Convex Functions (I)

Lecture 3, Convex Optimization (Part b)

National Taiwan University

March 10, 2023

Table of contents

Open Sets and Closed Sets

Open Sets

A set $C \subseteq \mathbf{R}^n$ is said to be **open** if every element in C is an interior point.

Closed Sets

A set $C \subseteq \mathbf{R}^n$ is said to be **closed** if the complement of C, (i.e., $\mathbf{R}^n \setminus C$), is open.

- An alternative definition for closed sets is: a set $C \subseteq \mathbf{R}^n$ is said to be closed if every convergent sequence in C converges to a point in C.

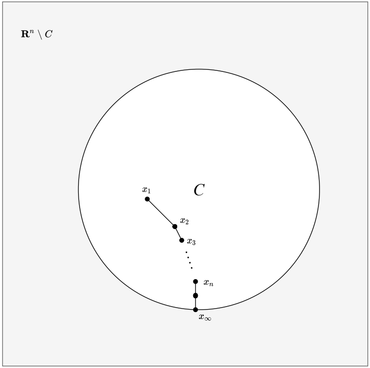
Examples of Open and Closed Sets

- The set $(1,2) = \{x \in \mathbf{R} \mid 1 < x < 2\}$ is open.
- The set $[1,2] = \{x \in \mathbf{R} \mid 1 \leq x \leq 2\}$ is closed.
- The empty set $\varnothing$ is open. It is also closed.
- The set $\mathbf{R}^n$ is open. It is also closed.

Properties of Closed Sets (1/3)

Properties of Closed Sets

A set $C \subseteq \mathbf{R}^n$ is closed if and only if the limit point of every convergent sequence is in C.

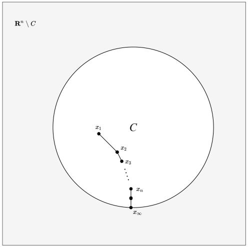


Proof:

Properties of Closed Sets (2/3)

Properties of Closed Sets

A set $C \subseteq \mathbf{R}^n$ is closed if and only if the limit point of every convergent sequence is in C.



Proof: “only if”: Suppose $\{x_k\}, k = 1, 2, ...$ is a convergent sequence in C. Then, there exists uniquely a limit point $x_\infty \in \mathbf{R}^n$, such that $\forall \epsilon > 0, \exists n_0 \in \mathbb{N}$ s.t. $\forall n_1 > n_0$,

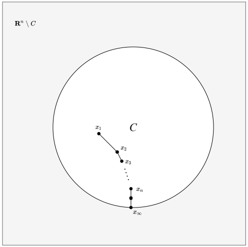
$$||x_{n_1} - x_\infty||_2 < \epsilon. \tag{1}$$

Suppose x_∞ is NOT in C. Then $x_\infty \in \mathbf{R}^n \setminus C$ instead. Then x_∞ must be an interior point in $\mathbf{R}^n \setminus C$ (since it is open). And this means $\exists \epsilon > 0$ such that the ball $B(x_\infty, \epsilon) \subseteq \mathbf{R}^n \setminus C$. But this is contradictory to (1).

Properties of Closed Sets (3/3)

Properties of Closed Sets

A set $C \subseteq \mathbf{R}^n$ is closed if and only if the limit point of every convergent sequence is in C.



Proof: "if": Suppose C is not closed. Then $\mathbf{R}^n \setminus C$ is not open, and $\exists x_0 \in \mathbf{R}^n \setminus C$ that is not in $\mathbf{int}\,(\mathbf{R}^n \setminus C)$. Thus, $\forall \epsilon > 0$, $B(x_0, \epsilon) \not\subseteq \mathbf{R}^n \setminus C$, i.e., $B(x_0, \epsilon) \bigcap C \neq \varnothing, \forall \epsilon > 0$. Construct a sequence $\{x_n\}$ and let the nth point $x_n \in C$ be chosen as any point in $B(x_0, \frac{1}{n}) \bigcap C$. Then, $\lim_{n\to\infty} x_n = x_0 \notin C$.

Closure

Closure

The closure of a set $C \subseteq \mathbf{R}^n$ is defined as

$$\mathbf{cl}\ C = \mathbf{R}^n \setminus \mathbf{int}\ (\mathbf{R}^n \setminus C)$$

- A point $x \in \mathbf{cl}\ C$ if $\forall \epsilon > 0$, $\exists y \in C$ such that $||x - y||_2 \leq \epsilon$.
- It can be shown that the closure of C is the set of all the limit points of convergent sequences in C.

Boundary

Boundary

The boundary of the set $C \subseteq \mathbf{R}^n$ is defined as

$$\mathbf{bd}\ C = \mathbf{cl}\ C \setminus \mathbf{int}\ C.$$

- $\forall x \in \mathbf{bd}\ C, \forall \epsilon > 0$, $\exists y \in C$ and $z \notin C$ s.t.

$$||y - x||_2 \leq \epsilon, \quad ||z - x||_2 \leq \epsilon.$$

- A set $C \subseteq \mathbf{R}^n$ is closed if it contains its boundary: $\mathbf{bd}\ C \subseteq C$.
- A set $C \subseteq \mathbf{R}^n$ is open if it contains no boundary points: $C \bigcap \mathbf{bd}\ C = \varnothing$.

Supremum

Upper bound

Suppose $C \subseteq \mathbf{R}$. A number a is an **upper bound** on C if for each $x \in C, x \leq a$.

The set of upper bounds on a set C is either

1. empty (in which case we say C is **unbounded above**),
2. all of $\mathbf{R}$ (only when $C = \varnothing$), or
3. a closed infinite interval $[b, \infty)$.

Supremum

The number b is called the **least upper bound** or **supremum** of the set C, and is denoted by $\sup C$.

- We take $\sup \varnothing = -\infty$, and $\sup C = \infty$ if C is unbounded above.
- When $\sup C \in C$, we say the supremum of C is **attained** or **achieved**.

Infimum

Lower Bound and Infimum

A number a is a **lower bound** on $C \subseteq \mathbf{R}$ if for each $x \in C, a \leq x$. The **infimum** (or **greatest lower bound**) of a set $C \subseteq \mathbf{R}$ is defined as $\inf C = -\sup(-C)$. When C is finite, the infimum is the minimum of its elements. We take $\inf \varnothing = \infty$, and $\inf C = -\infty$ if C is unbounded below, i.e., has no lower bound.

Derivative

Derivative

Suppose $f : \mathbf{R}^n \to \mathbf{R}^m$ and $x \in \textbf{int dom } f$. The function f is **differentiable** at x if there exists a matrix $Df(x) \in \mathbf{R}^{m\times n}$ that satisfies

$$\lim_{z\in \textbf{dom } f, z\neq x, z\to x} \frac{||f(z) - f(x) - Df(x)(z-x)||_2}{||z-x||_2} = 0.$$

The matrix $Df(x)$ is called the **derivative** (or **Jacobian**) of f at x.

- $Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, \quad i = 1, \cdots, m, \quad j = 1, \cdots, n,$ where $f_i : \mathbf{R}^n \to \mathbf{R}$ is the function such that $f_i(x)$ is the ith component of $f(x)$ for all $x \in \mathbf{R}^n$.

Gradient

Gradient

If f is real-valued (i.e., $f : \mathbf{R}^n \to \mathbf{R}$), the derivative $Df(x)$ is a $1 \times n$ matrix, i.e., it is a row vector. Its transpose, as a column vector in $\mathbf{R}^n$, is called the **gradient** of the function:

$$\nabla f(x) = Df(x)^T = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}.$$

The first-order approximation of f at a point $x \in \mathbf{int\,dom}\ f$ can be expressed as (the affine function of z)

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

Examples for Gradient

As a simple example, consider the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$,

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

where $P \in \mathbf{S}^n, q \in \mathbf{R}^n$, and $r \in \mathbf{R}$. Then, its gradient is

$$\nabla f(x) = Px + q.$$

Gradient of Functions Defined on Symmetric Matrices (1/3)

Gradient of functions on symmetric matrices

If $f : \mathbf{S}^n \to \mathbf{R}$, then the **gradient** of f is defined as

$$\nabla f(X) = \mathrm{svec}^{-1}\left(\nabla \tilde{f}(\mathrm{svec}(X))\right)$$

where $\tilde{f} : \mathbf{R}^{n(n+1)/2} \to \mathbf{R}$ is defined from f as $\tilde{f}(x) \triangleq f\left(\mathrm{svec}^{-1}(x)\right)$.

$$\begin{tikzcd}
\mathbf{R}^{n(n+1)/2} \arrow[r, shift left, "\mathrm{svec}^{-1}"] \arrow[rr, "\tilde{f}"'] & \mathbf{S}^n \arrow[l, shift left, "\mathrm{svec}"] \arrow[r, "f"] & \mathbf{R}
\end{tikzcd}$$

Gradient of Functions Defined on Symmetric Matrices (2/3)

Gradient of functions on symmetric matrices

If $f : \mathbf{S}^n \to \mathbf{R}$, then the **gradient** of f is defined as

$$\nabla f(X) = \operatorname{svec}^{-1}\left(\nabla \tilde{f}(\operatorname{svec}(X))\right)$$

where $\tilde{f} : \mathbf{R}^{n(n+1)/2} \to \mathbf{R}$ is defined from f as $\tilde{f}(x) \triangleq f\left(\operatorname{svec}^{-1}(x)\right)$.

Gradient w.r.t. Symmetric Matrices

It can be shown that

$$\nabla f(X) = \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \frac{1}{2}\frac{\partial f(X)}{\partial x_{12}} & \cdots & \frac{1}{2}\frac{\partial f(X)}{\partial x_{1n}} \\ \frac{1}{2}\frac{\partial f(X)}{\partial x_{12}} & \frac{\partial f(X)}{\partial x_{22}} & \cdots & \frac{1}{2}\frac{\partial f(X)}{\partial x_{2n}} \\ \vdots & & & \\ \frac{1}{2}\frac{\partial f(X)}{\partial x_{1n}} & \frac{1}{2}\frac{\partial f(X)}{\partial x_{2n}} & \cdots & \frac{\partial f(X)}{\partial x_{nn}} \end{bmatrix} \in \mathbf{S}^n.$$

Gradient of Functions Defined on Symmetric Matrices (3/3)

$$\nabla f(X) = \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \frac{1}{2}\frac{\partial f(X)}{\partial x_{12}} & \cdots & \frac{1}{2}\frac{\partial f(X)}{\partial x_{1n}} \\ \frac{1}{2}\frac{\partial f(X)}{\partial x_{12}} & \frac{\partial f(X)}{\partial x_{22}} & \cdots & \frac{1}{2}\frac{\partial f(X)}{\partial x_{2n}} \\ \vdots & & & \\ \frac{1}{2}\frac{\partial f(X)}{\partial x_{1n}} & \frac{1}{2}\frac{\partial f(X)}{\partial x_{2n}} & \cdots & \frac{\partial f(X)}{\partial x_{nn}} \end{bmatrix} \in \mathbf{S}^n.$$

- E.g. Let $f : \mathbf{S}^2 \to \mathbf{R}$ defined as $f\left(\begin{bmatrix} x & y \\ y & z \end{bmatrix}\right) = 2x + 2y + 3z$, then

$$\nabla f\left(\begin{bmatrix} x & y \\ y & z \end{bmatrix}\right) = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}.$$

- It can be shown that

$$\lim_{Z \in \mathbf{dom}\, f, Z \neq X, Z \to X} \frac{|f(Z) - f(X) - \mathbf{tr}\,(\nabla f(X)(Z - X))|}{||Z - X||_F} = 0.$$

Gradient of Functions on Symmetric Matrices – Example

- Let $f : \mathbf{S}^n \to \mathbf{R}$ be defined as $f(X) = \mathbf{tr}\ (AX)$ where $A \in \mathbf{R}^{n \times n}$. Then,

$$\nabla f(X) = \frac{A + A^T}{2}.$$

  If $A \in \mathbf{S}^n$, then $\nabla f(X) = A$.
- The first-order approximation of f at a point $X \in \mathbf{int\ dom}\ f$ can be expressed as (the affine function of Z)

$$f(X) + \mathbf{tr}\ (\nabla f(X)(Z - X)).$$

Gradient of the Log-Determinant Function (1/2)

Gradient of the Log-Determinant Function

Consider the function $f : \mathbf{S}^n \to \mathbf{R}$ given by

$$f(X) = \log \mathbf{det}\ X, \quad \mathbf{dom}\ f = \mathbf{S}^n_{++}.$$

We will show that

$$\nabla f(X) = X^{-1}.$$

Gradient of the Log-Determinant Function (2/2)

Notice that

$$
\begin{aligned}
\log \mathbf{det}\ Z &= \log \mathbf{det}\ (X + \Delta X) \\
&= \log \mathbf{det}\ (X^{1/2}(I + X^{-1/2}\Delta X X^{-1/2})X^{1/2}) \\
&= \log \mathbf{det}\ X + \log \mathbf{det}\ (I + X^{-1/2}\Delta X X^{-1/2}) \\
&= \log \mathbf{det}\ X + \sum_{i=1}^{n} \log(1+\lambda_i) \quad \text{(where } \lambda_i\text{'s are eig.vals. of } X^{-1/2}\Delta X X^{-1/2}\text{)} \\
&\approx \log \mathbf{det}\ X + \sum_{i=1}^{n} \lambda_i \\
&= \log \mathbf{det}\ X + \mathbf{tr}\ (X^{-1/2}\Delta X X^{-1/2}) \\
&= \log \mathbf{det}\ X + \mathbf{tr}\ (X^{-1}(Z - X)).
\end{aligned}
$$

So, $f(Z) \approx f(X) + \mathbf{tr}\ (X^{-1}(Z - X))$.
Therefore,

$$\nabla f(X) = X^{-1}.$$

Second derivative and Hessian matrix

Hessian matrix

Let $f : \mathbf{R}^n \to \mathbf{R}$. Then **second derivative** or **Hessian matrix** of f at $x \in \mathbf{int\,dom}\, f$, denoted $\nabla^2 f(x)$, is given by

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i = 1, \cdots, n, j = 1, \cdots, n.$$

- The Hessian matrix is a symmetric matrix as long as we assume that any second derivatives of f are continuous.
- The second-order approximation of f, at or near x, is the quadratic function of z defined by

$$f(z) \approx f(x) + \nabla f(x)^T (z - x) + (1/2)(z - x)^T \nabla^2 f(x)(z - x).$$

Second derivative and Hessian matrix

- Note that $D\nabla f(x) = \nabla^2 f(x) = (\nabla^2 f(x))^T$.
- Consider again the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$,

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

where $P \in \mathbf{S}^n, q \in \mathbf{R}^n$, and $r \in \mathbf{R}$. Then, its gradient is

$$\nabla f(x) = Px + q.$$

So, its Hessian matrix is

$$\nabla^2 f(x) = P.$$

Chain Rules

- Let $f : \mathbf{R}^n \to \mathbf{R}^m$ and $g : \mathbf{R}^m \to \mathbf{R}^p$. Consider the composite function $h : \mathbf{R}^n \to \mathbf{R}^p$, with $\mathbf{dom}\ h = \mathbf{dom}\ f \cap f^{-1}(\mathbf{dom}\ g)$, defined as $h(z) = g(f(z))$ for any $z \in \mathbf{dom}\ h$.
- Then,

$$Dh(x) = Dg(f(x)) \cdot Df(x).$$

To see this, note that

$$\begin{aligned} h(z) &= h(x + \Delta x) \\ &= g(f(x + \Delta x)) \\ &\approx g(f(x) + Df(x)\Delta x) \\ &\approx g(f(x)) + Dg(f(x))Df(x)\Delta x \\ &= h(x) + Dh(x)(z - x) \end{aligned}$$

for any Δx with a sufficiently small $\|\Delta x\|$.

1. Review of important concepts (II)
   - (§A.2) Open Sets and Closed Sets
   - (§A.4) Derivative, gradient, and Hessian

2. Convex functions – basics (§3.1)
   - Definition
   - First-order conditions
   - Second-order conditions
   - Examples

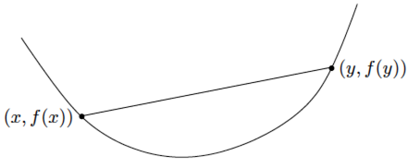
Definitions of Convex Functions

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[1]



[1] The term graph will be formally defined in a later section.

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 $\theta \in [0, 1]$,

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y). \qquad (2)$$

- A function f is **strictly convex** if strict inequality holds in (2) 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 (2), 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.[2]
- That is, f is convex if and only if $\forall x \in \mathbf{dom}\ f, v \in \mathbf{R}^n$, the function $g(t) = f(x + tv)$ is convex on $\{t \mid x + tv \in \mathbf{dom}\ f\}$.
- A convex function is continuous on the relative interior of its domain; it can have discontinuities only on its relative boundary.

[2] It would be more accurate to say "A function is convex if and only if it is convex when restricted to the intersection of any line and its domain (whenever such intersection is not empty)."

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

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

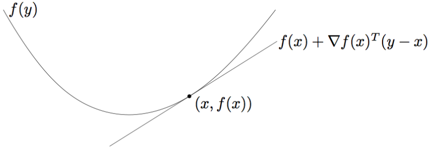
First-Order Conditions

First-Order Conditions

Suppose f is differentiable (implying that $\mathbf{dom}\ f$ is open). Then f is convex if and only if $\mathbf{dom}\ f$ is convex and

$$f(y) \geq f(x) + \nabla f(x)^T (y - x),\ \forall x, y \in \mathbf{dom}\ f.$$

- 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.



Remark: In this course, we do not define the derivative of a function at a boundary point.

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 $\mathbf{dom}\ f$ is convex and for $x, y \in \mathbf{dom}\ f, x \neq y$, 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 $\mathbf{dom}\ f$ is convex and for $x, y \in \mathbf{dom}\ f$, we have

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

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

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

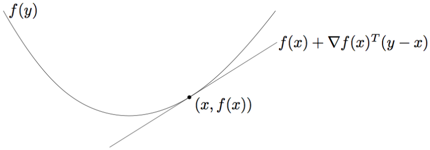
Proof of First-Order Conditions

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 $g(t) = f(ty + (1-t)x)$ and $\mathbf{dom}\ g = \{t \in \mathbf{R} \mid ty + (1-t)x \in \mathbf{dom}\ f\}$.

Second-Order Conditions

- Assume that $f : \mathbf{R} \to \mathbf{R}$ is twice differentiable with $\mathbf{dom}\ f = \mathbf{R}$, 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 $\mathbf{dom}\ f$ (open).

Second-Order Conditions

Then, f is convex if and only if $\mathbf{dom}\ f$ 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 $\mathbf{dom}\ f$ is convex.

Second-Order Conditions – Strict Convexity, Concavity

Second-Order Conditions for Concavity

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

Second-Order Conditions for Strict Convexity

If $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\ f$ where $\mathbf{dom}\ f$ is convex, 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 $\mathbf{dom}\ f = \mathbf{R}^n$, given by

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

  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.
- Max function. $f(x) = \max\{x_1, \cdots, x_n\}$ 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 $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbf{R}^n$.
    - Note that $\max\{x_1, \cdots, x_n\} \leq f(x) \leq \max\{x_1, \cdots, x_n\} + \log n$.
- 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 $f(X) = \log \mathbf{det}\ X$ is concave on $\mathbf{dom}\ f = \mathbf{S}^n_{++}$.

Norms and Max function

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

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

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

$$\begin{aligned}
\max_i(\theta x_i + (1-\theta)y_i) &\le \max_i \theta x_i + \max_i (1-\theta) y_i \quad ^3 \\
&= \theta \max_i x_i + (1-\theta) \max_i y_i. \\
&= \theta f(x) + (1-\theta) f(y).
\end{aligned}$$

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

[3] It is worthy to note that the index "i" that achieves the maximum of each of the three terms in this inequality may be different in general.

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 $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ 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 $z = (e^{x_1}, \cdots, e^{x_n})$, 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 (1/2)

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

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

and

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k x_l} (k \neq l).$$

Geometric mean (2/2)

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

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

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 $f : \mathbf{S}^n \to \mathbf{R}$, $f(X) = \log \mathbf{det}\ X$, with $\mathbf{dom}\ f = \mathbf{S}^n_{++}$, 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 $g(t) = f(Z + tV)$, $\mathbf{dom}\ g = \{t \mid Z + tV \succ 0\}$.
- Then it can be shown that

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

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

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