \neq C$, then there is a point $\mathbf{x}$ such that $\mathbf{x} \in \tilde{C}$ but $\mathbf{x} \notin C$. By Theorem 2.5, we know there exists a nonzero vector $\mathbf{p}$ and a scalar α such that $\mathbf{p^t x} > \alpha$ and $\mathbf{p^t y} \leq \alpha$ for each $\mathbf{y} \in C$. Then the half space $\mathbf{p^t y} \leq \alpha$ contains C. By the definition of $\tilde{C}$, we know the half space $\mathbf{p^t y} \leq \alpha$ contains $\tilde{C}$. Since $\mathbf{x} \in \tilde{C}$, then $\mathbf{p^t x} \leq \alpha$, which contradicts $\mathbf{p^t x} > \alpha$. Hence, we know $\tilde{C} = C$. □

Alternative Theorem: Farkas' Lemma

Lemma 2.8 (Farkas' Lemma)

Let $\mathbf{A}$ be an $m \times n$ matrix and $\mathbf{c}$ be an n-dimensional vector. Then exactly one of the following two systems has a solution:

$System\ 1$	$\mathbf{A}\mathbf{x} \leq \mathbf{0}$ *and* $\mathbf{c}^{\mathbf{t}}\mathbf{x} > 0$	*for some* $\mathbf{x} \in \mathbb{R}^n$
$System\ 2$	$\mathbf{A}^{\mathbf{t}}\mathbf{y} = \mathbf{c}$ *and* $\mathbf{y} \geq \mathbf{0}$	*for some* $\mathbf{y} \in \mathbb{R}^m$

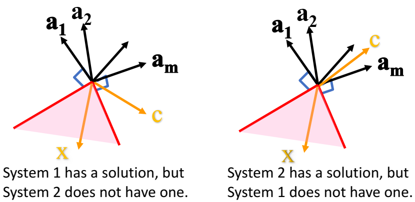

Figure 2: Illustration of Farkas' Lemma, where $\mathbf{A}^{\mathbf{t}} = (\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a_m})$.

Alternative Theorem: Farkas' Lemma continued

Let A be an $m \times n$ matrix and let

$$C = \{\mathbf{w} : \mathbf{w} = A\mathbf{x}, \mathbf{w} \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \geq \mathbf{0}\}.$$

Lemma 2.9

If the columns of A are linearly independent, then C is a closed convex cone.

Lemma 2.10

If the columns of A are not linearly independent, then any point $\mathbf{z}$ in C can be expressed as a nonnegative linear combination of $p < n$ linearly independent columns of A.

Lemma 2.11

C is a closed convex cone.

Alternative Theorem: Farkas' Lemma continued

Proof of Lemma 2.9.

A straightforward calculation verifies that C is a convex cone. Next, we show C is closed.

Let $\mathbf{w}_0$ be a limit point of C, and let $\{\mathbf{w}_k\}$, be a sequence of points in C converging to $\mathbf{w}_0$. Then there exists a sequence of points $\{\mathbf{x}_k\}$, in $\mathbb{R}^n$ with $\mathbf{x}_k \geq \mathbf{0}$ such that $A\mathbf{x}_k \to \mathbf{w}_0$. We show that $\{\mathbf{x}_k\}$ is bounded. If $\{\mathbf{x}_k\}$, were unbounded, there would exist a subsequence, again denoted by $\{\mathbf{x}_k\}$, such that $\|\mathbf{x}_k\| \to \infty$. All points in the sequence $\{\mathbf{x}_k/\|\mathbf{x}_k\|\}$, have norm 1, so there exists a subsequence, again denoted by $\{\mathbf{x}_k\}$, and a point $\mathbf{x}^*$ of norm 1 such that $\mathbf{x}_k/\|\mathbf{x}_k\| \to \mathbf{x}^*$. From $\|\mathbf{x}_k\| A(\mathbf{x}_k/\|\mathbf{x}_k\|) \to \mathbf{w}_0$, $\|\mathbf{x}_k\| \to \infty$, we see that we must have $A(\mathbf{x}_k/\|\mathbf{x}_k\|) \to \mathbf{0}$. On the other hand, $\mathbf{x}_k/\|\mathbf{x}_k\| \to \mathbf{x}^*$ implies that $A(\mathbf{x}_k/\|\mathbf{x}_k\|) \to A\mathbf{x}^*$. Hence, $A\mathbf{x}^* = \mathbf{0}$. Recall that $\|\mathbf{x}^*\| = 1$, so that $\mathbf{x}^* \neq \mathbf{0}$. Therefore the relation $A\mathbf{x}^* = \mathbf{0}$ contradicts the linear independence of the columns of A.

Since $\{\mathbf{x}_k\}$ is bounded, $\{\mathbf{x}_k\}$ has a convergent subsequence, which we denote again by $\{\mathbf{x}_k\}$. Let $\mathbf{x}_0 = \lim_{k\to\infty} \mathbf{x}_k$. Then $\mathbf{x}_0 \geq \mathbf{0}$. Also, $A\mathbf{x}_k \to A\mathbf{x}_0$, so that $A\mathbf{x}_0 = \mathbf{w}_0$. Thus $\mathbf{w}_0 \in C$, and C is closed. $\square$

Alternative Theorem: Farkas' Lemma continued

Proof of Lemma 2.10.

Let $A_j, j = 1, \ldots, n$, denote the jth column of A. Then there exists an $\mathbf{x} \geq \mathbf{0}$ such that $\mathbf{z} = \sum_{j=1}^{n} x_j A_j, x_j \geq 0$. Since the columns of A are linearly dependent, there exists a $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n) \neq \mathbf{0}$ such that $\sum_{j=1}^{n} \mu_j A_j = \mathbf{0}$. Therefore, for any $\rho \in \mathbb{R}$,

$$\mathbf{z} = \sum_{j=1}^{n} (x_j - \rho\mu_j) A_j = \sum_{\mu_j \neq 0} (x_j - \rho\mu_j) A_j + \sum_{\mu_j = 0} x_j A_j.$$

Since $\boldsymbol{\mu} \neq \mathbf{0}$, the first sum on the right exists. If $\mu_j < 0$, then $x_j - \rho\mu_j \geq 0$ whenever$\rho \geq x_j/\mu_j$. In particular, since $x_j \geq 0$, we have $x_j - \rho\mu_j \geq 0$ whenever $\rho \geq 0$. If $\mu_j > 0$, then $x_j/\mu_j \geq 0$ and $x_j - \rho\mu_j \geq 0$ if and only if $\rho \leq x_j/\mu_j$. If there exist indices j such that $\mu_j > 0$, set $\bar{\rho} = \min\{\frac{x_j}{\mu_j} : \mu_j > 0\}$. If $\mu_j < 0$ for all indices j for which $\mu_j \neq 0$, set $\bar{\rho} = \max\{\frac{x_j}{\mu_j} : \mu_j \neq 0\}$. Then $x_j - \rho\mu_j \geq 0$ for all $j = 1, \ldots, n$ and $x_j - \rho\mu_j = 0$ for at least one value of j. We have now expressed $\mathbf{z}$ as a nonnegative linear combination of $q < n$ columns of A. We continue the process until we express $\mathbf{z}$ as a nonnegative linear combination of $p < n$ linearly independent columns of A. □

Alternative Theorem: Farkas' Lemma continued

Proof of Lemma 2.11.

Let σ denote a choice of $p < n$ linearly independent columns of A and let $A(\sigma)$ denote the corresponding $m \times p$ matrix. There are a finite number of such choices. Let

$$C(\sigma) = \{\mathbf{w} : \mathbf{w} = A(\sigma)\mathbf{y}, \mathbf{w} \in \mathbb{R}^m, \mathbf{y} \in \mathbb{R}^p, \mathbf{y} \geq \mathbf{0}\}.$$

By Lemma 2.9 and Lemma 2.10, each set $C(\sigma)$ is closed and $C \subseteq \cup_\sigma C(\sigma)$. We now show the opposite inclusion. Let $\mathbf{w} \in C(\sigma)$ for some σ. By relabeling columns of A, we can assume without loss of generality that selects the first p columns of A. Then there exists a $\mathbf{y} \in \mathbb{R}^p, \mathbf{y} \geq \mathbf{0}$ such that

$$\mathbf{w} = \sum_{j=1}^{p} y_j A_j = \sum_{j=1}^{p} y_j A_j + \sum_{j=p+1}^{n} 0 * A_j, y_j \geq 0, j = 1, \ldots, p.$$

Let $\mathbf{x} = (y_1, \ldots, y_p, 0, \ldots, 0)$. Then $\mathbf{w} = A\mathbf{x}$, and so $\mathbf{x} \in C$. We have expressed C as the union of a finite number of closed sets. Hence C is closed and Lemma 2.11 is proved. $\square$

Alternative Theorem: Farkas' Lemma continued

Proof of Lemma 2.8(Farkas's Lemma).

If System 1 has a solution, then System 2 has no solution. If System 2 also had a solution, there would exist $\mathbf{x}$ and $\mathbf{y}$ such that

$$A\mathbf{x} \leq \mathbf{0}, \mathbf{c}^{\mathbf{t}}\mathbf{x} > 0, \mathbf{A}^{\mathbf{t}}\mathbf{y} = \mathbf{c} \text{ and } \mathbf{y} \geq \mathbf{0}.$$

Then $0 < \mathbf{c}^{\mathbf{t}}\mathbf{x} = \mathbf{y}^{\mathbf{t}}\mathbf{A}\mathbf{x} \leq 0$, which is a contradiction.
If System 1 has no solution, we will show System 2 has a solution. Let

$$C = \{\mathbf{w} : \mathbf{w} = \mathbf{A}^{\mathbf{t}}\mathbf{y}, \mathbf{w} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m, \mathbf{y} \geq \mathbf{0}\}.$$

If System 2 had no solution, then $\mathbf{c} \notin C$. By Lemma 2.11, C is a nonempty closed convex cone. From Lemma 2.5, there exists a nonzero vector $\mathbf{p}$ and a scalar α such that $\mathbf{c}^{\mathbf{t}}\mathbf{p} > \alpha$ and $\mathbf{p}^{\mathbf{t}}\mathbf{w} \leq \alpha$ for each $\mathbf{w} \in C$. Since $\mathbf{0} \in C$, then $\alpha \geq 0$. By the definition of C, then $(\mathbf{A}\mathbf{p})^{\mathbf{t}}\mathbf{y} \leq \alpha, \forall \mathbf{y} \geq \mathbf{0}$, which yields $\mathbf{A}\mathbf{p} \leq \mathbf{0}$. Therefore, we find a solution for System 1, which is a contradiction. □

Alternative Theorem: Farkas' Lemma continued

Corollary 2.12

Let $\mathbf{A}$ be an $m \times n$ matrix and $\mathbf{c}$ be an n-dimensional vector. Then exactly one of the following two systems has a solution:

$$\begin{array}{lll} System\ 1 & \mathbf{Ax} \leq \mathbf{0}, \mathbf{x} \geq \mathbf{0}, \mathbf{c}^{\mathbf{t}}\mathbf{x} > 0 & \textit{for some } \mathbf{x} \in \mathbb{R}^n \\ System\ 2 & \mathbf{A}^{\mathbf{t}}\mathbf{y} \geq \mathbf{c}, \mathbf{y} \geq \mathbf{0} & \textit{for some } \mathbf{y} \in \mathbb{R}^m \end{array}$$

Hint: The result follows by replacing $\mathbf{A}^{\mathbf{t}}$ in Farkas' Lemma by $[\mathbf{A}^{\mathbf{t}}, -\mathbf{I}]$.

Alternative theorem: Gordan's Lemma

Lemma 2.13 (Gordan's Lemma)

Let $\mathbf{A}$ *be an* $m \times n$ *matrix and* $\mathbf{c}$ *be an n-dimensional vector. Then exactly one of the following two systems has a solution:*

$$\begin{array}{lll} System\ 1 & \mathbf{A}\mathbf{x} < \mathbf{0} & \text{for some } \mathbf{x} \in \mathbb{R}^n \\ System\ 2 & \mathbf{A}^{\mathbf{t}}\mathbf{y} = \mathbf{0} \text{ and } \mathbf{y} \geq \mathbf{0} & \text{for some } \mathbf{y} \neq \mathbf{0} \in \mathbb{R}^m. \end{array}$$

Proof.

The $System\ 1$ has a solution if and only if the following system

$$System\ 1' \quad \mathbf{A}\mathbf{x} + \zeta\mathbf{e} \leq \mathbf{0}, \qquad \zeta > 0,$$

has a solution. The result follows by replacing $\mathbf{A}, \mathbf{c}^{\mathbf{t}}, \mathbf{x}$ in Farkas' Lemma by $[\mathbf{A}, \mathbf{e}], (\mathbf{0}^t, 1), (\mathbf{x}^t, \zeta)$. □

Separation Theorem: Two Disjoint Convex Sets

Theorem 2.14

Let S be a nonempty convex set in $\mathbb{R}^n$ and suppose $\bar{\mathbf{x}} \notin S$. Then there exists a hyperplane that separates S and $\bar{\mathbf{x}}$; that is, there exists a nonzero vector $\mathbf{p}$ in $\mathbb{R}^n$ such that

$$\mathbf{p}^{\mathbf{t}}(\mathbf{x} - \bar{\mathbf{x}}) \leq 0$$

for each $\mathbf{x} \in \bar{S}$.

Lemma 2.15

Let S be a nonempty convex set in $\mathbb{R}^n$ and suppose $\bar{\mathbf{x}} \notin S$. For any finite set $A = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} \subseteq S$, there exists a nonzero vector $\mathbf{p}$ in $\mathbb{R}^n$ such that $\mathbf{p}^{\mathbf{t}}(\mathbf{x} - \bar{\mathbf{x}}) < 0$ for each $\mathbf{x} \in$ co$(A) \subseteq S$.

Proof.

Since co$(A) = \{\sum_{i=1}^m \lambda_i \mathbf{x}_i | \mathbf{x}_i \in A, \lambda_i \geq 0, i = 1, 2, \ldots, m, \sum_{i=1}^m \lambda_i = 1, m \in \mathbb{N}\}$ and A is a finite set, then co(A) is a closed set. From Theorem 2.5, we can derive the result. □

Separation Theorem: Two Disjoint Convex Sets continued

Definition (Compact Set)

Let S be a set in $\mathbb{R}^n$. A collection of open sets $\{O_\alpha\}_{\alpha\in A}$ is said to be an open cover of S if every $\mathbf{x} \in S$ is contained in some O_α. A set S in $\mathbb{R}^n$ is said to be compact if for every open cover $\{O_\alpha\}_{\alpha\in A}$ of S there is a finite collection of sets $O_1, \ldots, O_m$ from the original collection $\{O_\alpha\}_{\alpha\in A}$ such that the finite collection $O_1, \ldots, O_m$ is also an open cover of S.

Theorem 2.16 (Finite-Intersection Property)

Let S be a compact set in $\mathbb{R}^n$. Suppose $\{C_\alpha\}_{\alpha\in A}$ is a collection of closed sets in S. If for any finite set $\{\alpha_1, \ldots, \alpha_k\}, \cap_{i=1}^k C_{\alpha_i} \neq \phi$, then $\cap_{\alpha\in A} C_\alpha \neq \phi$.

Proof of Theorem 2.14.

If $\cap_{\alpha\in A} C_\alpha = \phi$, then $S \subseteq \cup_{\alpha\in A}(\mathbb{R}^n / \ C_\alpha)$. Since C_α is closed, then $\mathbb{R}^n / \ C_\alpha$ is open. So $\{\mathbb{R}^n / \ C_\alpha\}_{\alpha\in A}$ is an open cover of S. Since S is compact, then there exists a finite collection of sets $C_{\alpha_1}, \ldots, C_{\alpha_m}$ such that $S \subseteq \cup_{i=1}^m (\mathbb{R}^n / \ C_{\alpha_i}) = \mathbb{R}^n / \ (\cap_{i=1}^k C_{\alpha_i})$. Since $\cap_{i=1}^k C_{\alpha_i} \subseteq S$, then $\cap_{i=1}^k C_{\alpha_i} = \phi$ and a contradiction. $\square$

Separation Theorem: Two Disjoint Convex Sets continued

Proof of Theorem 2.14.

For each $\mathbf{x}$ in S define

$$N_{\mathbf{x}} = \{\mathbf{z} : \|\mathbf{z}\| = 1, \mathbf{z}^{\mathbf{t}}(\mathbf{x} - \bar{\mathbf{x}}) \le 0.\}$$

The set $N_{\mathbf{x}}$ is nonempty since it contains the element $-(\mathbf{x} - \bar{\mathbf{x}})/\|\mathbf{x} - \bar{\mathbf{x}}\|$. To derive the theorem, it suffices to show that

$$\cap_{\mathbf{x} \in C} N_{\mathbf{x}} \neq \phi.$$

Note that $N_{\mathbf{x}}$ is closed and is contained in compact set $S(\mathbf{0}, 1) = \{\mathbf{u} : \|\mathbf{u}\| = 1\}$. From Lemma 2.15, for any set $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$, we have $\cap_{i=1}^{k} N_{\mathbf{x}_i} \neq \phi$. Therefore, we derive the result by the Theorem 2.16.

□

Separation Theorem: Two Disjoint Convex Sets continued

From Theorem 2.14, we can derive the separation theorem for two disjoint convex sets as follows.

Theorem 2.17

Let S_1 and S_2 be nonempty convex sets in $\mathbb{R}^n$ and suppose $S_1 \cap S_2 = \phi$. Then there exists a hyperplane that separates S_1 and S_2; that is, there exists a nonzero vector $\mathbf{p}$ in $\mathbb{R}^n$ such that

$$\inf\{\mathbf{p}^{\mathbf{t}}\mathbf{x} : \mathbf{x} \in \bar{S}_1\} \geq \sup\{\mathbf{p}^{\mathbf{t}}\mathbf{x} : \mathbf{x} \in \bar{S}_2\}.$$

Proof.

Let $S_0 = S_2 - S_1$. Then S_0 is convex. Since $S_1 \cap S_2 = \phi$, then $\mathbf{0} \notin S_0$. By the theorem 2.14, there exists a nonzero vector $\mathbf{p} \in \mathbb{R}^n$ such that $\mathbf{p}^{\mathbf{t}}(\mathbf{s} - \mathbf{0}) \leq 0, \forall \mathbf{s} \in S_0$, i.e. $\mathbf{p}^{\mathbf{t}}\mathbf{x} \leq \mathbf{p}^{\mathbf{t}}\mathbf{y}, \forall \mathbf{x} \in S_2, \mathbf{y} \in S_1$. □

Separation Theorem: Supporting Hyperplane

Corollary 2.18

Let S be a nonempty convex set in $\mathbb{R}^n$ and suppose $\bar{\mathbf{x}} \in \partial S$. Then there exists a hyperplane that supports S at $\bar{\mathbf{x}}$; that is, there exists a nonzero vector $\mathbf{p}$ in $\mathbb{R}^n$ such that

$$\mathbf{p}^{\mathbf{t}}(\mathbf{x} - \bar{\mathbf{x}}) \leq 0$$

for each $\mathbf{x} \in \bar{S}$.

Proof.

Since $\bar{\mathbf{x}} \in \partial S$, based on the definition of ∂S, we know their exists a sequence $\{\mathbf{x}_n\}_{n=1}^{\infty} \not\subseteq S$ and $\lim_{n\to\infty} \mathbf{x}_n = \bar{\mathbf{x}}$. For each $\mathbf{x}_n \notin S$, we know there exists $\bar{\mathbf{p}}_n$ such that $||\bar{\mathbf{p}}_n|| = 1$ and

$$\mathbf{p}_n^{\mathbf{t}}(\mathbf{x} - \mathbf{x}_n) \leq 0$$

for each $\mathbf{x} \in \bar{S}$. Without loss of generality, suppose $\lim_{n\to\infty} \mathbf{p}_n = \mathbf{p}$. Let n go to ∞ and then we have the result. □

2. Separation Theorems, and Optimization Problems
   - Separation Theorem
   - Optimization Problems
   - Unconstrained Optimization Problems

Optimization Problems

- Linear Optimization Problem:

$$\text{minimize } \mathbf{c}^{\mathbf{t}}\mathbf{x}$$
$$\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

- (Nonlinear) Optimization Problem:

$$\text{minimize } f(\mathbf{x})$$
$$\text{subject to } \mathbf{x} \in X$$

Here, X is a set in $\mathbb{R}^n$ and f is a real-valued function on $\mathbb{R}^n$. X is called the *feasible set* and elements of X are called *feasible points* or *feasible vectors*.

Definition

Definition 2.19

- Optimization Problem:

$$\text{minimize } f(\mathbf{x})$$
$$\text{subject to } \mathbf{x} \in X$$

  Here, X is a set in $\mathbb{R}^n$ and f is a real-valued function on $\mathbb{R}^n$. X is called the *feasible set* and elements of X are called *feasible points* or *feasible vectors*.

- **(Global) Minimizer**: We say that a point $\mathbf{x}_0$ in X is a (global) minimizer or that f has a (global) minimum at $\mathbf{x}_0$ if $f(\mathbf{x}_0) \leq f(\mathbf{x})$ for all $\mathbf{x}$ in X.
- **(Strict) Local Minimizer**: We say that a point $\mathbf{x}_0$ in S is a (strict) local minimizer or that f has a (strict) local minimum at $\mathbf{x}_0$ if there exists a $\delta > 0$ such that $f(\mathbf{x}_0) \leq (<) f(\mathbf{x})$ for all $\mathbf{x} (\neq \mathbf{x}_0)$ in $B(\mathbf{x}_0, \delta) \cap X$, where $B(\mathbf{x}_0, \delta) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_0\| < \delta\}$.
- Unconstrained Problem: X is an open set in $\mathbb{R}^n$
- Constrained Problem: $X = \{\mathbf{x} : \mathbf{x} \in X_0, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$, where $f : X_0 \to \mathbb{R}, \mathbf{g} : X_0 \to \mathbb{R}^m, \mathbf{h} : X_0 \to \mathbb{R}^k$.

Descent Direction

Definition 2.20 (Descent Direction)

Let f be differentiable at $\mathbf{x}$. If $\nabla f(\mathbf{x})\mathbf{d} < 0$, $\mathbf{d}$ is denoted as a descent direction of f at $\mathbf{x}$.

Lemma 2.21

Let X be an open set in $\mathbb{R}^n$ and let f be differentiable at $\mathbf{x}$. If there is a vector $\mathbf{d}$ such that $\nabla f(\mathbf{x})\mathbf{d} < 0$, then there exists a $\delta > 0$ such that $f(\mathbf{x} + \lambda\mathbf{d}) < f(\mathbf{x})$ for all $\lambda \in (0, \delta)$, so that $\mathbf{d}$ is a descent direction of f at $\mathbf{x}$.

Proof.

By differentiability of f at $\mathbf{x}$, we must have

$$f(\mathbf{x} + \lambda\mathbf{d}) = f(\mathbf{x}) + \lambda\nabla f(\mathbf{x})\mathbf{d} + \lambda\|\mathbf{d}\|\alpha(\mathbf{x}; \lambda\mathbf{d})$$

where $\alpha(\mathbf{x}; \lambda\mathbf{d}) \to 0$ as $\lambda \to 0$. You can readily derive the lemma. $\square$

If $\nabla f(\mathbf{x}) \neq 0$, then $-\nabla f(\mathbf{x})$ is the direction of steepest descent at $\mathbf{x}$. Denote by $\mathcal{D}_{\mathbf{x}}$ the cone of descent directions at $\mathbf{x}$ or

$$\mathcal{D}_{\mathbf{x}} = \{\mathbf{d} \in \mathbb{R}^n : \nabla f(\mathbf{x})\mathbf{d} < 0\}.$$

Feasible Direction

Definition 2.22 (Feasible Direction)

Let X be a set in $\mathbb{R}^n$ and $\mathbf{x} \in X$. Nonzero vector $\mathbf{d}$ is denoted as a feasible direction of X at $\mathbf{x}$ if there is a δ and a differentiable curve $\boldsymbol{\xi}(t)$ such that

$$\boldsymbol{\xi}(0) = \mathbf{x}, \boldsymbol{\xi}'(0) = \mathbf{d}, \boldsymbol{\xi}(t) \in X, t \in (0, \delta).$$

At feasible point $\mathbf{x}$, denote the cone of feasible directions at $\mathbf{x}$ as $\mathcal{F}_{\mathbf{x}}$. If there is a $\gamma > 0$ such that $\mathbf{x} + t\mathbf{d} \in X$ for all $t \in (0, \gamma)$, then $\mathbf{d}$ is a feasible direction of X at $\mathbf{x}$. Therefore, $\{\mathbf{d} \in \mathbb{R}^n : \mathbf{d} \neq \mathbf{0}, \mathbf{x} + t\mathbf{d} \in X$ for all $t \in (0, \gamma)$ for some $\gamma > 0\} \subseteq \mathcal{F}_{\mathbf{x}}$.

Example 2.23

Let $X = \{\mathbf{x} : A\mathbf{x} = \mathbf{0}\}$. Then $\mathcal{F}_{\mathbf{x}} = \{\mathbf{d} : \mathbf{d} \neq \mathbf{0}, A\mathbf{d} = \mathbf{0}.\}$.

Example 2.24

Let $X = \{\mathbf{x} : \mathbf{x} \in X_0, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$. Let $\mathbf{x}^* \in X$ and $E = \{i : g_i(\mathbf{x}^*) = 0\}$. If $\nabla \mathbf{g}_E(\mathbf{x}^*)\mathbf{d} < \mathbf{0}$, then $\mathbf{d}$ is a feasible direction of X at $\mathbf{x}^*$.

Hint: $\mathbf{x}^* + t\mathbf{d} \in X$ with small enough t.

Example for Feasible Direction

Consider

$$\text{minimize } x_1 + x_2$$
$$\text{subject to } g(\mathbf{x}) = 1 - x_1^2 - x_2^2 \geq 0$$

For the interior point $\mathbf{x}^1, \mathcal{F}_{\mathbf{x}^1} = \mathbb{R}^2$.
For the boundary point $\mathbf{x}^2 = (-1, 0), \mathcal{F}_{\mathbf{x}^2} = \{\mathbf{d} \in \mathbb{R}^2 | d_1 \geq 0, \mathbf{d} \neq \mathbf{0}\}$.

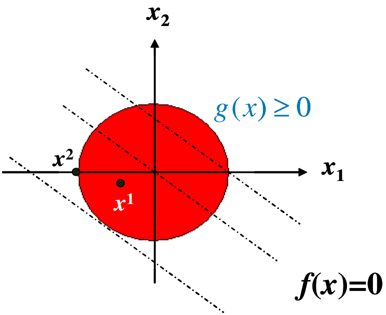

Figure 3: Feasible Direction.

Geometric Optimality Conditions

A fundamental question in Optimization is: what are the necessary conditions in order to have $\bar{\mathbf{x}}$ as a local optimizer?

A general answer is: the intersection of the descent and feasible direction sets at $\bar{\mathbf{x}}$ must be empty. That is, $\mathcal{D}_{\bar{\mathbf{x}}} \cap \mathcal{F}_{\bar{\mathbf{x}}} = \phi$ can be regarded as a geometric condition for $\bar{\mathbf{x}}$ to be a local minimizer. It is a necessary condition.

Unconstrained Problems–Univariate Function

Theorem 2.25 (Necessary Condition)

Let $\mathbf{X}$ be an open interval in $\mathbb{R}$, let f be a real-valued function defined on $\mathbb{R}$, and let x_0 in $\mathbf{X}$ be a local minimizer for f. If f is differentiable at x_0, then $f'(x_0) = 0$. If f is of class $C^{(2)}$ on some interval about x_0, then $f''(x_0) \geq 0$.

Hint: By Taylor's theorem,
$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(\bar{x})(x - x_0)^2$, where $\bar{x}$ lies between x_0 and x.

Definition 2.26

Critical Point: A point c in X such that $f'(c) = 0$ is called a critical point of f.

Unconstrained Problems–Univariate Function

Theorem 2.27 (Sufficient Condition)

Let $\mathbf{X}$ be an open interval in $\mathbb{R}$ and let f be a real-valued function defined on $\mathbb{R}$ and of class $C^{(2)}$ on some interval about x_0 in X. If $f'(x_0) = 0$ and $f''(x_0) > 0$, then x_0 is a strict local minimizer for f.

Proof.

Since f is of class $C^{(2)}$ on some interval about x_0 in X and $f''(x_0) > 0$, then there is a $\delta > 0$ such that $f''(x) > 0$ if $x \in B(x_0, \delta)$. For $x \in B(x_0, \delta)$ and $x \neq x_0$, by Taylor's theorem, $f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(\bar{x})(x - x_0)^2 > f(x_0)$, where $\bar{x}$ lies between x_0 and x. □

Corollary 2.28

Let $f'(x_0) = 0$. If $\underline{f''(x) \geq 0 \text{ on } X}$, then x_0 is a minimizer. If $f''(x) > 0$ on X, then x_0 is a strict minimizer. If $f''(x) \geq 0$ on some interval about x_0, then x_0 is a local minimizer.

Unconstrained Problems–Multivariate Function

Notation: f is a real-valued function and $\mathbf{g}^t = (g_1, \dots, g_m)$ is a vector-valued function that is differentiable at $\mathbf{x}_0$, then

$$\nabla f(\mathbf{x}_0) = (\frac{\partial f}{\partial x_1}(\mathbf{x}_0), \frac{\partial f}{\partial x_2}(\mathbf{x}_0), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x}_0))$$

and

$$\nabla \mathbf{g}(\mathbf{x}_0) = \begin{pmatrix} \nabla g_1(\mathbf{x}_0) \\ \nabla g_2(\mathbf{x}_0) \\ \dots \\ \nabla g_m(\mathbf{x}_0) \end{pmatrix} = (\frac{\partial g_i(\mathbf{x}_0)}{\partial x_j})$$

Unconstrained Problems–Multivariate Function

Theorem 2.29 (Necessary Condition)

Let X be an open set in $\mathbb{R}^n$, let f be a real-valued function defined on X, and let $\mathbf{x}_0$ be a local minimizer. If f has first-order partial derivatives at $\mathbf{x}_0$, then

$$\frac{\partial f}{\partial x_i}(\mathbf{x}_0) = 0, \quad i = 1, \ldots, n.$$

If f is of class $C^{(2)}$ on some open ball $B(\mathbf{x}_0, \delta)$ centered at $\mathbf{x}_0$, then

$$H(\mathbf{x}_0) = \left(\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}_0)\right)$$

is positive semidefinite.

Hint: Consider $\varphi(t; \mathbf{d}) = f(\mathbf{x}_0 + t\mathbf{d})$. Then $\varphi'(0; \mathbf{d}) = \nabla f(\mathbf{x}_0)\mathbf{d}$ and $\varphi''(0; \mathbf{d}) = \langle \mathbf{d}, H(\mathbf{x}_0)\mathbf{d} \rangle$.

Definition 2.30

Critical Point: A point $\mathbf{c}$ in X such that $\nabla f(\mathbf{c}) = \mathbf{0}$ is called a critical point of f.

Unconstrained Problems–Multivariate Function

Theorem 2.31 (Sufficient Condition)

Let X be an open set in $\mathbb{R}^n$ and let f be a real-valued function of class $C^{(2)}$ on some ball $B(\mathbf{x}_0, \delta_1)$. If $H(\mathbf{x}_0)$ is positive definite and $\nabla f(\mathbf{x}_0) = \mathbf{0}$, then $\mathbf{x}_0$ is a strict local minimizer. If X is convex, f is of class $C^{(2)}$ on X, and $H(\mathbf{x})$ is positive semidefinite for all $\mathbf{x}$ in X, then $\mathbf{x}_0$ is a minimizer for f. If $H(\mathbf{x})$ is positive definite for all $\mathbf{x}$ in X, then $\mathbf{x}_0$ is a strict minimizer.

Example

Example 2.32

Let

$$f(x,y) = x^4 + y^4 - 32y^2.$$

Minimize f on $\mathbb{R}^2$.

Solution 2.33

Since $\nabla f(x,y) = (4x^3, 4y^3 - 64y)$, *then the critical points are* $(0,0), (0,-4), (0,4)$. *Note the Hessian matrix* $H(x,y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & 12y^2 - 64 \end{pmatrix}$. *At no critical point, the Hessian matrix is positive definite. But it is positive semidefinite when* (x,y) *is close enough to* $(0,-4)$ *or* $(0,4)$. *Therefore,* $(0,-4)$ *and* $(0,4)$ *are the local minimizers of* f. *Note* $f(x,y) = x^4 + (y^2 - 16)^2 - 16^2 \geq -16^2$, *which equals the value of* f *at* $(0,-4)$ *or* $(0,4)$. *Then* $(0,-4)$ *and* $(0,4)$ *are also the global minimizers of* f.

Optimization of Convex Functions

Theorem 2.34

Let f be a convex function defined on a convex set C.

- *If f attains a local minimum at $\mathbf{x}_0$, then f attains a minimum at $\mathbf{x}_0$.*
- *The set of points at which f attains a minimum is either empty or convex.*
- *If f is strictly convex and f attains a minimum at $\mathbf{x}^*$, then $\mathbf{x}^*$ is unique.*
- *If f is not a constant function and if f attains a maximum at some point $\mathbf{x}$ in C, then $\mathbf{x}$ must be a boundary point of C.*

Proof.

Since f is not a constant, there exists $\mathbf{y} \in C$ such that $f(\mathbf{y}) < f(\mathbf{x})$. If $\mathbf{x}$ is not a boundary point of C, there exists a δ such that $B(\mathbf{x}, \delta) \subseteq C$. Then $\mathbf{z} = \mathbf{x} + \frac{\delta}{2}(\mathbf{x} - \mathbf{y}) \in B(\mathbf{x}, \delta)$ and $\mathbf{x} = \frac{1}{1+\frac{\delta}{2}}\mathbf{z} + \frac{\frac{\delta}{2}}{1+\frac{\delta}{2}}\mathbf{y}$. Since f is convex, then we have the contradiction $f(\mathbf{x}) \le \frac{1}{1+\frac{\delta}{2}} f(\mathbf{z}) + \frac{\frac{\delta}{2}}{1+\frac{\delta}{2}} f(\mathbf{y}) < \frac{1}{1+\frac{\delta}{2}} f(\mathbf{x}) + \frac{\frac{\delta}{2}}{1+\frac{\delta}{2}} f(\mathbf{x}) = f(\mathbf{x})$. □

Optimization of Convex Functions

Theorem 2.35

Let f be convex and differentiable on an open convex set C. Then f attains a minimum at $\mathbf{x}^*$ in C if and only if $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Theorem 2.36

Let f be of class $C^{(2)}$ on an open convex set D. Then f is convex on D if and only if the Hessian matrix

$$H(\mathbf{x}) = \left(\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}_0)\right)$$

is positive semidefinite at each point $\mathbf{x}$ in D. If $H(\mathbf{x})$ is positive definite at each $\mathbf{x}$, then f is strictly convex.

Note: The sufficient condition for the unconstrained problems is equivalent to say that f is convex near the local minimizer.

Example

Example 2.37

Show that $f(x, y) = x^2 + y^2 - 3x - 7y$ attains a minimum on $\mathbb{R}^2$ at a unique point and find the minimum.

Solution 2.38

Since $\nabla f(x, y) = (2x - 3, 2y - 7)$, *then the critical point is* $(\frac{3}{2}, \frac{7}{2})$. *Note the Hessian matrix* $H(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, *which is positive definite for all* (x, y). *Therefore,* $f(x, y)$ *is a convex function on* $\mathbb{R}^2$. *Then* $(\frac{3}{2}, \frac{7}{2})$ *is the global minimizers of* f *and the minimum is* $f(\frac{3}{2}, \frac{7}{2}) = -\frac{29}{2}$.