

Generalized inequalities

- Minimum and minimal elements

We say that $\mathbf{x} \in S$ is the *minimum element* of S (with respect to the generalized inequality \preceq_K) if for every $\mathbf{y} \in S$ we have $\mathbf{x} \preceq_K \mathbf{y}$. We define the *maximum element* of a set S , with respect to a generalized inequality, in a similar way. If a set has a minimum (maximum) element, then it is unique. A related concept is minimal element. We say that $\mathbf{x} \in S$ is a *minimal element* of S (with respect to the generalized inequality \preceq_K) if $\mathbf{y} \in S$, $\mathbf{y} \preceq_K \mathbf{x}$ only if $\mathbf{y} = \mathbf{x}$. We define *maximal element* in a similar way. A set can have many different minimal (maximal) elements.

total ordering vs. partial ordering:

1. Reflexivity: $a \preceq a$, for all $a \in \mathcal{A}$;
2. Antisymmetry: $a \preceq b$ and $b \preceq a$ imply $a = b$;
3. Transitivity: $a \preceq b$ and $b \preceq c$ imply $a \preceq c$;
4. Comparability: for all a and b in \mathcal{A} , either $a \preceq b$ or $b \preceq a$.

Generalized inequalities

- Minimum and minimal elements: set notation

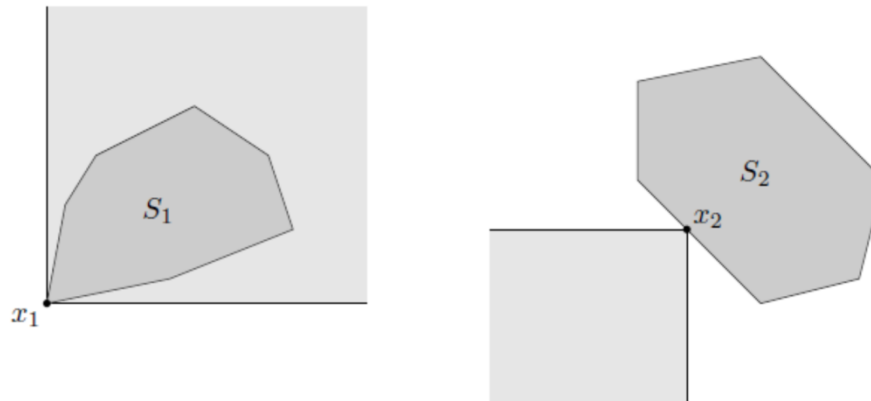
A point $\mathbf{x} \in S$ is the *minimum element* of S if and only if

$$S \subseteq \mathbf{x} + K.$$

Here $\mathbf{x} + K$ denotes all the points that are comparable to \mathbf{x} and greater than or equal to \mathbf{x} (according to \preceq_K). A point $\mathbf{x} \in S$ is a *minimal element* if and only if

$$(\mathbf{x} - K) \cap S = \{\mathbf{x}\}.$$

Here $\mathbf{x} - K$ denotes all the points that are comparable to \mathbf{x} and less than or equal to \mathbf{x} (according to \preceq_K), the only point in common with S is \mathbf{x} .



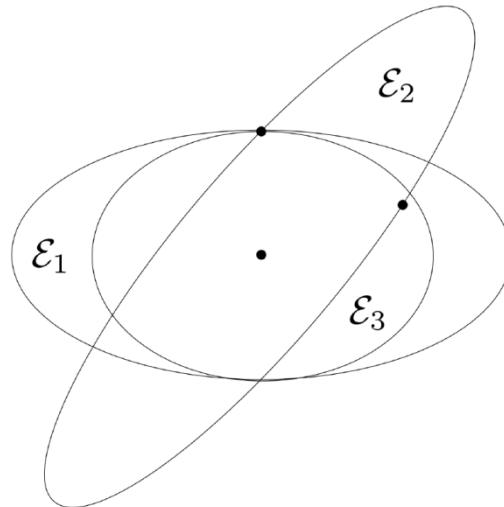
Generalized inequalities

- Minimum and minimal elements: examples

We associate with each $\mathbf{A} \in \mathbb{S}_{++}^n$ an ellipsoid centered at the origin, given by

$$\varepsilon_{\mathbf{A}} = \{\mathbf{x} \mid \mathbf{x}^T \mathbf{A}^{-1} \mathbf{x} \leq 1\}.$$

We have $\mathbf{A} \preceq \mathbf{B}$ if and only if $\varepsilon_{\mathbf{A}} \subseteq \varepsilon_{\mathbf{B}}$. Let $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ be given and define S to be the set of ellipsoids that contain these points. The set S does not have a minimum element: for any ellipsoid that contains the points $\mathbf{v}_1, \dots, \mathbf{v}_k$ we can find another one that contains the points, and is not comparable to it. An ellipsoid is minimal if it contains the points, but no smaller ellipsoid does.

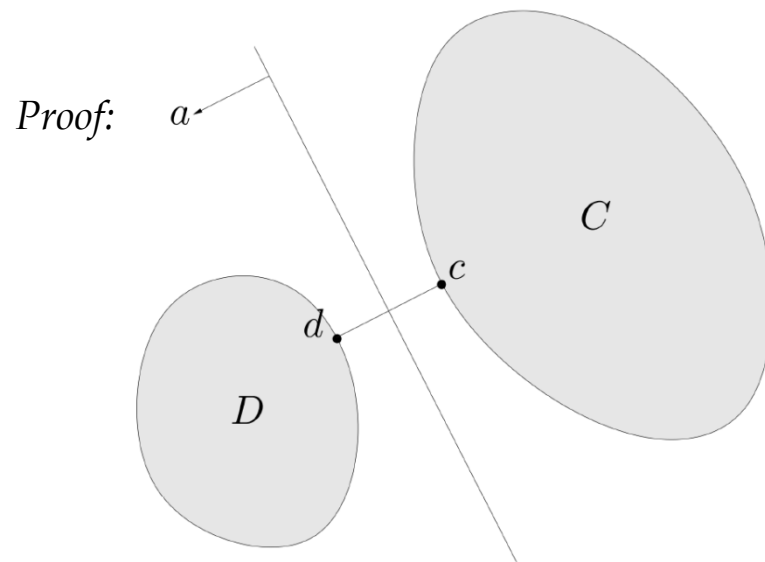
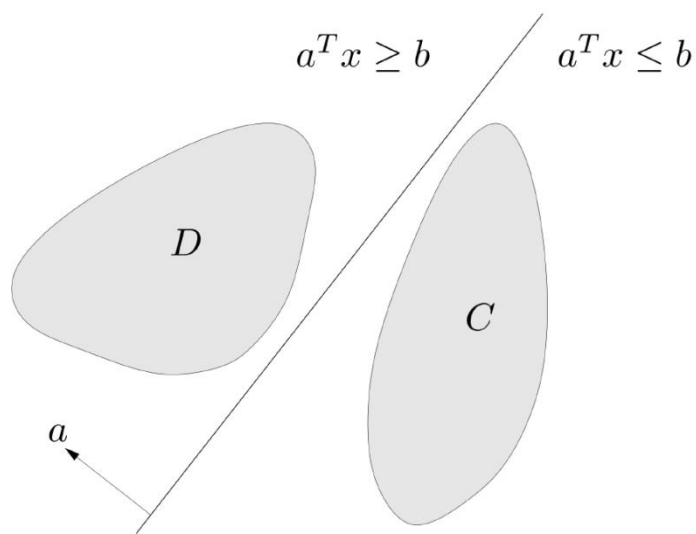


Separating and supporting hyperplanes

- Separating hyperplane theorem

Theorem 1. Suppose C and D are two convex sets that do not intersect, i.e., $C \cap D = \emptyset$. Then there exist $\mathbf{a} \neq \mathbf{0}$ and b such that $\mathbf{a}^T \mathbf{x} \leq b$ for all $\mathbf{x} \in C$ and $\mathbf{a}^T \mathbf{x} \geq b$ for all $\mathbf{x} \in D$. In other words, the affine function $\mathbf{a}^T \mathbf{x} - b$ is nonpositive on C and nonnegative on D .

The hyperplane $\{\mathbf{x} | \mathbf{a}^T \mathbf{x} = b\}$ is called a *separating hyperplane* for the sets C and D , or is said to separate the sets C and D .



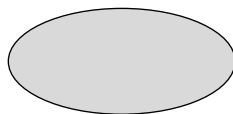
Separating and supporting hyperplanes

- Separating hyperplane theorem: example

Suppose C is convex and D is affine, i.e., $D = \{\mathbf{F}\mathbf{u} + \mathbf{g} \mid \mathbf{u} \in \mathbb{R}^m\}$, where $\mathbf{F} \in \mathbb{R}^{n \times m}$. Suppose C and D are disjoint, so by the separating hyperplane theorem there are $\mathbf{a} \neq 0$ and b such that $\mathbf{a}^T \mathbf{x} \leq b$ for all $\mathbf{x} \in C$ and $\mathbf{a}^T \mathbf{x} \geq b$ for all $\mathbf{x} \in D$.

Now $\mathbf{a}^T \mathbf{x} \geq b$ for all $\mathbf{x} \in D$ means $\mathbf{a}^T \mathbf{F}\mathbf{u} \geq b - \mathbf{a}^T \mathbf{g}$ for all $\mathbf{u} \in \mathbb{R}^m$. But a linear function is bounded below on \mathbb{R}^m only when it is zero, so we conclude $\mathbf{a}^T \mathbf{F} = \mathbf{0}$ (and hence, $b \leq \mathbf{a}^T \mathbf{g}$).

Thus we conclude that there exists $\mathbf{a} \neq 0$ such that $\mathbf{F}^T \mathbf{a} = \mathbf{0}$ and $\mathbf{a}^T \mathbf{x} \leq \mathbf{a}^T \mathbf{g}$ for all $\mathbf{x} \in C$.

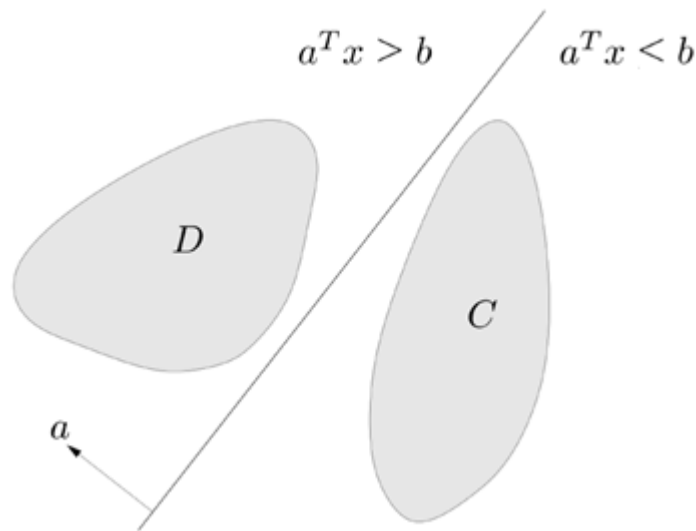


When is the separation hyperplane unique?

Separating and supporting hyperplanes

- Strict separation

If the separating hyperplane satisfies the stronger condition that $\mathbf{a}^T \mathbf{x} < b$ for all $\mathbf{x} \in C$ and $\mathbf{a}^T \mathbf{x} > b$ for all $\mathbf{x} \in D$, then the sets C and D are called *strictly separated*.



Disjoint convex sets need not be strictly separable by a hyperplane (even when the sets are closed)

Example: a point and a closed convex set; a closed convex set is the intersection of all halfspaces that contain it

Separating and supporting hyperplanes

- Converse separating hyperplane theorems

Theorem 1. *Any two convex sets C and D , at least one of which is open, are disjoint if and only if there exists a separating hyperplane.*

Example: (Theorem of alternatives for strict linear inequalities) We derive the necessary and sufficient conditions for solvability of a system of strict linear inequalities $\mathbf{Ax} < \mathbf{b}$.

pair of alternatives

These inequalities are infeasible if and only if the (convex) sets

$$C = \{\mathbf{b} - \mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^n\}, \quad D = \mathbb{R}_{++}^m = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} \succ \mathbf{0}\}$$

do not intersect. The set D is open, C is an affine set. Hence by the above theorem, C and D are disjoint iff there exists a separating hyperplane, i.e., a nonzero $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\mu \in \mathbb{R}$ such that $\boldsymbol{\lambda}^T \mathbf{y} \leq \mu$ on C and $\boldsymbol{\lambda}^T \mathbf{y} \geq \mu$ on D .

$\mu \leq 0$ and $\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \neq \mathbf{0}$.



$\exists \boldsymbol{\lambda}$ s.t. $\boldsymbol{\lambda} \neq \mathbf{0}, \boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{A}^T \boldsymbol{\lambda} = \mathbf{0}, \boldsymbol{\lambda}^T \mathbf{b} \leq 0$.

Separating and supporting hyperplanes

- Converse separating hyperplane theorems

Theorem 1 (Theorem of the Alternative (Farkas' Lemma)). *For $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ the following are strong alternatives:*

1. $\exists \mathbf{x} \in \mathbb{R}_+^n$ such that $\mathbf{Ax} = \mathbf{b}$,
2. $\exists \mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{A}^T \mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^T \mathbf{y} < 0$.

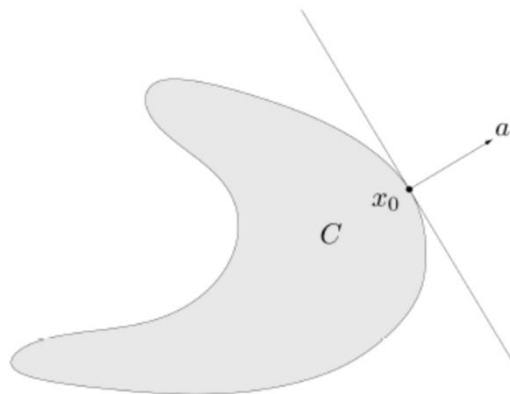
Proof. $1) \implies \neg 2)$: For $\mathbf{x} \in \mathbb{R}_+^n$ with $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{y} \in \mathbb{R}^m$ with $\mathbf{A}^T \mathbf{y} \geq \mathbf{0}$ we have $\mathbf{b}^T \mathbf{y} = \mathbf{x}^T \mathbf{A}^T \mathbf{y} \geq 0$.

$\neg 1) \implies 2)$: $C := \text{cone}(\mathbf{A})$ is a closed convex cone which does not contain the vector \mathbf{b} : by the Separating Hyperplane Theorem there exists a $\mathbf{y} \in \mathbb{R}^m$ with $\langle \mathbf{y}, \mathbf{x} \rangle \geq 0 > \langle \mathbf{y}, \mathbf{b} \rangle$ for all $\mathbf{x} \in C$, in particular $\mathbf{A}_i^T \mathbf{y} = \langle \mathbf{y}, \mathbf{A}_i \rangle \geq 0, \forall i$, that is, $\mathbf{A}^T \mathbf{y} \geq \mathbf{0}$.

Separating and supporting hyperplanes

- Supporting hyperplanes

Suppose $C \subseteq \mathbb{R}^n$, and \mathbf{x}_0 is a point in its boundary ∂C . If $\mathbf{a} \neq \mathbf{0}$ satisfies $\mathbf{a}^T \mathbf{x} \leq \mathbf{a}^T \mathbf{x}_0$ for all $\mathbf{x} \in C$, then the hyperplane $\{\mathbf{x} \mid \mathbf{a}^T \mathbf{x} = \mathbf{a}^T \mathbf{x}_0\}$ is called a *supporting hyperplane* to C at the point \mathbf{x}_0 .



Theorem 1 (Supporting Hyperplane Theorem). *For any nonempty convex set C , and any $\mathbf{x}_0 \in \partial C$, there exists a supporting hyperplane to C at \mathbf{x}_0 .*

Proof: Two cases: $C^\circ \neq \emptyset$ and $C^\circ = \emptyset$.

Dual cones and generalized inequalities

- Dual cones

Let K be a cone. The set

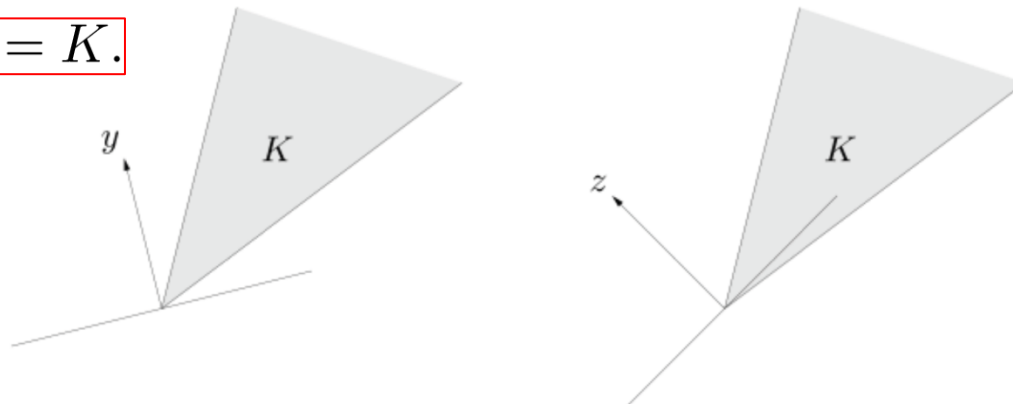
$$K^* = \{\mathbf{y} \mid \mathbf{x}^T \mathbf{y} \geq 0 \text{ for all } \mathbf{x} \in K\}$$

is called the dual cone of K .

K^* is a cone, and is always convex, even when the original cone K is not.

Geometrically, $\mathbf{y} \in K^*$ if and only if $-\mathbf{y}$ is the normal of a hyperplane that supports K at the origin.

Self-dual cone: $K^ = K$.*



Example: subspace, nonnegative orthant, positive semidefinite cone, norm cone

Dual cones and generalized inequalities

- Properties of dual cones
- K^* is closed and convex.
- $K_1 \subseteq K_2$ implies $K_2^* \subseteq K_1^*$.
- If K has nonempty interior, then K^* is pointed.
- If the closure of K is pointed then K^* has nonempty interior.
- K^{**} is the closure of the convex hull of K . (Hence if K is convex and closed, $K^{**} = K$.)

These properties show that if K is a proper cone, then so is its dual K^* , and moreover, that $K^{**} = K$.

Dual cones and generalized inequalities

- Dual generalized inequalities

Suppose that the convex cone K is proper, so it induces a generalized inequality \preceq_K . Then its dual cone K^* is also proper, and therefore induces a generalized inequality. We refer to the generalized inequality \preceq_{K^*} as the dual of the generalized inequality \preceq_K . Some important properties relating a generalized inequality and its dual are:

- $\mathbf{x} \preceq_K \mathbf{y}$ if and only if $\boldsymbol{\lambda}^T \mathbf{x} \leq \boldsymbol{\lambda}^T \mathbf{y}$ for all $\boldsymbol{\lambda} \succeq_{K^*} \mathbf{0}$.
- $\mathbf{x} \prec_K \mathbf{y}$ if and only if $\boldsymbol{\lambda}^T \mathbf{x} < \boldsymbol{\lambda}^T \mathbf{y}$ for all $\boldsymbol{\lambda} \succeq_{K^*} \mathbf{0}, \boldsymbol{\lambda} \neq \mathbf{0}$.

Since $K = K^{**}$, the dual generalized inequality associated with \preceq_{K^*} is \preceq_K , so these properties hold if the generalized inequality and its dual are swapped. As a specific example, we have $\boldsymbol{\lambda} \preceq_{K^*} \boldsymbol{\mu}$ if and only if $\boldsymbol{\lambda}^T \mathbf{x} \leq \boldsymbol{\mu}^T \mathbf{x}$ for all $\mathbf{x} \succeq_K \mathbf{0}$.

Dual cones and generalized inequalities

- Theorem of alternatives for linear strict generalized inequalities

Suppose $K \subseteq \mathbb{R}^m$ is a proper cone. Consider the strict generalized inequality

$$\mathbf{Ax} \prec_K \mathbf{b}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$. Then the inequality systems (1) and

$$\exists \boldsymbol{\lambda} \text{ s.t. } \boldsymbol{\lambda} \neq \mathbf{0}, \boldsymbol{\lambda} \succeq_{K^*} \mathbf{0}, \mathbf{A}^T \boldsymbol{\lambda} = \mathbf{0}, \boldsymbol{\lambda}^T \mathbf{b} \leq 0. \tag{2}$$

are alternatives.

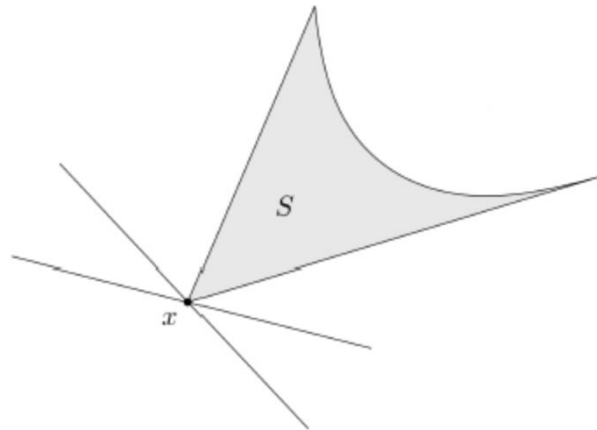
Dual cones and generalized inequalities

- Minimum and minimal elements via dual inequalities
 - Dual characterization of minimum element

\mathbf{x} is the minimum element of S , with respect to the generalized inequality \preceq_K , iff for all $\boldsymbol{\lambda} \succ_{K^*} \mathbf{0}$, \mathbf{x} is the unique minimizer of $\boldsymbol{\lambda}^T \mathbf{z}$ over $\mathbf{z} \in S$. Geometrically, this means that for any $\boldsymbol{\lambda} \succ_{K^*} \mathbf{0}$, the hyperplane

$$\{\mathbf{z} | \boldsymbol{\lambda}^T (\mathbf{z} - \mathbf{x}) = 0\}$$

is a strict supporting hyperplane to S at \mathbf{x} .

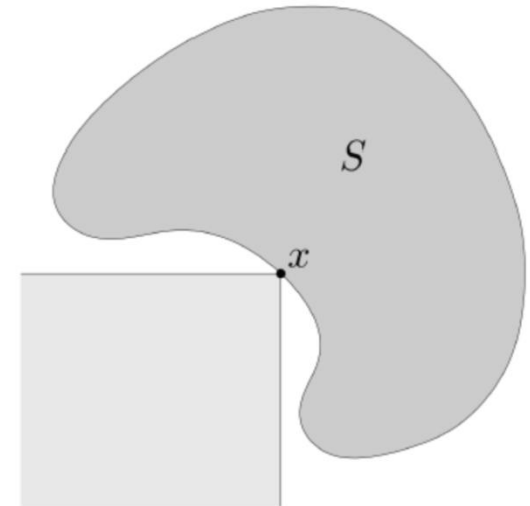
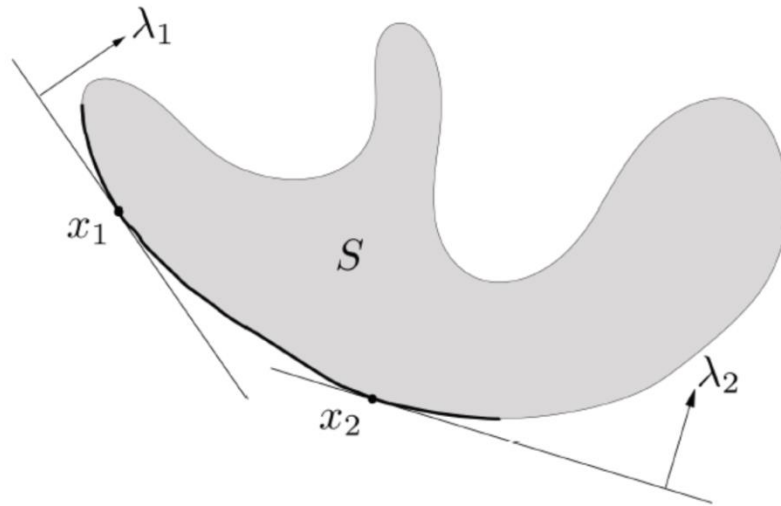


Dual cones and generalized inequalities

- Minimum and minimal elements via dual inequalities
 - Dual characterization of minimal element

only sufficient!

If $\lambda \succ_{K^*} 0$ and \mathbf{x} minimizes $\lambda^T \mathbf{z}$ over $\mathbf{z} \in S$, then \mathbf{x} is minimal.



Convexity matters!

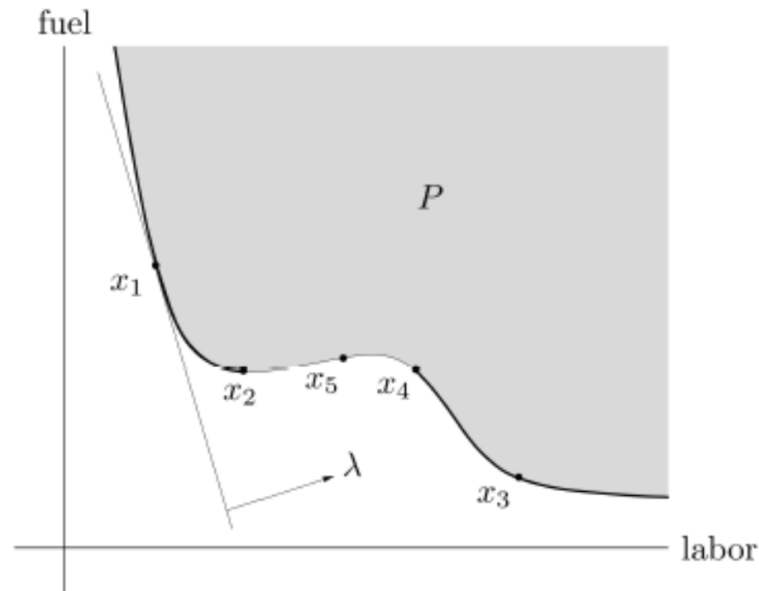
Dual cones and generalized inequalities

- Example: Pareto optimal production frontier

Minimize

$$\lambda^T \mathbf{x} = \lambda_1 \mathbf{x}_1 + \dots + \lambda_n \mathbf{x}_n$$

over the set P of production vectors, using any $\lambda > 0$.



Chapter 4: Convex Functions

- Basic properties and examples
- Operations that preserve convexity
- The conjugate function
- A little about nonconvex analysis

Basic properties and examples

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is *convex* if $\text{dom } f$ is a convex set and if for all $\mathbf{x}, \mathbf{y} \in \text{dom } f$, and θ with $0 \leq \theta \leq 1$, we have

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

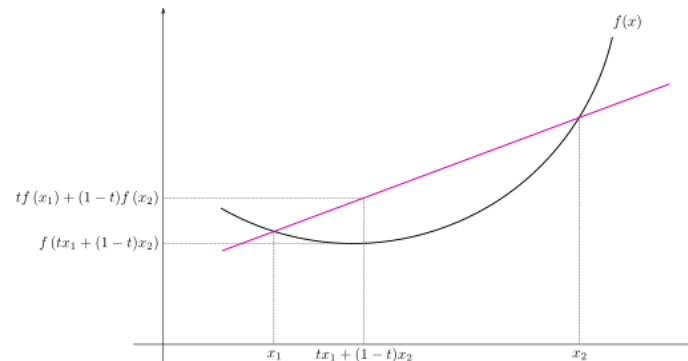
A function f is *strictly convex* if strict inequality holds in (1) whenever $\mathbf{x} \neq \mathbf{y}$ and $0 < \theta < 1$.

A function f is *strongly convex* if

$$f(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y}) - \frac{\theta(1 - \theta)\mu}{2} \|\mathbf{y} - \mathbf{x}\|^2, \quad \forall \theta \in [0, 1]. \quad (2)$$

modulus
μ

f is *concave*, *strictly concave*, *strongly concave* if $-f$ is convex, strictly convex, strongly convex. A function is both convex and concave iff it is an affine function.



Basic properties and examples

A convex function is continuous on the relative interior of its domain; it can have discontinuities only on its relative boundary.

Theorem 1 (Rademacher's Theorem). *A convex function is differentiable almost everywhere on the relative interior of its domain.*

- Extended-value extensions

If f is convex we define its *extended-value extension* $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ by

$$\tilde{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \mathbf{x} \in \text{dom } f \\ \infty, & \mathbf{x} \notin \text{dom } f. \end{cases}$$

We will use the same symbol to denote a convex function and its extension.

Example: Indicator function of a convex set

$$\begin{array}{ll} \min_{\mathbf{x}} f(\mathbf{x}), & \\ s.t. \mathbf{x} \in \mathcal{C}. & \end{array} \quad \longleftrightarrow \quad \min_{\mathbf{x}} f(\mathbf{x}) + \tilde{I}_{\mathcal{C}}(\mathbf{x}).$$

Basic properties and examples

- First-order conditions

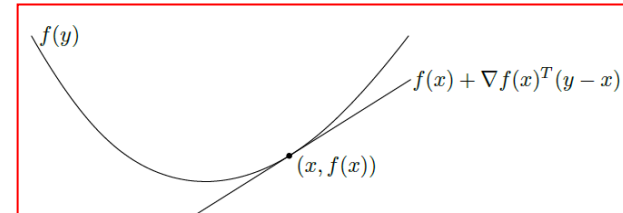
Suppose f is differentiable. Then f is convex iff $\text{dom } f$ is convex and

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \quad (1)$$

holds for all $\mathbf{x}, \mathbf{y} \in \text{dom } f$.

Proof. If f is convex, then $f((1 - \alpha)\mathbf{x} + \alpha\mathbf{y}) \leq (1 - \alpha)f(\mathbf{x}) + \alpha f(\mathbf{y})$, which can be rewritten as

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \frac{f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\alpha},$$



Letting $\alpha \rightarrow 0^+$, we have (1). If (1) holds, we have

$$f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}) \leq f(\mathbf{x}) - (1 - \alpha)\langle \nabla f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle,$$

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

Multiplying the first inequality with α and the second with $(1 - \alpha)$ and adding them together, we obtain $f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y})$.

Basic properties and examples

If $\nabla f(\mathbf{x}) = \mathbf{0}$, then for all $\mathbf{y} \in \text{dom } f$, $f(\mathbf{y}) \geq f(\mathbf{x})$, *i.e.*, \mathbf{x} is a global minimizer of f .

Strictly convex:

$$f(\mathbf{y}) > f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle, \quad \text{if } \mathbf{y} \neq \mathbf{x}. \quad (1)$$

Proof. $f(\mathbf{y}) > f(\mathbf{x}) + \frac{f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\alpha}$, $\forall \alpha \in (0, 1)$. For all $\alpha \in (0, 1)$ by the convexity we have $f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) - f(\mathbf{x}) \geq \alpha \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$. Thus $\langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle = \inf_{\alpha \in (0, 1)} \frac{f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\alpha}$. If there exists $\alpha \in (0, 1)$ such that $\frac{f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\alpha} > \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$, then (1) holds. Otherwise,

$$\frac{f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\alpha} = \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle, \quad \forall \alpha \in (0, 1).$$

So $f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}))$ is a linear function of $\alpha \in (0, 1)$ and f cannot be strictly convex.

Strongly convex: $f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{\mu}{2} \|\mathbf{y} - \mathbf{x}\|^2$.

Proof. Follow the proof of convexity.

Basic properties and examples

- Second-order conditions

Assume that f is twice differentiable. Then f is convex iff $\text{dom } f$ is convex and its Hessian is positive semidefinite: for all $\mathbf{x} \in \text{dom } f$,

$$\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}.$$

For a function on \mathbb{R} , this reduces to the simple condition $f''(x) \geq 0$ (and $\text{dom } f$ convex, *i.e.*, an interval), which means that the derivative is nondecreasing. The condition $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ can be interpreted geometrically as the requirement that the graph of the function have positive (upward) curvature at \mathbf{x} .

Strictly convex: $\nabla^2 f(\mathbf{x}) \succ \mathbf{0}$.

Strongly convex: $\nabla^2 f(\mathbf{x}) \succeq \mu \mathbf{I}$.

Basic properties and examples

- Examples
- *Exponential.* e^{ax} is convex on \mathbb{R} , for any $a \in \mathbb{R}$.
- *Powers.* x^a is convex on \mathbb{R}_{++} when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$.
- *Powers of absolute value.* $|x|^p$, for $p \geq 1$, is convex on \mathbb{R} .
- *Logarithm.* $\log x$ is concave on \mathbb{R}_{++} .
- *Negative entropy.* $x \log x$ (either on \mathbb{R}_{++} , or on \mathbb{R}_+ , defined as 0 for $x = 0$) is convex.

Basic properties and examples

- Examples
- *Norms.* Every norm on \mathbb{R}^n is convex.
- *Max function.* $f(\mathbf{x}) = \max\{x_1, \dots, x_n\}$ is convex on \mathbb{R}^n .
- $f(x, y) = x^2/y$ with $\text{dom } f = \mathbb{R} \times \mathbb{R}_{++} = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$ is convex.
- *Log-sum-exp.* $f(\mathbf{x}) = \log(e^{x_1} + \dots + e^{x_n})$ is convex on \mathbb{R}^n .
- *Geometric mean.* $f(\mathbf{x}) = (\prod_{i=1}^n x_i)^{1/n}$ is concave on $\text{dom } f = \mathbb{R}_{++}^n$.
- *Log-determinant.* $f(\mathbf{X}) = \log \det \mathbf{X}$ is concave on $\text{dom } f = \mathbb{S}_{++}^n$.

