Mathematical Notations and Analyses

LI-PING ZHANG
Department of Mathematical Sciences
Tsinghua University, Beijing 100084
Office: New Science Building #A302, Tel: 62798531
E-mail: lipingzhang@tsinghua.edu.cn

Outline

- Euclidean Space
- Convex Set, Separating Hyperplane Theorems
- Convex Function
- Farkas' Lemma

Euclidean Space

- $\mathcal{R}$: real numbers; $\mathcal{R}^n$: n-dimensional Euclidean space
- Column vector: $\mathbf{x} = (x_1; x_2; \ldots; x_n)$
- Row vector: $\mathbf{x} = (x_1, x_2, \ldots, x_n)$
- Transpose operation: ^{T}
- $\mathbf{x} \geq \mathbf{y}$ means $x_j \geq y_j$ for $j = 1, 2, ..., n$
- $\mathbf{0}$: vector of all zeros; $\mathbf{e}$: vector of all ones

Metric of an Euclidean Space

- Inner-product of two vectors:

$$\mathbf{x} \bullet \mathbf{y} := \mathbf{x}^T \mathbf{y} = \sum_{j=1}^{n} x_j y_j$$

- Euclidean norm: $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{x}}$,

  Infinity-norm: $\|\mathbf{x}\|_\infty = \max\{|x_1|, |x_2|, ..., |x_n|\}$,

  p-norm: $\|\mathbf{x}\|_p = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}$

Basis of an Euclidean Space

- A set of vectors $\mathbf{a}_1, ..., \mathbf{a}_m$ is said to be linearly dependent if there are scalars $\lambda_1, ..., \lambda_m$, not all zero, such that the linear combination

$$\sum_{i=1}^{m} \lambda_i \mathbf{a}_i = \mathbf{0}$$

- A linearly independent set of vectors that span R^n is a basis.

Plane and Half-Spaces and Polyhedron

$$H = \{\mathbf{x} : \ \mathbf{ax} = \sum_{j=1}^{n} a_j x_j = b\}$$

$$H^+ = \{\mathbf{x} : \ \mathbf{ax} = \sum_{j=1}^{n} a_j x_j \leq b\}$$

$$H^- = \{\mathbf{x} : \ \mathbf{ax} = \sum_{j=1}^{n} a_j x_j \geq b\}$$

Polyhedron: intersection of finite many closed halfspaces 多面体

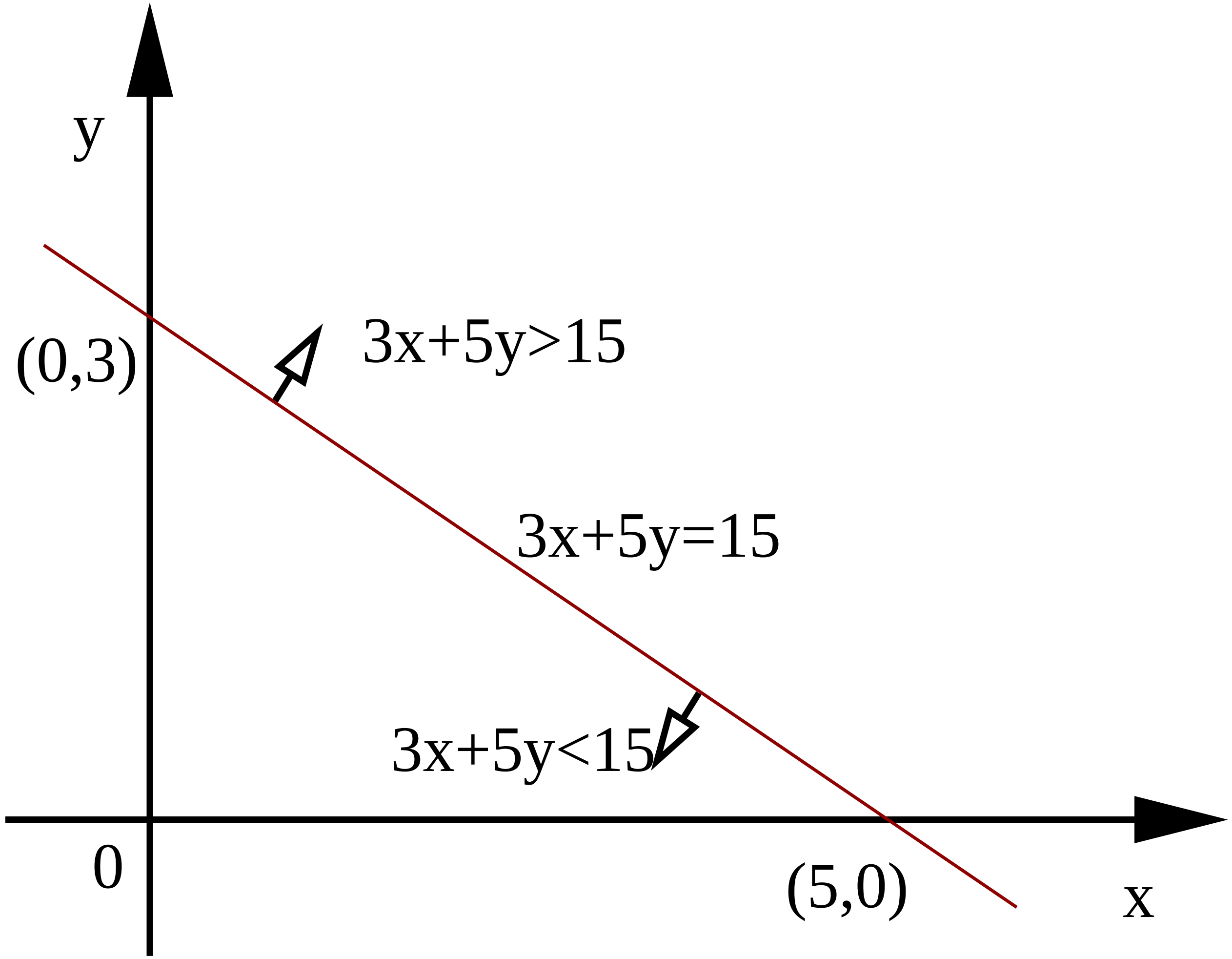

Figure 1: Plane and Half-Spaces and Polyhedron

Matrices

- Matrix: $A \in \mathcal{R}^{m\times n}$, ith row: $A_{i.}$, jth column: $A_{.j}$, ijth element: a_{ij}
- A_I denotes the submatrix of A whose rows belong to index set I, A_J denotes the submatrix whose columns belong to index set J, A_{IJ} denotes the submatrix whose rows belong to index set I and columns belong to index set J.
- All-zero matrix: $\mathbf{0}$, and identity matrix: I

Matrices

- Diagonal matrix: $X = \textbf{diag}(\mathbf{x})$
- Symmetric matrix: $Q = Q^T$
- Positive Definite: $Q \succ 0$ iff $\mathbf{x}^T Q \mathbf{x} > 0, \quad \textbf{for all} \quad \mathbf{x} \neq \mathbf{0}$
- Positive Semidefinite: $Q \succeq 0$ iff $\mathbf{x}^T Q \mathbf{x} \geq 0, \quad \textbf{for all} \quad \mathbf{x}$

Line and Convex Combination

When $\mathbf{x}$ **and** $\mathbf{y}$ are two distinct points in R^n and α runs over R ,

$$\{\mathbf{z} : \mathbf{z} = \alpha\mathbf{x} + (1 - \alpha)\mathbf{y}\}$$

is the line determined by $\mathbf{x}$ **and** $\mathbf{y}$.

When $0 \leq \alpha \leq 1$, it is called the convex combination of $\mathbf{x}$ **and** $\mathbf{y}$ and it is the line segment between $\mathbf{x}$ **and** $\mathbf{y}$.

Face of a Polyhedron

Let P be a polyhedron in $\mathcal{R}^n$. F is a face of P if and only if there is a vector $\mathbf{b}$ for which F is the set of points attaining $\max\{\mathbf{b}^T\mathbf{y} : \mathbf{y} \in P\}$ provided this maximum is finite.

That is, if $P \subset \{\mathbf{x} : \mathbf{px} \leq \beta\}$, then $F = \{\mathbf{x} \in P : \mathbf{px} = \beta\}$.

Remark: A polyhedron has only finite many faces; each face is a non-empty polyhedron.

Extreme Point of a Polyhedron

- P: a polyhedron in $\mathcal{R}^n$
- A vector $\mathbf{y} \in P$ is an extreme point or a vertex of P if

  $\mathbf{y}$ is not a convex combination of more than one distinct points.

顶点，不能表示为p中任意其他两点的凸组合

Convex Set

- Ω is said to be a convex set if for every $\mathbf{x}^1, \mathbf{x}^2 \in \Omega$ and every real number $\alpha \in [0,1]$, the point $\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2 \in \Omega$.
- The convex hull of a set Ω is the intersection of all convex sets containing Ω **凸包**
- Intersection of convex sets is convex

♣ Let $K \subset \mathcal{R}^n$ be a convex set. Then, $a \in K$ is an extreme point iff $K \backslash \{a\}$ is convex.

任意n个点的凸包可以写为至多n+1个点的凸组合表示

Proof of Convex Set

- All solutions to the system of linear equations, $\{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}\}$, form a convex set.
- All solutions to the system of linear inequalities, $\{\mathbf{x} : \ A\mathbf{x} \leq \mathbf{b}\}$, form a convex set (polyhedron).

Proof of Convex Set

- All solutions to the system of linear equations and inequalities, $\{\mathbf{x}: \ A\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}\}$, form a convex set.
- Every polyhedron is a closed convex set.

Convex Cones

- A set C is a cone if $\mathbf{x} \in C$ implies $\alpha\mathbf{x} \in C$ for all $\alpha > 0$
- A convex cone is cone plus convex-set.

Dual Cone and Polar

- Dual cone:

$$C^* := \{\mathbf{y} : \ \mathbf{y} \bullet \mathbf{x} \geq 0 \quad \textbf{for all} \quad \mathbf{x} \in C\} \quad 对偶锥$$

- Self-Dual cone: $C^* = C$

- The polar of C:

$$C^P := \{\mathbf{y} : \ \mathbf{y} \bullet \mathbf{x} \leq 0 \quad \textbf{for all} \quad \mathbf{x} \in C\}$$

Cone Examples

- The n-dimensional nonnegative orthant, $\mathcal{R}^n_+ = \{\mathbf{x} \in \mathcal{R}^n : \mathbf{x} \geq 0\}$, is a convex cone.
- The set $\{(t; \mathbf{x}) \in \mathcal{R}^{n+1} : t \geq \|\mathbf{x}\|\}$ is a convex cone in $\mathcal{R}^{n+1}$, called the second-order cone.

二阶锥都是自对偶的

♣ Show that second-order cone is a self-dual cone.

Polyhedral Convex Cones

- A cone C is (convex) polyhedral if C can be represented by

$$C = \{\mathbf{x} : A\mathbf{x} \leq 0\}$$

for some matrix A.

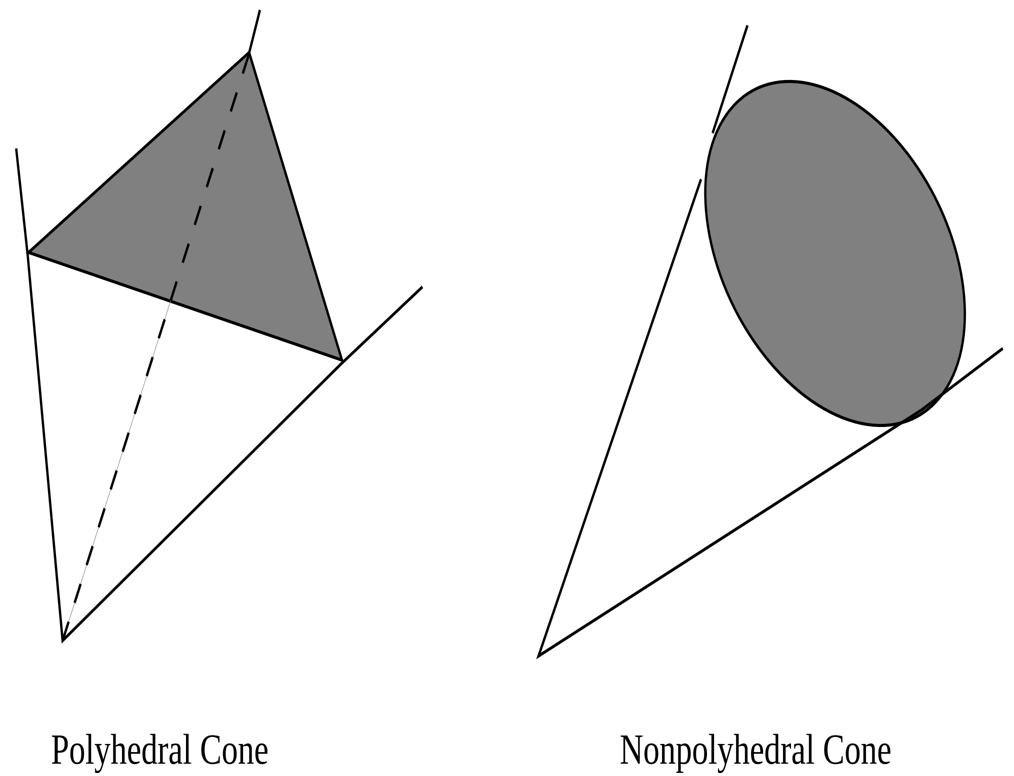

Figure 2: Polyhedral and non-polyhedral cones.

- The nonnegative orthant is a polyhedral cone but the second-order cone is not polyhedral.

Separating Hyperplane

- **Definition** Let $H = \{\mathbf{x} : \mathbf{p}^T\mathbf{x} = \alpha\}$ be a hyperplane in $\mathcal{R}^n$ (hence $\mathbf{p} \neq \mathbf{0}$). Let S_1 and S_2 be two nonempty subsets of $\mathcal{R}^n$. Then H separates S_1 and S_2 if $\mathbf{p}^T\mathbf{x} \geq \alpha$ for all $\mathbf{x} \in S_1$ and $\mathbf{p}^T\mathbf{x} \leq \alpha$ for all $\mathbf{x} \in S_2$.

- Proper separation requires that $S_1 \cup S_2 \not\subset H$ (there exists $\mathbf{x} \in S_1 \cup S_2$ such that $\mathbf{p}^T\mathbf{x} \neq \alpha$). 真分离

Separating Hyperplane

- The sets S_1 and S_2 are strictly separated by H if $\mathbf{p}^T\mathbf{x} > \alpha$ for all $\mathbf{x} \in S_1$ and $\mathbf{p}^T\mathbf{x} < \alpha$ for all $\mathbf{x} \in S_2$. 严格分离
- The sets S_1 and S_2 are strongly separated by H if there exists a number $\varepsilon > 0$ such that $\mathbf{p}^T\mathbf{x} > \alpha + \varepsilon$ for all $\mathbf{x} \in S_1$ and $\mathbf{p}^T\mathbf{x} < \alpha - \varepsilon$ for all $\mathbf{x} \in S_2$. **强分离**

Examples in $\mathcal{R}^2$

Let $H = \{x \in \mathcal{R}^2 : x_2 = 0\}$.

1. $X = [0, 2] \times \{0\}$ and $Y = [1, 3] \times \{0\}$ are separated by H but are not properly separated by H.

2. $X = [0, 2] \times \{0\}$ and $Y = [1, 3] \times [0, 1]$ are properly separated by H but are not strictly separated by H.

3. $X = [0, 2] \times [-1, 0)$ and $Y = [1, 3] \times (0, 1]$ are strictly separated by H but are not strongly separated by H.

4. $X = [0, 2] \times [-1, -\varepsilon)$ and $Y = [1, 3] \times (\varepsilon, 1]$ are strongly separated by H.

Separating Hyperplane Theorem

The most important theorem about the convex set is the following separating hyperplane theorem (Figure 3).

Theorem 1 *(Separating hyperplane theorem) Let $C \subset \mathcal{R}^n$ be a closed convex set, and let $\mathbf{b} \notin C$. Then there is a vector $\mathbf{a} \neq \mathbf{0}$ such that*

$$\mathbf{a} \bullet \mathbf{b} > \sup_{\mathbf{x} \in C} \mathbf{a} \bullet \mathbf{x}$$

where $\mathbf{a}$ is the norm direction of the hyperplane.

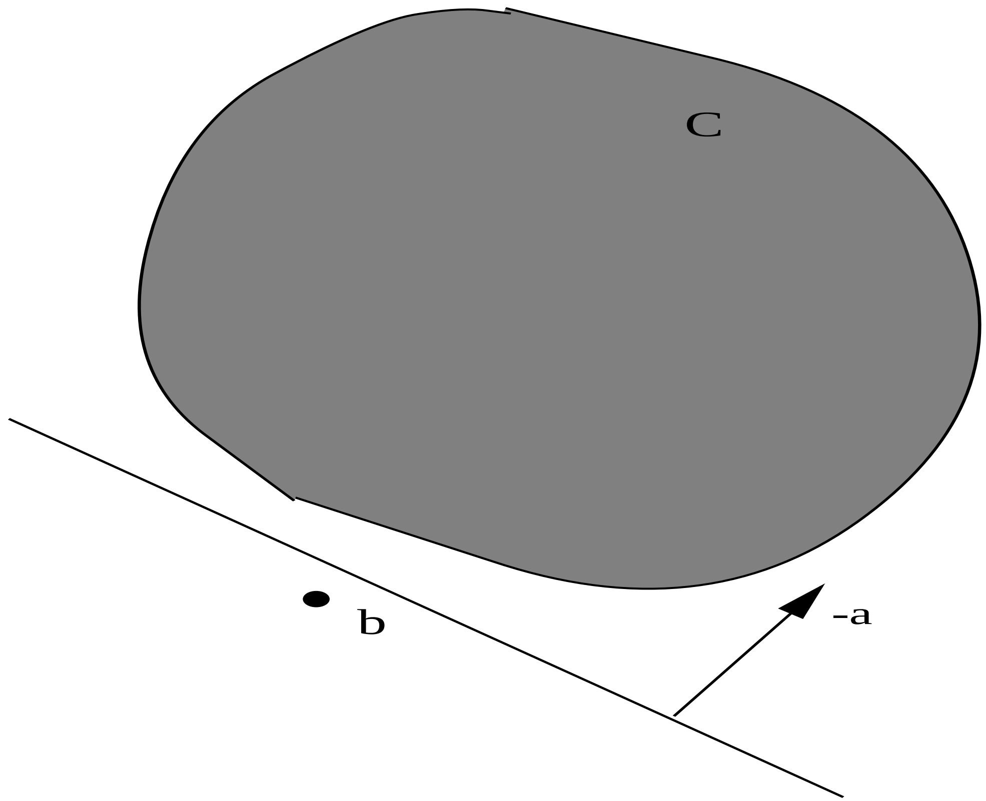

Figure 3: Illustration of the separating hyperplane theorem; an exterior point $\mathbf{b}$ is separated by a hyperplane from a convex set C.

Proof of Separating Hyperplane Theorem

- We first show that there exists a unique point $\bar{\mathbf{x}} \in C$ such that

$$\|\mathbf{b} - \bar{\mathbf{x}}\| = \inf_{\mathbf{x} \in C} \|\mathbf{b} - \mathbf{x}\|.$$

  Furthermore, $\bar{\mathbf{x}}$ is the point of C closest to $\mathbf{b}$ if and only if

$$(\mathbf{b} - \bar{\mathbf{x}})^T (\mathbf{x} - \bar{\mathbf{x}}) \leq 0$$

  for all $\mathbf{x} \in C$.

Proof of Separating Hyperplane Theorem

Let $\gamma = \inf_{\mathbf{x}\in C} \|\mathbf{b}-\mathbf{x}\| > 0.$ Then, there exists a sequence $\{\mathbf{x}_k\} \in C$ such that $\|\mathbf{b}-\mathbf{x}_k\| \to \gamma$.

Since

$$\begin{aligned}
\|\mathbf{x}_k - \mathbf{x}_m\|^2 &= 2\|\mathbf{x}_k - \mathbf{b}\|^2 + 2\|\mathbf{x}_m - \mathbf{b}\|^2 - \|\mathbf{x}_k + \mathbf{x}_m - 2\mathbf{b}\|^2 \\
&= 2\|\mathbf{x}_k - \mathbf{b}\|^2 + 2\|\mathbf{x}_m - \mathbf{b}\|^2 - 4\|\frac{\mathbf{x}_k + \mathbf{x}_m}{2} - \mathbf{b}\|^2 \\
&\leq 2\|\mathbf{x}_k - \mathbf{b}\|^2 + 2\|\mathbf{x}_m - \mathbf{b}\|^2 - 4\gamma^2 \to 0,
\end{aligned}$$

thus there exists a point $\bar{\mathbf{x}} \in C$ such that $\gamma = \|\mathbf{b}-\bar{\mathbf{x}}\|$.

Since C is convex, $\bar{\mathbf{x}}$ is unique.

Proof of Separating Hyperplane Theorem

Suppose that $\mathbf{x} \in C,\ (\mathbf{b} - \bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}}) \leq 0$. Since

$$\begin{aligned}
\|\mathbf{b} - \mathbf{x}\|^2 &= \|\mathbf{b} - \bar{\mathbf{x}} + \bar{\mathbf{x}} - \mathbf{x}\|^2 \\
&= \|\mathbf{b} - \bar{\mathbf{x}}\|^2 + \|\mathbf{x} - \bar{\mathbf{x}}\|^2 - 2(\mathbf{b} - \bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}}) \\
&\geq \|\mathbf{b} - \bar{\mathbf{x}}\|^2,
\end{aligned}$$

then $\bar{\mathbf{x}}$ is the point of C closest to $\mathbf{b}$.

Conversely, it follows from $\|\mathbf{b}-\bar{\mathbf{x}}\| = \inf_{\mathbf{x}\in C}\|\mathbf{b}-\mathbf{x}\|$ that for sufficiently small number $\lambda > 0$,

$$\|\mathbf{b}-\bar{\mathbf{x}}-\lambda(\mathbf{x}-\bar{\mathbf{x}})\|^2 \geq \|\mathbf{b}-\bar{\mathbf{x}}\|^2.$$

Thus,

$$\begin{aligned} 0 &\leq \|\mathbf{b}-\bar{\mathbf{x}}-\lambda(\mathbf{x}-\bar{\mathbf{x}})\|^2 - \|\mathbf{b}-\bar{\mathbf{x}}\|^2 \\ &= \lambda^2\|\mathbf{x}-\bar{\mathbf{x}}\|^2 - 2\lambda(\mathbf{b}-\bar{\mathbf{x}})^T(\mathbf{x}-\bar{\mathbf{x}}). \end{aligned}$$

Hence, $(\mathbf{b}-\bar{\mathbf{x}})^T(\mathbf{x}-\bar{\mathbf{x}}) \leq 0$ for all $\mathbf{x} \in C$.

Proof of Separating Hyperplane Theorem

- We next show that there exists a vector $\mathbf{a} \neq \mathbf{0}$ such that

$$\mathbf{a} \bullet \mathbf{b} > \sup_{\mathbf{x} \in C} \mathbf{a} \bullet \mathbf{x}.$$

There is a point $\bar{\mathbf{x}} \in C$ such that

$$(\mathbf{b} - \bar{\mathbf{x}})^T (\mathbf{x} - \bar{\mathbf{x}}) \leq 0, \quad \forall \mathbf{x} \in C.$$

Let $\mathbf{a} = \mathbf{b} - \bar{\mathbf{x}}$. Then $\mathbf{a} \neq \mathbf{0}$ and for all $\mathbf{x} \in C$,

$$\begin{aligned} \mathbf{a}^T (\mathbf{b} - \mathbf{x}) &= \mathbf{a}^T (\mathbf{b} - \bar{\mathbf{x}}) + \mathbf{a}^T (\bar{\mathbf{x}} - \mathbf{x}) \\ &\geq \|\mathbf{a}\|^2 > 0. \end{aligned}$$

♣ Let C be a nonempty cone in $\mathcal{R}^n$. Then the polar of its polar iff C is closed and convex.

Proof: The polar of C is $C^P = \{y : y^T x \leq 0, \forall x \in C\}$. To show that $C = (C^P)^P$ if C is closed and convex.

Take any $x \in C$, we have

$$y^T x \leq 0 \quad \forall y \in C^P,$$

which implies $x \in (C^P)^P$. So, $C \subseteq (C^P)^P$.

Let $b \in (C^P)^P, b \notin C$. Since C is closed and convex, by the separating hyperplane theorem, there exists $\bar{x} \in C$ such that

$$(b - \bar{x})^T (x - \bar{x}) \leq 0 \quad \forall x \in C.$$

Let $a = b - \bar{x}$. We have

$$b^T a > a^T \bar{x}, \quad a^T x \leq a^T \bar{x} \ \forall x \in C.$$

Since C is a cone, $a^T \bar{x} = 0$. Hence,

$$b^T a > 0, \quad a^T x \leq 0 \ \forall x \in C.$$

The second inequality implies $a \in C^P$. Since $b \in (C^P)^P$, we have $b^T a \leq 0$. This contradicts with the first inequality. So, $(C^P)^P \subseteq C$.

Interior and Boundary

- Let $B_\delta(\bar{\mathbf{x}})$ denote the open ball of radius δ centered at the point $\bar{\mathbf{x}}$, i.e.,

$$B_\delta(\bar{\mathbf{x}}) = \{\mathbf{x} : \ \|\mathbf{x} - \bar{\mathbf{x}}\| < \delta\}.$$

- The **interior** of a set S is the set

$$int(S) = \{\mathbf{x} \in S : \ B_\delta(\mathbf{x}) \subset S \ \textbf{ for some } \ \delta > 0\}.$$

- A point $\bar{\mathbf{x}}$ belongs to the **boundary** ∂S of S if every open ball centered at $\bar{\mathbf{x}}$ meets both S and its complement.

Theorem 2 *Let* $C \subset \mathcal{R}^n$ *be a convex set. Then both* $int(C)$ *and* $cl(C)$ *are convex.*

Proof. We first prove for any $\lambda \in (0, 1)$

$$\mathbf{x} \in int(C), \mathbf{y} \in cl(C) \quad \Rightarrow \quad \mathbf{u} = \lambda \mathbf{x} + (1 - \lambda)\mathbf{y} \in int(C).$$

Since $\mathbf{x} \in int(C)$, there exists $\delta > 0$ such that $B_\delta(\mathbf{x}) \subset C$.

Let

$$\mathcal{Y} = \frac{1}{1-\lambda}(\mathbf{u} - \lambda B_\delta(\mathbf{x})).$$

Then $\mathcal{Y} \cap C \neq \emptyset$. Take

$$\mathbf{v} = \frac{1}{1-\lambda}(\mathbf{u} - \lambda \bar{\mathbf{x}}) \in \mathcal{Y} \cap C$$

where $\bar{\mathbf{x}} \in B_\delta(\mathbf{x})$. Thus, $\mathbf{u} = (1-\lambda)\mathbf{v} + \lambda \bar{\mathbf{x}} \in C$. Let

$$\mathcal{U} = \lambda B_\delta(\mathbf{x}) + (1-\lambda)\mathbf{v}.$$

$\mathcal{U} \subset C$ is a neighborhood of $\mathbf{u}$. Hence $\mathbf{u} \in int(C)$. Clearly, $int(C)$ is convex.

Other Separating Hyperplane Theorems

Theorem 3 *Let $S \in \mathcal{R}^n$ be a nonempty convex set and $\bar{\mathbf{x}} \in \partial S$. Then there exists a vector $\mathbf{a} \neq \mathbf{0}$ such that $\mathbf{a}^T\mathbf{x} \leq \mathbf{a}^T\bar{\mathbf{x}}$ for each $\mathbf{x} \in cl(S)$.*

Proof: Since $\bar{\mathbf{x}} \in \partial S$, there exists a sequence $\{\mathbf{y}^k\}$ not in $cl(S)$ such that $\mathbf{y}^k \to \bar{\mathbf{x}}$. By the separating hyperplane theorem, corresponding to each $\mathbf{y}^k$ there exists a vector $\mathbf{a}^k$ with $\|\mathbf{a}^k\| = 1$ such that $(\mathbf{a}^k)^T(\mathbf{y}^k - \mathbf{x}) > 0$ for all $\mathbf{x} \in cl(S)$. Since $\{\mathbf{a}^k\}$ is bounded, it has a convergent subsequence $\{\mathbf{a}^k\}_{\mathcal{K}} \to \mathbf{a}$ with $\|\mathbf{a}\| = 1$. Thus, we have $\mathbf{a}^T\mathbf{x} \leq \mathbf{a}^T\bar{\mathbf{x}}$ for each $\mathbf{x} \in cl(S)$.

Corollary of Separating Hyperplane Theorem

- **Corollary 1** *Let* $S \in \mathcal{R}^n$ *be a nonempty convex set and* $\bar{\mathbf{x}} \notin int(S)$*. Then there exists a vector* $\mathbf{a} \neq \mathbf{0}$ *such that* $\mathbf{a}^T\mathbf{x} \leq \mathbf{a}^T\bar{\mathbf{x}}$ *for each* $\mathbf{x} \in cl(S)$*.*

- **Corollary 2** *Let* $S \in \mathcal{R}^n$ *be a nonempty convex set and* $\bar{\mathbf{x}} \notin S$*. Then there exists a vector* $\mathbf{a} \neq \mathbf{0}$ *such that* $\mathbf{a}^T\mathbf{x} \leq \mathbf{a}^T\bar{\mathbf{x}}$ *for each* $\mathbf{x} \in cl(S)$*.*

Separating Hyperplane Theorems for Two Convex Sets

- **Theorem 4** *If S_1 and S_2 are two nonempty closed convex subsets of $\mathcal{R}^n$, $S_1 \cap S_2 = \emptyset$, S_1 is bounded, then there exists a vector $\mathbf{a} \neq \mathbf{0}$ and $\varepsilon > 0$ such that*

强分离

$$\textbf{inf}\{\mathbf{a}^T\mathbf{x} : \mathbf{x} \in S_1\} \geq \varepsilon + \textbf{sup}\{\mathbf{a}^T\mathbf{x} : \mathbf{x} \in S_2\}.$$

- **Theorem 5** *If S_1 and S_2 are two nonempty convex subsets of $\mathcal{R}^n$, $S_1 \cap S_2 = \emptyset$, then there exists a vector $\mathbf{a} \neq \mathbf{0}$ such that*

分离

$$\textbf{inf}\{\mathbf{a}^T\mathbf{x} : \mathbf{x} \in S_1\} \geq \textbf{sup}\{\mathbf{a}^T\mathbf{x} : \mathbf{x} \in S_2\}.$$

Proof of Separating Hyperplane Theorems for Two Convex Sets

Let $\mathcal{A} = S_2 - S_1$. Then $\mathcal{A}$ is a nonempty closed convex set and $\mathbf{0} \notin \mathcal{A}$. Obviously, $\mathcal{A}$ is nonempty and convex. We now show that $\mathcal{A}$ is closed. Let $z^k \in \mathcal{A}$ and tend to z. Then we have

$$z^k = y^k - x^k, \quad y^k \in S_2, \quad x^k \in S_1$$

Since S_1 is bounded, then the sequence $\{x^k\}$ has a convergent subsequence. WLOG, assume $x^k \to x$. Thus, y^k has a limit point $x + z$. Let $y = x + z$. we have $y \in S_2$ due to S_2 is closed. So, $z = y - x \in \mathcal{A}$. This implies that $\mathcal{A}$ is closed. Hence, there exists a nonzero vector $\mathbf{a}$ such that

$$\textbf{sup}\{\mathbf{a}^T\mathbf{x} : \mathbf{x} \in \mathcal{A}\} < 0.$$

We complete the proof.

Real Functions

- Continuous functions C
- The gradient vector C^1:

$$\nabla f(\mathbf{x}) = \{\partial f/\partial x_i\}, \quad \textbf{for} \quad i = 1, ..., n.$$

- The Hessian matrix C^2:

$$\nabla^2 f(\mathbf{x}) = \left\{ \frac{\partial^2 f}{\partial x_i \partial x_j} \right\} \quad \textbf{for} \quad i = 1, ..., n;\ j = 1, ..., n.$$

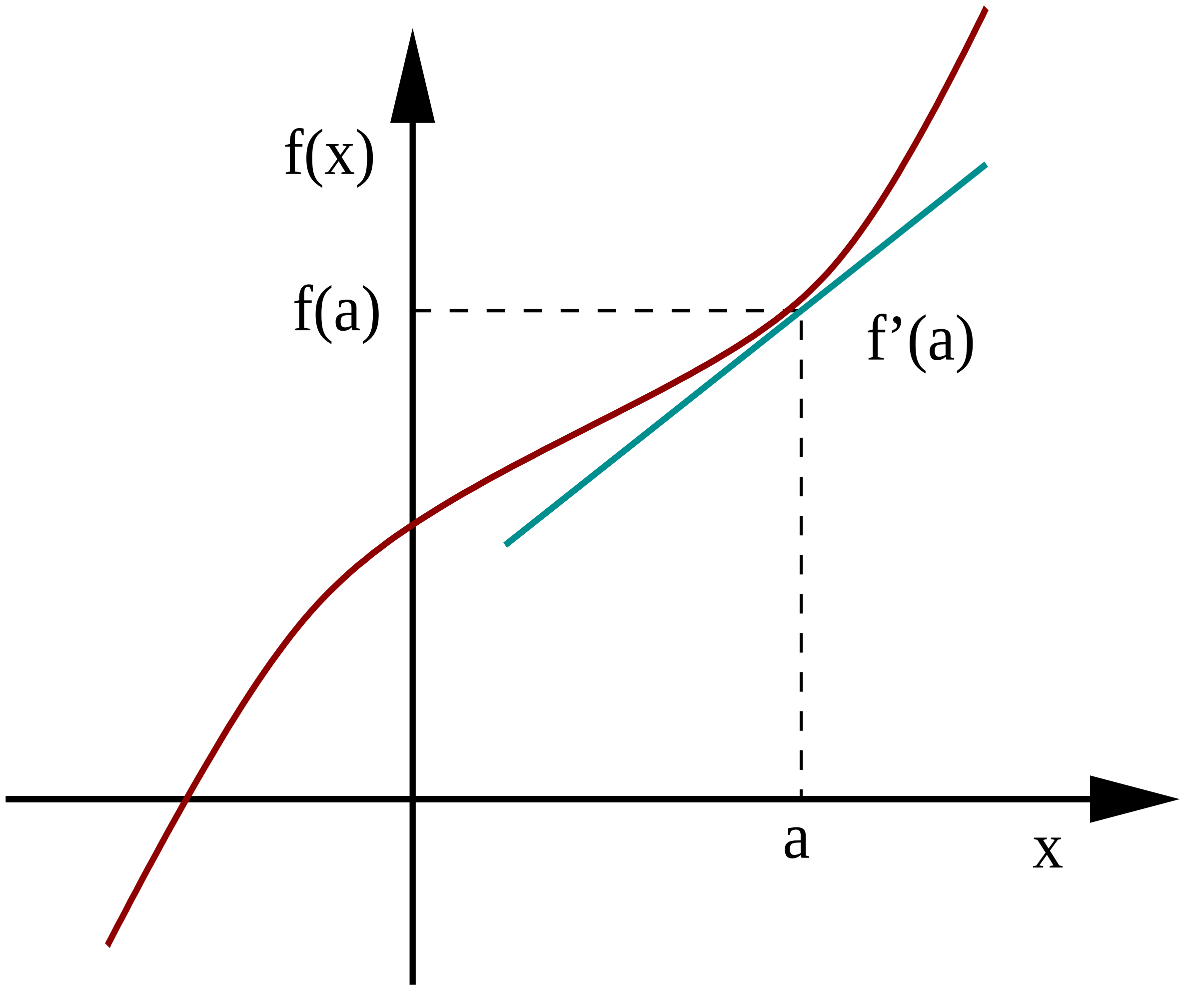

Figure 4: Derivative and slope

Real Functions

- Weierstrass theorem: a continuous function $f(\mathbf{x})$ defined on a compact set (bounded and closed) $\Omega \subset \mathcal{R}^n$ has a minimizer in Ω.
- The least upper bound or supremum of f over Ω

$$\sup\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$$

and the greatest lower bound or infimum of f over Ω

$$\inf\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$$

Real Functions

- Vector function: $\mathbf{f} = (f_1; f_2; ...; f_m)$
- The Jacobian matrix of $\mathbf{f}$ is

$$\nabla \mathbf{f}(\mathbf{x}) = \begin{pmatrix} \nabla f_1(\mathbf{x}) \\ ... \\ \nabla f_m(\mathbf{x}) \end{pmatrix}.$$

Convex Function

- f convex function iff for $0 \leq \alpha \leq 1$,

$$f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}).$$

  If strict inequality holds whenever $\mathbf{x} \neq \mathbf{y}$, then f is said to be strictly convex.

- The level set of convex function f

$$L(z) = \{\mathbf{x} : \quad f(\mathbf{x}) \leq z\}$$

  is a convex set. The converse is not true; e.g., $f(x) = x^3$.

称这样的函数为拟凸函数

quasi-convex

R 上的所有单调函数均拟凸函数

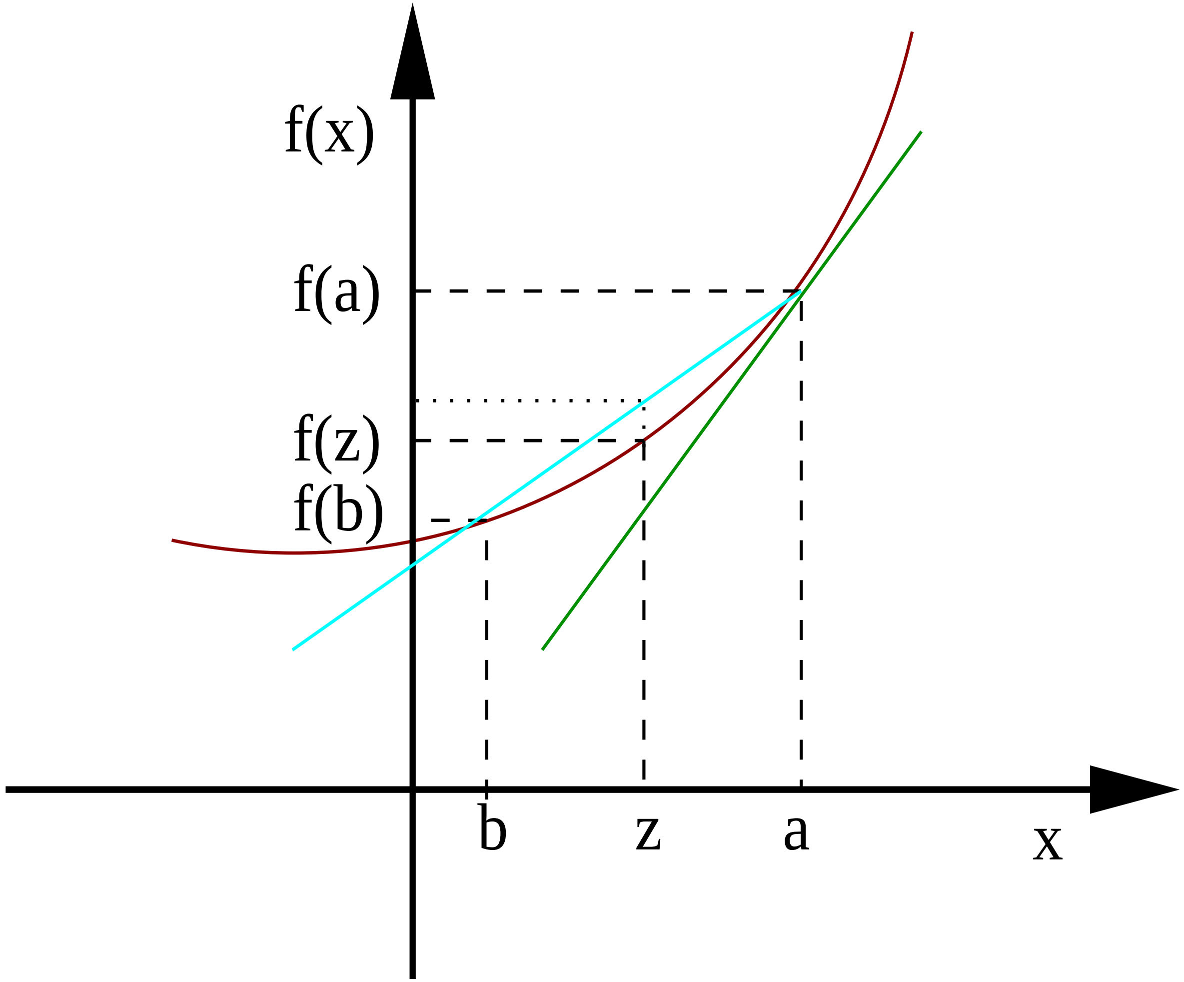

Figure 5: Properties of convex function

Examples

- Linear functions are both convex and concave. 线性函数
- Positive semidefinite quadratic forms are convex. 半正定二次型
- Positively-weighted sums of convex functions are convex. 凸函数的正权的平均

Jensen's Inequality

Theorem 6 *If $f : C \to \mathcal{R}$ is convex, then*

$$f(\lambda_1 \mathbf{x}^1 + \ldots + \lambda_m \mathbf{x}^m) \leq \lambda_1 f(\mathbf{x}^1) + \ldots + \lambda_m f(\mathbf{x}^m)$$

for any $\mathbf{x}^1, \ldots, \mathbf{x}^m \in C$ and any $\lambda_1 \geq 0, \ldots, \lambda_m \geq 0$ such that $\lambda_1 + \ldots + \lambda_m = 1$.

Proof of Jensen's Inequality

Proof. For $m = 2$ the inequality is just the definition of convexity. Arguing inductively, we now assume $m > 2$ and that the inequality holds for $m - 1$ points. For m points we have

$$\lambda_1 \mathbf{x}^1 + \ldots + \lambda_m \mathbf{x}^m = \lambda_1 \mathbf{x}^1 + (1 - \lambda_1) \left[\frac{\lambda_2}{1 - \lambda_1} \mathbf{x}^2 + \ldots + \frac{\lambda_m}{1 - \lambda_1} \mathbf{x}^m \right],$$

and then the rest of the proof is obtained by using the convexity of f together with the inductive hypothesis.

Theorems on Functions

Taylor's theorem or the mean-value theorem:

Theorem 7 *Let* $f \in C^1$ *be in a region containing the line segment* $[\mathbf{x}, \mathbf{y}]$*. Then there is a* α*,* $0 \leq \alpha \leq 1$*, such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y})(\mathbf{y} - \mathbf{x}).$$

Furthermore, if $f \in C^2$ *then there is a* α*,* $0 \leq \alpha \leq 1$*, such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y}-\mathbf{x}) + (1/2)(\mathbf{y}-\mathbf{x})^T \nabla^2 f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y})(\mathbf{y}-\mathbf{x}).$$

Taylor's Theorem for Multivariate Functions

Theorem 8 *Suppose $X \subseteq \mathcal{R}^n$ is open, $x \in X$, and $f : X \to \mathcal{R}$ is differentiable. Then*

$$f(x+h) = f(x) + \nabla f(x)h + o(\|h\|) \textbf{ as } h \to 0.$$

If $f \in C^2$, then

$$f(x+h) = f(x) + \nabla f(x)h + \frac{1}{2}h^T \nabla^2 f(x)h + o(\|h\|^2) \textbf{ as } h \to 0.$$

Theorems on Convex Functions

Theorem 9 *Let $f \in C^1$. Then f is convex over a convex set Ω if and only if the gradient inequality holds, i.e.,*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

for all $\mathbf{x},\ \mathbf{y} \in \Omega$.

Theorem 10 *Let $f \in C^2$. Then f is convex over a open convex set Ω if and only if the Hessian matrix of f is positive semi-definite throughout Ω.*

Example $f(x_1, x_2) = x_1^2 - 2x_1x_2 + x_2^2 + x_1 + x_2$ is convex.

Remarks

Remark 1 Let f be differentiable on the convex set Ω. Then f is strictly convex on Ω iff strict inequality holds in the gradient inequality for all pairs of distinct points $\mathbf{x},\ \mathbf{y} \in \Omega$.

Example This result can be used to prove that the univariate function $f(x) = \dfrac{1}{x}$ is strictly convex when Ω is the positive real line.

Remark 2 If the Hessian matrix of f is positive definite throughout Ω, then f is strictly convex on Ω. But the converse is false, as shown by the function $f(x) = x^4$ with domain $\Omega = \mathcal{R}$.

Remark

• A convex function need not be differentiable. As a matter of fact, it is not even necessary for a convex function to be continuous.

$$f(x) = |x|, \qquad g(x) = \begin{cases} 1 & \textbf{if } x = 0 \\ 0 & \textbf{if } 0 < x < 1 \\ 1 & \textbf{if } x = 1 \end{cases}$$

It should be noticed, however, that the function $g : [0, 1] \to \mathcal{R}$ is continuous on the interior of its domain. This is a consequence of a general result. Every convex function is continuous on the relative interior of its domain.

Known Inequalities

- Cauchy-Schwarz: given $\mathbf{x}, \mathbf{y} \in \mathcal{R}^n$, $\mathbf{x}^T\mathbf{y} \le \|\mathbf{x}\|\|\mathbf{y}\|$.
- Arithmetic-geometric mean: given $\mathbf{x} > \mathbf{0}$,

$$\frac{\sum x_j}{n} \ge \left(\prod x_j\right)^{1/n}.$$

- Harmonic: given $\mathbf{x} > \mathbf{0}$,

$$\left(\sum x_j\right)\left(\sum 1/x_j\right) \ge n^2.$$

System of Linear Equations

Solve for $\mathbf{x} \in \mathcal{R}^n$ from:

$$\left.\begin{array}{rcl} \mathbf{a}_1\mathbf{x} & = & b_1 \\ \mathbf{a}_2\mathbf{x} & = & b_2 \\ \ldots & . & . \\ \mathbf{a}_m\mathbf{x} & = & b_m \end{array}\right. \quad \Rightarrow A\mathbf{x} = \mathbf{b}$$

Basic solution: select m columns from A to form a square matrix A_B such that

$$A_B\mathbf{x}_B = \mathbf{b}, \quad \textbf{the rest of} \quad \mathbf{x}_N = \mathbf{0}$$

where B is the index set of selected m columns.

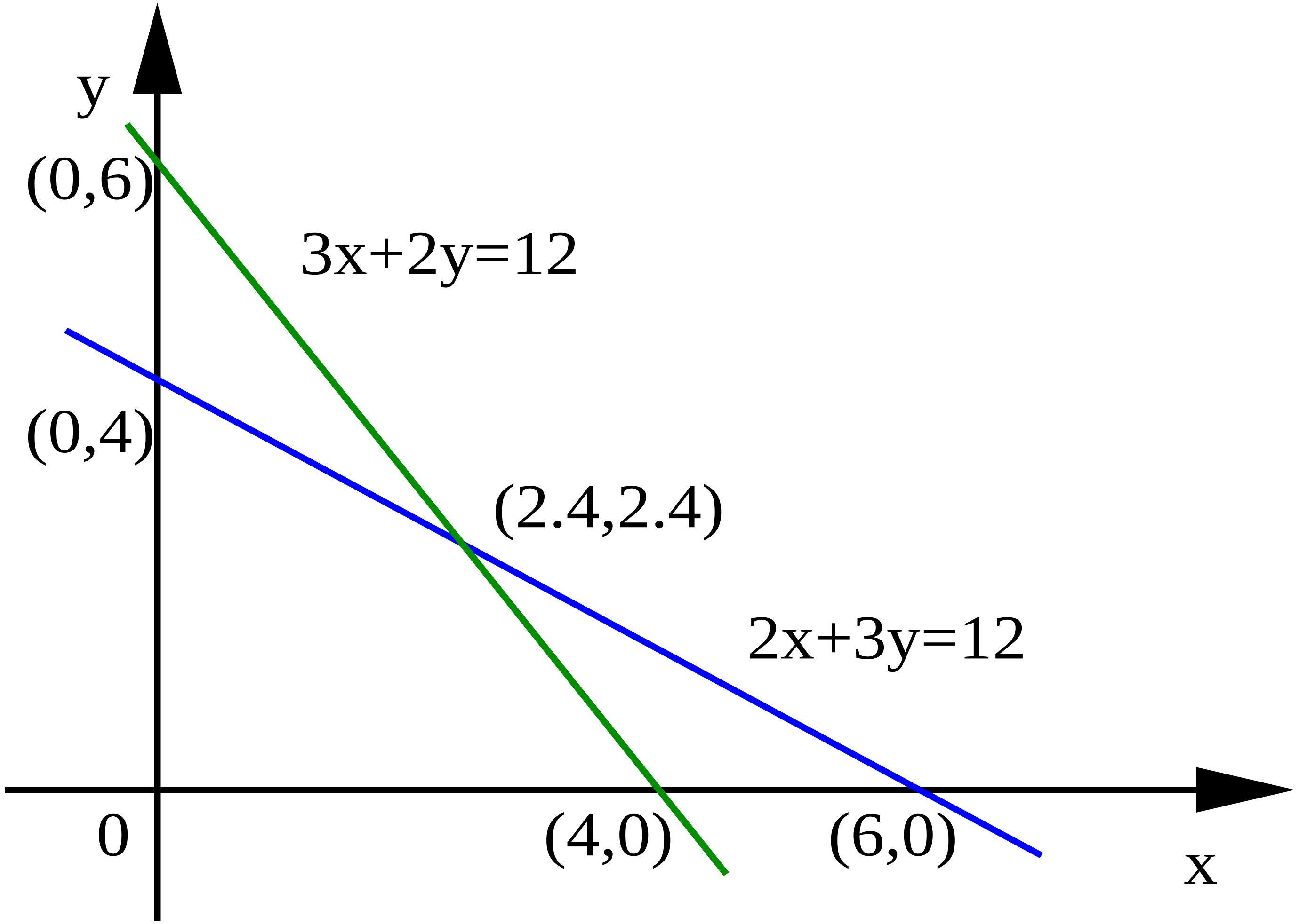

Figure 6: System of Linear Equations

Fundamental Theorem of Linear Equations

Theorem 11 *Let $A \in \mathcal{R}^{m\times n}$ and $\mathbf{b} \in \mathcal{R}^m$. The system $\{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}\}$ has a solution if and only if that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{b}^T\mathbf{y} \neq 0$ has no solution.*

择一系统

Proof: We assume that there exists a vector $\mathbf{y}$ such that

$$A^T\mathbf{y} = \mathbf{0}, \quad \mathbf{b}^T\mathbf{y} \neq 0.$$

Let $\bar{\mathbf{x}}$ be a solution of the system $\{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}\}$. Then

$$0 \neq \mathbf{b}^T\mathbf{y} = \bar{\mathbf{x}}^T A^T\mathbf{y} = 0,$$

which is a contradiction.

Proof of Fundamental Theorem of Linear Equations

Conversely, suppose that $\mathbf{b} \notin \{A\mathbf{x} : \mathbf{x} \in \mathcal{R}^n\}$. Define

$$C := \{A\mathbf{x} : \mathbf{x} \in \mathcal{R}^n\}.$$

Then C is a non-empty, closed and convex set (Why?). From the separating hyperplane theorem, there exists a vector $\mathbf{a} \neq \mathbf{0}$ such that

$$\mathbf{b}^T\mathbf{a} > \sup_{\mathbf{x} \in \mathcal{R}^n} \mathbf{x}^T(A^T\mathbf{a}).$$

Making an arbitrary choice of $\mathbf{x}$, we have $A^T\mathbf{a} = \mathbf{0}, \mathbf{b}^T\mathbf{a} \neq 0$.

Another Way of Proof

Suppose the system $\{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}\}$ has no solution. From the fact that the space $\mathcal{R}^m$ can be expressed as the range of A plus the nullspace of A^T. Then there exist vectors $\mathbf{x}$ and **u** such that

$$b = A\mathbf{x} + Z^T\mathbf{u}$$

where Z^T is a basis for the nullspace of A^T.

Take $\mathbf{a} := Z^T\mathbf{u} \neq \mathbf{0}$ (*otherwise the first system has a solution*). Thus, we have $A^T\mathbf{a} = \mathbf{0}$ and

$$\mathbf{b}^T\mathbf{a} = \mathbf{x}^T A^T Z^T\mathbf{u} + \mathbf{u}^T Z Z^T\mathbf{u} = 0 + \mathbf{a}^T\mathbf{a} > 0,$$

which is to say that the second system has a solution.

System of Linear Inequalities

Solve for $\mathbf{x} \in \mathcal{R}^n$ from:

$$\begin{array}{rcl} \mathbf{a}_1\mathbf{x} & \leq & b_1 \\ \mathbf{a}_2\mathbf{x} & \leq & b_2 \\ \cdots & . & . \\ \mathbf{a}_m\mathbf{x} & \leq & b_m \end{array} \quad \Rightarrow A\mathbf{x} \leq \mathbf{b}$$

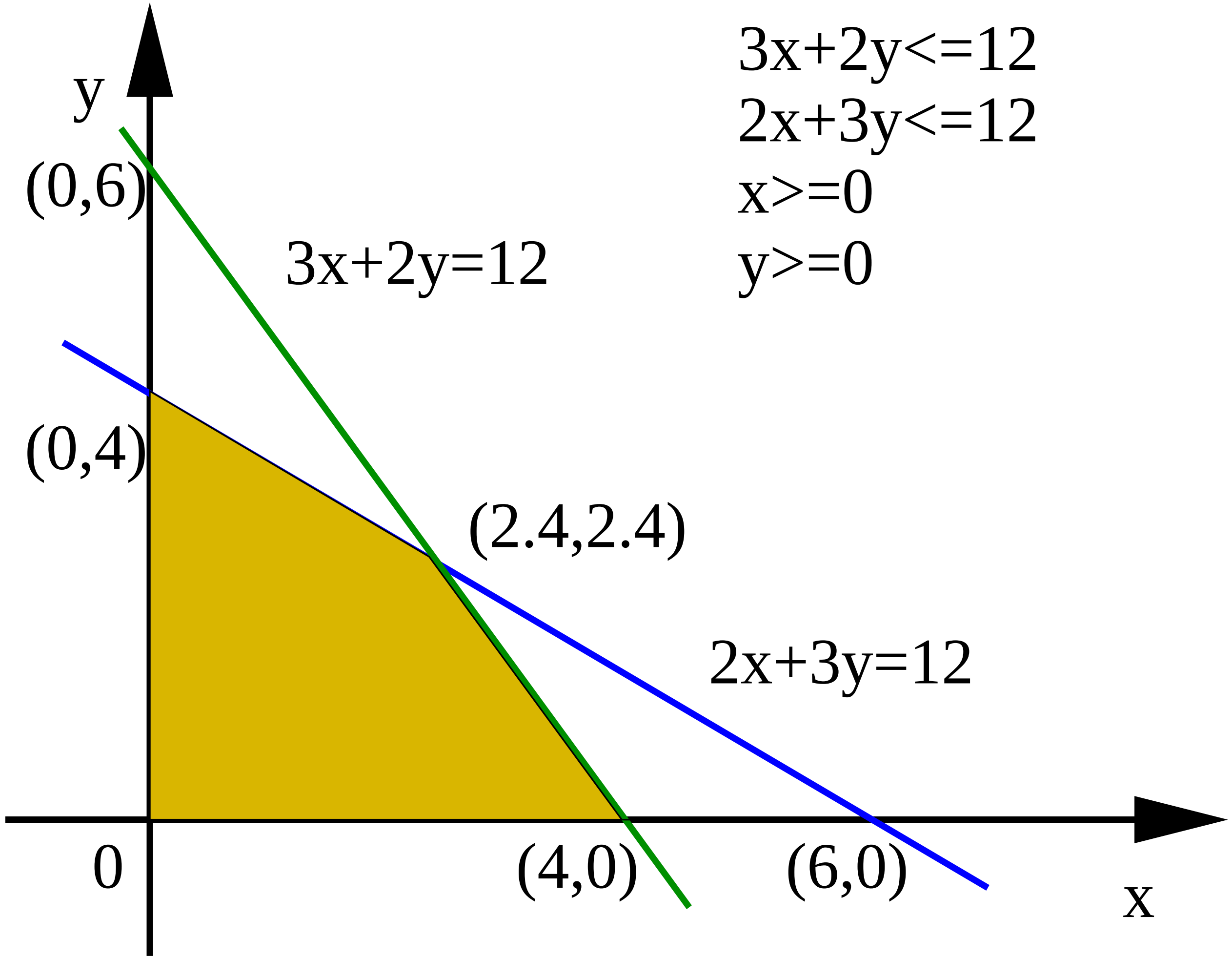

Figure 7: System of Linear Inequalities

Fundamental Theorem of Linear Inequalities

Theorem 12 *(Farkas' Lemma) Let $A \in \mathcal{R}^{m\times n}$ and $\mathbf{c} \in \mathcal{R}^n$. The system $\{\mathbf{x} : \; A\mathbf{x} \leq \mathbf{0}, \mathbf{c}^T\mathbf{x} > 0\}$ has a solution if and only if that $A^T\mathbf{y} = \mathbf{c}$ and $\mathbf{y} \geq \mathbf{0}$ has no solution.*

Proof: Suppose, reasoning by contradiction, that the system

$$\{\mathbf{y} : A^T\mathbf{y} = \mathbf{c}, \; \mathbf{y} \geq \mathbf{0}\}$$

has a solution $\bar{\mathbf{y}}$. Let $\bar{\mathbf{x}} \in \{\mathbf{x} : \; A\mathbf{x} \leq \mathbf{0}, \; \mathbf{c}^T\mathbf{x} > 0\}$. Then

$$0 < \mathbf{c}^T\bar{\mathbf{x}} = \bar{\mathbf{x}}^T(A^T\bar{\mathbf{y}}) = (A\bar{\mathbf{x}})^T\bar{\mathbf{y}} \leq 0,$$

which is a contradiction.

Proof of Farkas' Lemma

闭凸集

Conversely, consider convex cone $C := \{A^T\mathbf{y} : \mathbf{y} \geq \mathbf{0}\}$. We must have $\mathbf{c} \notin C$. From the separating hyperplane theorem, there is a vector $\mathbf{x} \neq \mathbf{0}$ such that

$$\mathbf{c}^T\mathbf{x} > \sup_{\mathbf{y} \geq \mathbf{0}} \mathbf{y}^T(A\mathbf{x}).$$

We must have (i) $\mathbf{c}^T\mathbf{x} > 0$ since $\mathbf{y} = \mathbf{0}$ makes $\mathbf{y}^T(A\mathbf{x}) = 0$. (ii) $A\mathbf{x} \leq \mathbf{0}$ since, otherwise, say the first element of $A\mathbf{x}$ is positive, we can choose $\mathbf{y} = (\beta; 0; \cdots; 0)$ and let $\beta \to +\infty$. Then, $\mathbf{y}^T(A\mathbf{x}) \to +\infty$ which is a contradiction to that $\mathbf{y}^T(A\mathbf{x})$ is bounded above for any $\mathbf{y} \geq \mathbf{0}$.

Alternative System

Given matrix $A \in \mathcal{R}^{m\times n}$ and vector $\mathbf{b} \in \mathcal{R}^m$, exactly one of the following two systems is feasible:

$$A\mathbf{x} \leq \mathbf{b},$$

or

$$A^T\mathbf{y} = \mathbf{0},\ \mathbf{y} \geq \mathbf{0},\ \mathbf{b}^T\mathbf{y} < 0.$$

Proof of the Alternative Theorem

The following system

$$A\mathbf{x} \le \mathbf{b}$$

can be reformulated as

$$A(\mathbf{u} - \mathbf{v}) + \mathbf{s} = \mathbf{b}, \quad \mathbf{u}, \mathbf{v}, \mathbf{s} \ge \mathbf{0}.$$

By Farkas' Lemma, its alternative system is

$$(A \quad -A \quad I_m)^T \mathbf{w} \le \mathbf{0}, \quad \mathbf{b}^T \mathbf{w} > 0.$$

Let $\mathbf{y} = -\mathbf{w}$, the equivalent formulation of the alternative system is

$$A^T \mathbf{y} = \mathbf{0}, \ \mathbf{y} \ge \mathbf{0}, \ \mathbf{b}^T \mathbf{y} < 0.$$

Alternative System Continued

Gordan Theorem Given matrix $A \in \mathcal{R}^{m\times n}$, exactly one of the following two systems is feasible:

$$A\mathbf{x} < \mathbf{0},$$

or

$$\mathbf{y} \geq \mathbf{0},\ \mathbf{y} \neq \mathbf{0},\ A^T\mathbf{y} = \mathbf{0}.$$

Proof of Gordan Theorem

We rewrite the system $A\mathbf{x} < \mathbf{0}$ as

$$(I)\ A\mathbf{x} + \alpha\mathbf{e} \leq \mathbf{0}, \quad \alpha > 0.$$

By Farkas' Lemma, the alternative system of (I) is

$$(A \quad \mathbf{e})^T\mathbf{y} = (\mathbf{0} \quad 1)^T, \quad \mathbf{y} \geq \mathbf{0},$$

i.e.,

$$A^T\mathbf{y} = \mathbf{0},\ \mathbf{y} \geq \mathbf{0},\ \mathbf{y} \neq \mathbf{0}.$$

Economic Applications of Farkas' Lemma

Example: Suppose a stock's current price is 10. Its price tomorrow can be either 0 or 15. What is the probability that it is 15 tomorrow (Ignore discounting)?

- Assume that today's price is the expected price tomorrow. Hence,

$$10 = Prob(p_{tomorrow} = 0) \times 0 + Prob(p_{tomorrow} = 15) \times 15$$

which gives

$$Prob(p_{tomorrow} = 15) = \frac{10}{15} = 67\%.$$

Two Stocks Case

Suppose now there are two stocks, X and Y, and two possible states tomorrow: G and B.

The price of X today is 10, and its prices tomorrow are 0 when the state is bad, and 15 when the state is good, that is, $p_{X_0} = 10$, $p_{X_B} = 0$, and $p_{X_G} = 15$.

The price of Y today is 15, and its prices tomorrow are 0 when the state is bad, and 20 when the state is good, that is, $p_{Y_0} = 15$, $p_{Y_B} = 0$, and $p_{Y_G} = 20$.

Then what is the probability that tomorrow is good?

Two Stocks Case Continued

- Using the data of X, this probability should be 67%. But if we use the data of Y, this probability should be 75%. What goes wrong here?

The problem is that these assets are not priced right. If we (long) buy today four shares of X (paying 40) and (short) sell three shares of Y (receiving 45), we net a profit 5 today.

套利

When tomorrow is bad, we can buy back Y and sell X without any cost. When tomorrow is good, we can swap our four shares of X for three shares of Y. So our portfolio will not have a negative value no matter which state obtains tomorrow! Of course, this is impossible since any well-run market should not allow any investor to have such arbitrage opportunities.

No Arbitrage Condition and Asset Pricing

Suppose there are n stocks. The i-th stock's current price S_i. There are m possible states tomorrow. In state j, the i-th stock's price will be T_{ij}. Under what conditions can we find interest rate R and probability number $Prob(j)$ for every state j such that, every stock's current price S_i is the discounted expected value of its prices tomorrow:

$$(EV) \qquad S_i = \frac{1}{1+R} \sum_{j=1}^{m} Prob(j) T_{ij}?$$

When (EV) hold, the number R is called the market-implied interest rate and $Prob(j)$ the market-implied probability.

n Stocks Case

Solution: A portfolio of stocks can be represented by a vector $p \in \mathcal{R}^n$. For example, the vector $(3, -4, \cdots, 1)$ is a portfolio that consist of long 3 shares of stock 1, short 4 shares of stock 2,$\ldots$, and long 1 share of stock n. The cost basis of a portfolio p today is $p^T S$ (S is the vector with S_i as coordinates). When state j realizes tomorrow, the value of the portfolio will be $p^T T_j$ (T_j is the column vector of the matrix (T_{ij})).

We say there is an arbitrage opportunity if one can find a portfolio p that has a negative cost basis today and has a non-negative value tomorrow no matter which state obtains. Obviously, a well-run market will not allow such opportunities at equilibrium.

No Arbitrage Theorem

Theorem 13 *Given the market price data S and S_i, (EV) holds for some interest rate R and (subjective) probability $Prob(j)$ if and only if there is no arbitrage opportunity, that is, $p^T S \geq 0$ for any portfolio p with* $p^T T_j \geq 0\,(j = 1, \cdots, m)$.

Linear Least-Squares Problem

Given $A \in \mathcal{R}^{m\times n}$ and $\mathbf{c} \in \mathcal{R}^n$,

$$
\begin{array}{lll}
(LS) & \textbf{minimize} & \|A^T\mathbf{y} - \mathbf{c}\|^2 \\
 & \textbf{subject to} & \mathbf{y} \in \mathcal{R}^m.
\end{array}
$$

$$AA^T\mathbf{y} = A\mathbf{c} \quad \textbf{or} \quad \mathbf{y} = (AA^T)^{-1}A\mathbf{c}$$

with the projection:

$$A^T\mathbf{y} = A^T(AA^T)^{-1}A\mathbf{c}$$

Projection matrix: $P = A^T(AA^T)^{-1}A$ or $P = I - A^T(AA^T)^{-1}A$

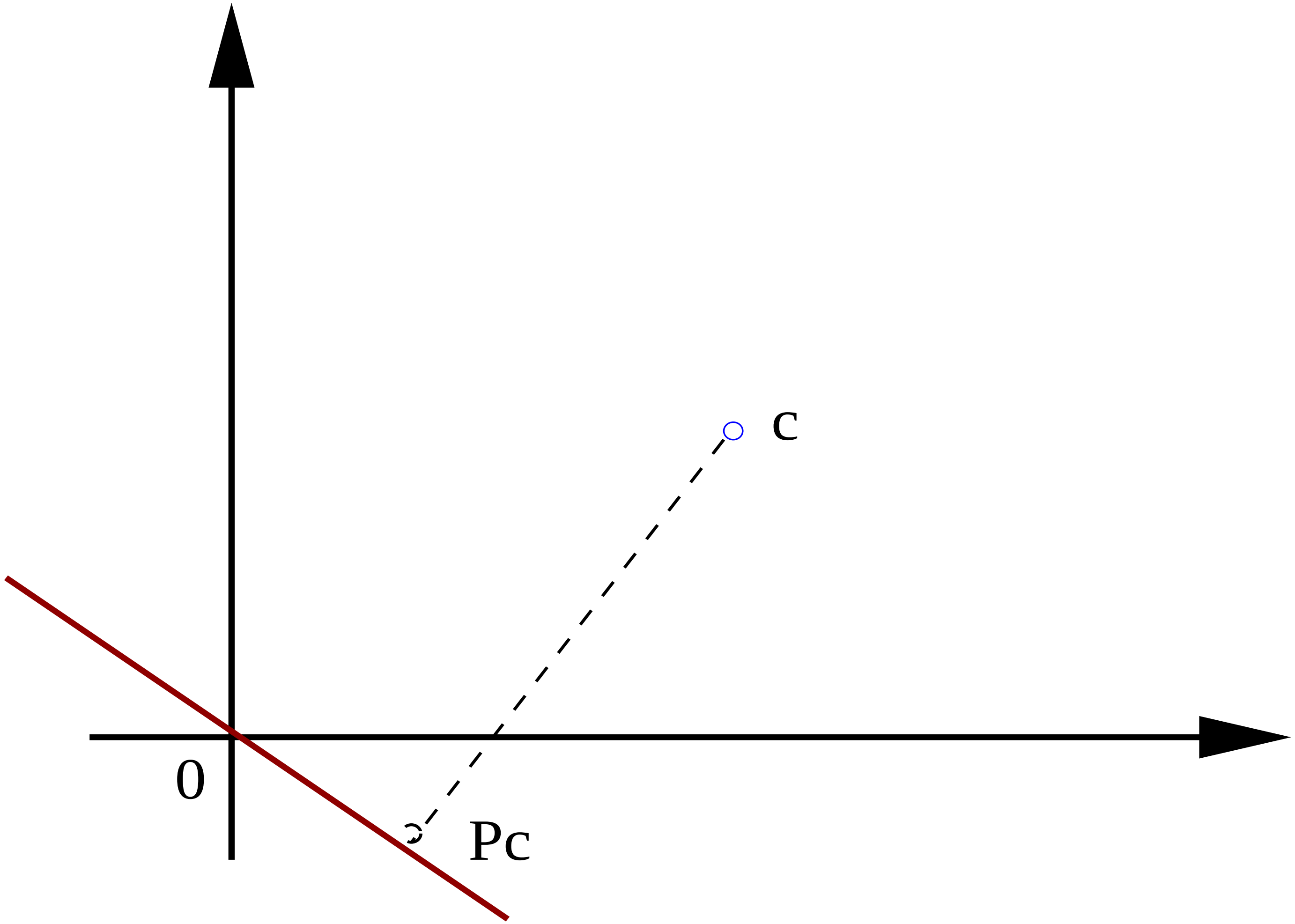

Figure 8: Projection of **c** onto a subspace

System of Nonlinear Equations

Given $\mathbf{f}(\mathbf{x}) : \mathcal{R}^n \to \mathcal{R}^n$, the problem is to solve n equations for n unknowns:

$$\mathbf{f}(\mathbf{x}) = \mathbf{0}.$$

Given a point $\mathbf{x}^k$, Newton's Method sets

$$f(\mathbf{x}) \simeq f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) = \mathbf{0}.$$

$$\mathbf{x}^{k+1} = \mathbf{x}^k - (\nabla f(\mathbf{x}^k))^{-1} f(\mathbf{x}^k)$$

or solve for direction vector $\mathbf{d}_x$:

$$\nabla f(\mathbf{x}^k)\mathbf{d}_x = -f(\mathbf{x}^k) \quad \textbf{and} \quad \mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{d}_x.$$

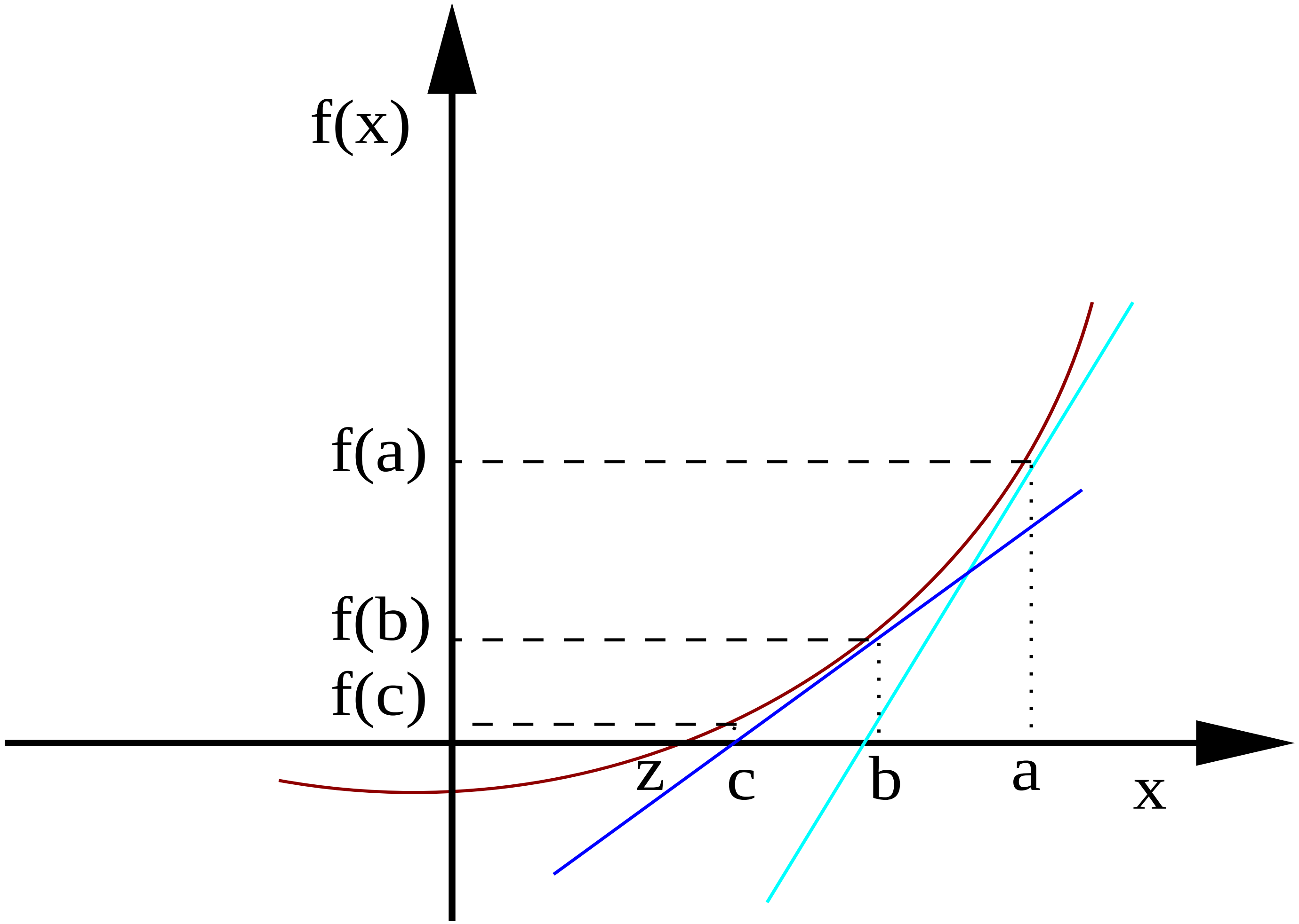

Figure 9: Newton's method for root finding