Convex Sets (III) + Convex Functions (I)

Lecture 3, Nonlinear Programming

National Taiwan University

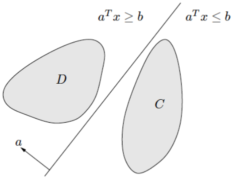
October 4, 2016

Table of contents

Separating Hyperplane Theorem

Separating Hyperplane

The hyperplane $\{x \mid a^T x = b\}$ is called a **separating hyperplane** for the sets C and D, or is said to **separate** the sets C and D if $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$.



Separating Hyperplane Theorem

Suppose C and D are two convex sets that do not intersect, i.e., $C \cap D = \phi$. Then there exist $a \neq 0$ and b such that the hyperplane $\{x \mid a^T x = b\}$ separates C and D.

Separating Hyperplane Theorem – Proof

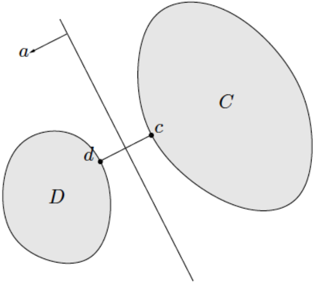
Proof of a special case

- Consider that C and D are both convex, closed, and bounded.
- Assume that the **Euclidean distance** between C and D, defined as

$$\textbf{dist}(C, D) = \inf\{||u - v||_2 \mid u \in C, v \in D\},$$

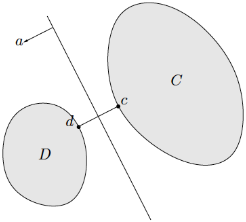
  is positive.
- Since C and D are both closed and bounded, there exist $c \in C$ and $d \in D$ such that

$$||c - d||_2 = \textbf{dist}(C, D).$$



Separating Hyperplane Theorem – Proof

Proof of a special case



- Let

$$a = d - c, \quad b = \frac{||d||_2^2 - ||c||_2^2}{2}.$$

- Then, it can be shown that the affine function

$$f(x) = a^T x - b = (d - c)^T \left(x - \frac{d + c}{2} \right)$$

is nonpositive on C and nonnegative on D.

Example – A Convex Set and An Affine Set

- Suppose C is convex and D is affine, i.e., $D = \{Fu + g | u \in \mathbf{R}^m\}$, where $F \in \mathbf{R}^{n \times m}$.
- Suppose C and D are disjoint, so by the separating hyperplane theorem there are $a \neq 0$ and b such that $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$.
- $\because a^T x \geq b$ for all $x \in D$, $\therefore a^T Fu \geq b - a^T g$ for all $u \in \mathbf{R}^m$.
- But a linear function is bounded below on $\mathbf{R}^m$ only when it is zero, so we conclude $a^T F = 0$ (and hence, $b \leq a^T g$).
- Thus we conclude that there exists $a \neq 0$ such that $F^T a = 0$ and $a^T x \leq a^T g$ for all $x \in C$.

Strict Separation of Convex Sets

Strict separation of convex sets

For two sets $C, D \subseteq \mathbf{R}^n$, if there exists $a \in \mathbf{R}^n$, $b \in \mathbf{R}$ such that

$$a^T x < b \ \forall x \in C \text{ and } a^T x > b \ \forall x \in D,$$

then C and D are said to be **strictly separable**, and the hyperplane $\{x \mid a^T x = b\}$ is called **strict separation** of C and D.

- Remark: The separating hyperplane theorem only dictates that two convex sets that are disjoint to be separated by a hyperplane. A strict separation is not guaranteed (even when the sets are closed).

Example – A Point and A Closed Convex Set

- Let C be a closed convex set and $x_0 \notin C$. Then there exists a hyperplane that strictly separates $\{x_0\}$ from C.
- Proof idea:
  - The two sets C and $B(x_0, \epsilon)$ do not intersect for some $\epsilon > 0$.
  - Apply the separating hyperplane theorem on C and $B(x_0, \epsilon)$ (getting a^T and b), and let $f(x)$.
  - The affine function

    $$f(x) = a^T x - b - \epsilon||a||_2/2$$

    strictly separates C and $\{x_0\}$.
- Corollary: a closed convex set is the intersection of all halfspaces that contain it. (*Hint: proof by contradiction*)

Converse of Separating Hyperplane Theorems

- Question: If there exists a hyperplane that separates convex sets C and D, does this imply C and D are disjoint?
  - (No. Consider $C = D = \{0\} \subseteq \mathbf{R}$.)
- Suppose C and D are convex sets, with C open, and there exists an affine function f that is nonpositive on C and nonnegative on D. Then C and D are disjoint.
  - *Hint: f is negative on C.*

Theorem

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

Theorem of alternatives for strict linear inequalities

Theorem of alternatives for strict linear inequalities

Let $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$. The inequalities

$$Ax \prec b$$

are infeasible if and only if there exists $\lambda \in \mathbf{R}^m$ such that

$$\lambda \neq 0, \quad \lambda \succeq 0, \quad A^T\lambda = 0, \quad \lambda^T b \leq 0.$$

- Proof idea: consider the open convex set

$$D = \mathbf{R}^m_{++} = \{y \in \mathbf{R}^m \mid y \succ 0\}$$

and the affine set (hence convex)

$$C = \{b - Ax \mid x \in \mathbf{R}^n\}.$$

Supporting Hyperplanes

Supporting hyperplanes

Suppose $C \subseteq \mathbf{R}^n$, and x_0 is a point in its boundary **bd** C, i.e.,

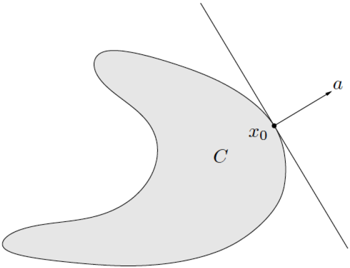
$$x_0 \in \mathbf{bd}\ C = \mathbf{cl}\ C \backslash \mathbf{int}\ C.$$

If $a \neq 0$ satisfies $a^T x \leq a^T x_0$ for all $x \in C$, then the hyperplane $\{x | a^T x = a^T x_0\}$ is called a supporting hyperplane to C at the point x_0.

- This is equivalent to the statement that $\{x_0\}$ and C are separated by the hyperplane $\{x | a^T x = a^T x_0\}$.
- The hyperplane is tangent to C at x_0, and the halfspace $\{x | a^T x \leq a^T x_0\}$ contains C.

Supporting Hyperplanes

- This is equivalent to the statement that $\{x_0\}$ and C are separated by the hyperplane $\{x|a^Tx = a^Tx_0\}$.
- The hyperplane is tangent to C at x_0, and the halfspace $\{x|a^Tx \le a^Tx_0\}$ contains C.



Supporting Hyperplane Theorem

Supporting Hyperplane Theorem

For any nonempty convex set C, and any $x_0 \in \textbf{bd } C$, there exists a supporting hyperplane to C at x_0.

Proof: Use the separating hyperplane theorem.

- If $\textbf{int } C \neq \phi$: then by applying the separating hyperplane theorem on $\{x_0\}$, the statement is proved.
- If $\textbf{int } C = \phi$, then C lies in an affine set of dimension less than n. Then any hyperplane that contains this affine set contains both C and x_0 and therefore is a supporting hyperplane.

(Partial) Converse of the Supporting Hyperplane Theorem

Converse of the Supporting Hyperplane Theorem

If a set C is closed, has nonempty interior, and has a supporting hyperplane at any $x_0 \in \mathbf{bd}\ C$, then C is convex.

Dual Cones

Dual Cones

Let K be a cone. The set

$$K^* = \left\{ y \;\middle|\; x^T y \geq 0 \text{ for all } x \in K \right\}$$

is called the **dual cone** of K.

Basic Properties of Dual Cones

- K^* is a cone.
- K^* is convex (even when K is not convex).

Dual Cones

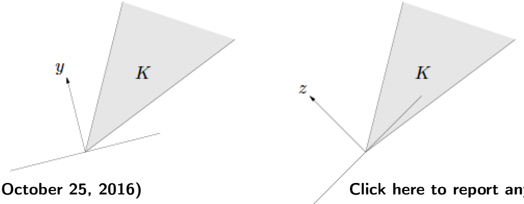
Dual Cones

Let K be a cone. The set

$$K^* = \left\{ y \;\middle|\; x^T y \geq 0 \text{ for all } x \in K \right\}$$

is called the **dual cone** of K.

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



Dual Cones – Examples

Dual Cones

Let K be a cone. The set

$$K^* = \left\{ y \;\middle|\; x^T y \geq 0 \text{ for all } x \in K \right\}$$

is called the **dual cone** of K.

Subspace

The dual cone of a subspace $V \subseteq \mathbf{R}^n$ (which is a cone) is its orthogonal complement

$$V^\perp = \left\{ y \;\middle|\; y^T v = 0 \quad \forall v \in V \right\}.$$

Dual Cones – Examples

Dual Cones

Let K be a cone. The set

$$K^* = \left\{ y \;\middle|\; x^T y \geq 0 \text{ for all } x \in K \right\}$$

is called the **dual cone** of K.

Nonnegative orthant

The cone $\mathbf{R}^n_+$ is its own dual:

$$y^T x \geq 0, \ \forall x \succeq 0 \Leftrightarrow y \succeq 0.$$

We call such a cone **self-dual**.

Dual of a Positive Semidefinite Cone

Positive semidefinite cone

The positive semidefinite cone $\mathbf{S}_+^n$ is self-dual, i.e., for $X, Y \in \mathbf{S}^n$,

$$\mathbf{tr}\,(XY) \geq 0 \text{ for all } X \succeq 0 \Leftrightarrow Y \succeq 0.$$

Here, we used the standard inner product $\mathbf{tr}\,(XY) = \sum_{i,j=1}^n X_{ij} Y_{ij}$ on the set of symmetric $n \times n$ matrices $\mathbf{S}^n$.

Proof:

- If $Y \notin \mathbf{S}_+^n$, then $\exists q \in \mathbf{R}^n$ s.t. $q^T Y q < 0$.
  Let $X = qq^T$, then $\mathbf{tr}\,(XY) = \mathbf{tr}\,(qq^T Y) = \mathbf{tr}\,(q^T Y q) < 0$.
- If $Y \in \mathbf{S}_+^n$, then for any $X \in \mathbf{S}_+^n$, i.e., $X \succeq 0$, X can be expressed as $X = \sum_{i=1}^n \lambda_i q_i q_i^T$, with $\lambda_i \geq 0$ and $q_i \neq 0$.
  Then,
  $\mathbf{tr}\,(XY) = \sum_{i=1}^n \lambda_i \mathbf{tr}\,(q_i q_i^T Y) = \sum_{i=1}^n \lambda_i \mathbf{tr}\,(q_i^T Y q_i) \geq 0.$

Properties of Dual Cones

Properties of Dual Cones

Dual cones satisfy the following properties:

- 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.
- If K is convex and closed, then $K^{**} = K$.

Moreover, if K is a proper cone, then

- K^* is also a proper cone.
- $K^{**} = K$.

Dual generalized inequalities

Dual generalized inequalities

If K is a proper cone which induces a generalized inequality $\preceq_K$. Then we refer to the generalized inequality $\preceq_{K^*}$ as the **dual** of the generalized inequality $\preceq_K$. (Note that K^* is also a proper cone).

Properties of dual generalized inequalities

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

Theorem of alternatives for strict generalized inequalities

Theorem of alternatives for linear strict generalized inequalities

Suppose $K \subseteq \mathbf{R}^m$ is a proper cone. The strict generalized inequality

$$Ax \prec_K b,$$

where $x \in \mathbf{R}^n$, is infeasible iff there exists $\lambda \in \mathbf{R}^m$ s.t.

$$\lambda \neq 0, \quad \lambda \succeq_{K^*} 0, \quad A^T\lambda = 0, \quad \lambda^T b \leq 0.$$

Proof idea:

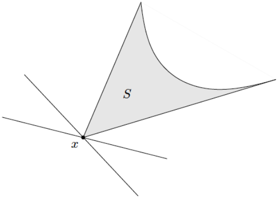
- Apply the separating hyperplane theorem on $\{b - Ax \mid x \in \mathbf{R}^n\}$ and $\mathbf{int}\ K$.
- Converse: Proof by contradiction.

Dual characterization of minimum element

Dual characterization of minimum element

Consider a set $S \subseteq \mathbf{R}^n$, not necessarily convex. Then, x is the minimum element of S, with respect to the generalized inequality $\preceq_K$, if and only if for all $\lambda \succ_{K^*} 0$, x is the unique minimizer of $\lambda^T z$ over $z \in S$.

- This means that for any $\lambda \succ_{K^*} 0$, the hyperplane $\{z \mid \lambda^T(z - x) = 0\}$ is a **strict supporting hyperplane** to S at x (i.e., the hyperplane intersects S only at the point x).



Dual characterization of minimum element

Dual characterization of minimum element

Consider a set $S \subseteq \mathbf{R}^n$, not necessarily convex. Then, x is the minimum element of S, with respect to the generalized inequality $\preceq_K$, if and only if for all $\lambda \succ_{K^*} 0$, x is the unique minimizer of $\lambda^T z$ over $z \in S$.

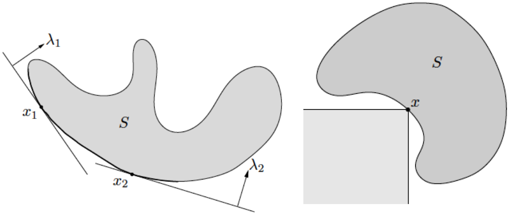
Proof:

Dual characterization of minimal elements

Dual Characterization of minimal elements

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

- The converse is not true: if x is minimal in S, there does not necessarily exist $\lambda \succ_{K^*} 0$ that minimizes $\lambda^T z$ over $z \in S$. (Example: if S is not convex).



Dual characterization of minimal elements

Dual Characterization of minimal elements in convex sets

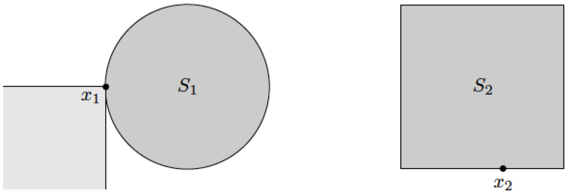
Suppose the set $S \subseteq \mathbf{R}^n$ is convex. Then, for any minimal element x in S there exists a nonzero $\lambda \succeq_{K^*} 0$ such that x minimizes $\lambda^T z$ over $z \in S$.

Proof: Suppose x is minimal: $((x-K)\setminus\{x\}) \cap S = \phi$. By applying the separating hyperplane theorem to the convex sets $(x-K)\setminus\{x\}$ and S, we have $\exists \lambda \neq 0$ and μ such that $\lambda^T(x-y) \leq \mu$ for all $y \in K$ and $\lambda^T z \geq \mu$ for all $z \in S$. Note that $\lambda^T x = \mu$ since $x \in S$ and $x \in x - K$. So $\lambda^T y \geq 0$ for all $y \in K$ and therefore $\lambda \in K^*$ (i.e., $\lambda \succeq_{K^*} 0$). Finally, $\lambda^T z \geq \mu = \lambda^T x$ for all $z \in S$ suggests that x minimizes $\lambda^T z$ over $z \in S$.

Question: Must there exist a $\lambda \succ_{K^*} 0$ such that x minimizes $\lambda^T z$ over $z \in S$?

Dual characterization of minimal elements

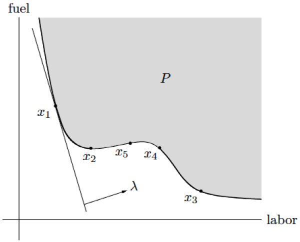
- Suppose the set $S \subseteq \mathbf{R}^n$ is convex. There does not necessarily exist a nonzero $\lambda \succ_{K^*} 0$ such that a minimal point in S, x, minimizes $\lambda^T z$ over $z \in S$.



Pareto optimal production frontier

Pareto optimal production frontier

- Consider a production set $P \subseteq \mathbf{R}^n$ consisting of vectors that correspond to some production method: $x \in P$ corresponds to a method that consumes x_i units of resource i, $i = 1, ..., n$.
- Production methods with resource vectors that are minimal elements of P, with respect to component-wise inequality, are called Pareto optimal or efficient.
- The set of minimal elements of P is called the efficient production frontier.



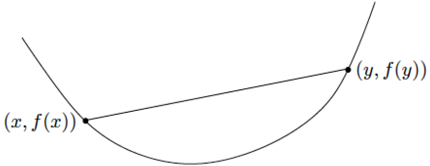
Definitions of Convex Functions

Convex functions

A function $f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\ f$ is a convex set and if for all $x, y \in \mathbf{dom}\ f$ and for all $\theta \in [0, 1]$,

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y).$$

- The line segment between $(x, f(x))$ and $(y, f(y))$, which is the chord from x to y, lies above the graph of f.



Definitions of Convex Functions

Convex functions

A function $f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\ f$ is a convex set and if for all $x, y \in \mathbf{dom}\ f$ and for all $\theta \in [0, 1]$,

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

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

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y).$$

- We say f is **concave** if $-f$ is convex, and **strictly concave** if $-f$ is strictly convex.

Affine Functions

Affine Functions

For an affine function we always have equality in (1), i.e.,

$$f(\theta x + (1-\theta)y) = \theta f(x) + (1-\theta)f(y),$$

so all affine functions are both convex and concave.

- Conversely, any function that is convex and concave is affine.

Convexity

- A function is convex if and only if it is convex when restricted to any line that intersects its domain:
- That is, f is convex if and only if $\forall x \in \textbf{dom } f, v \in \mathbf{R}^n$, the function $g(t) = f(x + tv)$ is convex on $\{t \mid x + tv \in \textbf{dom } f\}$.
- A convex function is continuous on the relative interior of its domain; it can have discontinuities only on its relative boundary.

Extended-Value Extensions

Extended-Value Extensions

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

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

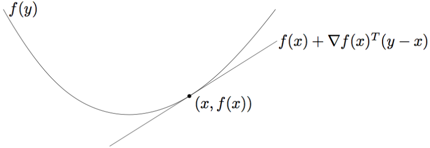
First-Order Conditions

First-Order Conditions

Suppose f is differentiable (implying that **dom** f is open). Then f is convex if and only if **dom** f is convex and

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

- Observation: the first-order Taylor approximation is a global underestimator of the function.
- Conversely, if the first-order Taylor approximation of a function is always a global underestimator of the function, then the function is convex.



First-Order Conditions

- A convex function f satisfies

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

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

- This shows that from local information about a convex function (i.e., $f(x), \nabla f(x)$), we can derive global information (i.e., a global underestimator).
- Example: if $\nabla f(x) = 0$, then for all $y \in \mathbf{dom}\ f$, $f(y) \geq f(x)$. (x is the global minimizer of f.)

First-Order Conditions – Strict Convexity, Concavity

First-Order Conditions for strict convexity

f is strictly convex if and only if **dom** f is convex and for $x, y \in$ **dom** $f, x \neq y$, we have

$$f(y) > f(x) + \nabla f(x)^T (y - x).$$

First-Order Conditions for (strict) concavity

f is concave if and only if **dom** f is convex and for $x, y \in$ **dom** f, we have

$$f(y) \leq f(x) + \nabla f(x)^T (y - x).$$

f is strictly concave if and only if **dom** f is convex and for $x, y \in$ **dom** $f, x \neq y$, we have

$$f(y) < f(x) + \nabla f(x)^T (y - x).$$

Proof of First-Order Conditions

Proof ideas:

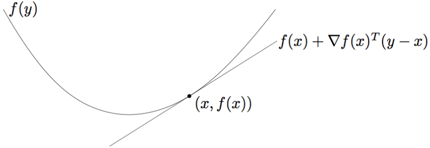
- Consider the special case $n = 1$ first.
  - Then we only need to prove that f is convex if and only if

$$f(y) \geq f(x) + f'(x)(y - x).$$

- For the general case $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\ f$ convex, consider the line passing by any two points $x, y \in \mathbf{dom}\ f, x \neq y$, and define a function $g : \mathbf{R} \to \mathbf{R}$ with $g(t) = f(ty + (1 - t)x)$.

Second-Order Conditions

- Assume that $f : \mathbf{R} \to \mathbf{R}$ is twice differentiable with **dom** $f = \mathbf{R}$, then it is convex if and only if its second derivative is nonnegative.



Second-Order Conditions

- Assume that f is twice differentiable, that is, its **Hessian** or second derivative $\nabla^2 f$ exists at each point in **dom** f (open).

Second-Order Conditions

Then, f is convex if and only if **dom** f is convex and its Hessian is positive semidefinite:

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \mathbf{dom}\ f.$$

- For a function on $\mathbf{R}$, this means $f''(x) \geq 0$, and **dom** f is convex.

Second-Order Conditions – Strict Convexity, Concavity

Second-Order Conditions for Concavity

A function f is concave if and only if $\textbf{dom}\ f$ is convex and $\nabla^2 f(x) \preceq 0$ for all $x \in \textbf{dom}\ f$.

Second-Order Conditions for Strict Convexity

If $\nabla^2 f(x) \succ 0$ for all $x \in \textbf{dom}\ f$, then f is strictly convex.

- If f is strictly convex, do we have $\nabla^2 f(x) \succ 0$? (e.g., think $f(x) = x^4$)
- Is $f : \mathbf{R} \to \mathbf{R}, f(x) = 1/x^2$ a convex function? Why?

Example – Quadratic Functions

- Consider the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$, with $\textbf{dom}\ f = \mathbf{R}^n$, given by

$$f(x) = (1/2)x^T Px + q^T x + r,$$

with $P \in \mathbf{S}^n, q \in \mathbf{R}^n$, and $r \in \mathbf{R}$.
  - Note that $\nabla^2 f(x) = P$.
- The function f is convex if and only if $P \succeq 0$.
- The function f is concave if and only if $P \preceq 0$.
- The function f is strictly convex if and only if $P \succ 0$.
- The function f is strictly concave if and only if $P \prec 0$.

Example Convex functions on $\mathbf{R}$

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

Example Convex Functions on $\mathbf{R}^n$

- Norms. Every norm on $\mathbf{R}^n$ is convex.

Norms

A function $f : \mathbf{R}^n \to \mathbf{R}$ (with $\mathbf{dom}\ f = \mathbf{R}^n$) is called a **norm** if for any $x, y \in \mathbf{R}^n$, $t \in \mathbf{R}$, we have

- $f(x) \geq 0$ (f is nonnegative).
- $f(x) = 0$ only if $x = 0$ (f is definite).
- $f(tx) = |t| f(x)$ (f is homogeneous).
- $f(x + y) \leq f(x) + f(y)$ (f satisfies the triangle inequality).

Example Convex Functions on $\mathbf{R}^n$

- Norms. Every norm on $\mathbf{R}^n$ is convex.
- Max function. $f(x) = \max\{x_1, \ldots, x_n\}$ is convex on $\mathbf{R}^n$.
- Quadratic-over-linear function. The function $f(x, y) = x^2/y$, with $\mathbf{dom}\ f = \mathbf{R} \times \mathbf{R}_{++} = \left\{(x, y) \in \mathbf{R}^2 \mid y > 0\right\}$, is convex.
- Log-sum-exp. The function $f(x) = \log(e^{x_1} + \ldots + e^{x_n})$ is convex on $\mathbf{R}^n$.
  - Note that $\max\{x_1, \ldots, x_n\} \leq f(x) \leq \max\{x_1, \ldots, x_n\} + \log n$.
- Geometric mean. The geometric mean $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\mathbf{dom}\ f = \mathbf{R}^n_{++}$.
- Log-determinant. The function $f(X) = \log\det X$ is concave on $\mathbf{dom}\ f = \mathbf{S}^n_{++}$.

More on Norms

Norms

A function $f : \mathbf{R}^n \to \mathbf{R}$ (with $\mathbf{dom}\ f = \mathbf{R}^n$) is called a **norm** if for any $x, y \in \mathbf{R}^n$, $t \in \mathbf{R}$, we have

- $f(x) \geq 0$ (f is nonnegative).
- $f(x) = 0$ only if $x = 0$ (f is definite).
- $f(tx) = |t| f(x)$ (f is homogeneous).
- $f(x + y) \leq f(x) + f(y)$ (f satisfies the triangle inequality).

l_p-norm

Let $p \geq 1$. Then the **l_p-norm** is defined as

$$||x||_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$$

Question: When $p < 1$, is $||x||_p$ still a norm?

Examples of l_p-norm

- When $p = 2$, the l_2-norm is actually the Euclidean norm:

$$||x||_2 = (x_1^2 + \cdots + x_n^2)^{1/2}.$$

- When $p = 1$, the l_1-norm is the sum-absolute-value:

$$||x||_1 = |x_1| + \cdots + |x_n|.$$

- When $p \to \infty$, the l_∞-norm is:

$$||x||_\infty \triangleq \lim_{p \to \infty} ||x||_p = \lim(|x_1|^p + \cdots + |x_n|^p)^{1/p}.$$

It can be shown that $||x||_\infty = \max\{|x_1|, ..., |x_n|\}$.

Other Examples of Norms

- For $P \in \mathbf{S}^n_{++}$, the **P-quadratic norm** is defined as

$$||x||_P = (x^T P x)^{1/2} = ||P^{1/2} x||_2.$$

  - The **unit ball** of a quadratic norm,

$$\{x \in \mathbf{R}^n \mid ||x||_P \leq 1\},$$

    is an ellipsoid.

- The **Frobenius norm**, defined on $\mathbf{R}^{m \times n}$, is

$$||X||_F = (\mathbf{tr}\ X^T X)^{1/2} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}^2 \right)^{1/2}.$$

Norms and Max function

- If $f : \mathbf{R}^n \to \mathbf{R}$ is a norm, and $0 \leq \theta \leq 1$, then

$$f(\theta x + (1-\theta)y) \leq f(\theta x) + f((1-\theta)y) = \theta f(x) + (1-\theta)f(y)$$

since f satisfies the triangle inequality and f is homogeneous.

  - Therefore any norm is convex.

- The function $f(x) = \max_i x_i$ is convex since

$$\begin{aligned} \max_i(\theta x_i + (1-\theta)y_i) &\leq \max_i \theta x_i + \max_i (1-\theta)y_i \\ &= \theta \max_i x_i + (1-\theta)\max_i y_i. \end{aligned}$$

  - In addition, $f(|x|) = \max_i |x_i|$ is a norm.

Quadratic-Over-Linear Function

- The quadratic-over-linear function $f : \mathbf{R}^2 \to \mathbf{R}, \mathbf{dom}\ f = \mathbf{R} \times \mathbf{R}_{++}$, $f(x, y) = x^2/y$, is convex since:

$$\nabla^2 f(x, y) = \frac{2}{y^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0.$$

Log-Sum-Exp

- The log-sum-exp function $f(x) = \log(e^{x_1} + \ldots + e^{x_n})$ is convex on $\mathbf{R}^n$ since

$$\nabla^2 f(x) = \frac{1}{(\mathbf{1}^T z)^2}\left((\mathbf{1}^T z)\mathbf{diag}\,(z) - zz^T\right),$$

where $z = (e^{x_1}, \ldots, e^{x_n})$, and

- for all v,

$$v^T \nabla^2 f(x) v = \frac{1}{(\mathbf{1}^T z)^2}\left(\left(\sum_{i=1}^n z_i\right)\left(\sum_{i=1}^n v_i^2 z_i\right) - \left(\sum_{i=1}^n v_i z_i\right)^2\right) \geq 0.$$

Geometric mean

- The **geometric mean** function $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\textbf{dom}\ f = \mathbf{R}^n_{++}$ since its Hessian $\nabla^2 f(x)$ can be shown to be negative semidefinite.
- Note that

$$\frac{\partial f(x)}{\partial x_k} = \frac{1}{n}\frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{\prod_{i=1}^n x_i} \cdot \prod_{i=1, i\neq k}^{n} x_i = \frac{1}{n}\frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{x_k}$$

and

$$\frac{\partial^2 f(x)}{\partial x_k^2} = -(n-1)\frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{n^2 x_k^2}, \frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{n^2 x_k x_l}$$

Geometric mean

- The geometric mean function $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\mathbf{dom}\ f = \mathbf{R}^n_{++}$ since its Hessian $\nabla^2 f(x)$ can be shown to be negative semidefinite.
- So,

$$\nabla^2 f(x) = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n\ \mathbf{diag}\ (1/x_1^2, ..., 1/x_n^2) - qq^T\right)$$

where $q_i = 1/x_i$

- For any $v \in \mathbf{R}^n$, we have

$$v^T \nabla^2 f(x) v = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n\sum_{i=1}^n v_i^2/x_i^2 - \left(\sum_{i=1}^n v_i/x_i\right)^2\right) \leq 0$$

Log-Determinant

- The function $f : \mathbf{S}^n \to \mathbf{R}$, $f(X) = \log\det X$, with **dom** $= \mathbf{S}^n_{++}$ is concave.
- Proof idea: consider an arbitrary line in $\mathbf{S}^n$ (that passes through some point in $\mathbf{S}^n_{++}$) given by $X = Z + tV$, where $Z \in \mathbf{S}^n_{++}$, $V \in \mathbf{S}^n$, and define $g(t) = f(Z + tV)$, **dom** $g = \{t \mid Z + tV \succ 0\}$.
- Then it can be shown that

$$g(t) = \sum_{i=1}^{n} \log(1 + t\lambda_i) + \log\det Z$$

where λ_i are the eigenvalues of $Z^{-1/2} V Z^{-1/2}$.
- So,

$$g''(t) = -\sum_{i=1}^{n} \frac{\lambda_i^2}{(1 + t\lambda_i)^2} \leq 0.$$