Lecture 5. Convex Functions

5.1 Definition

We now introduce *convex functions*. For convenience, use $\bar{\theta}$ to denote $1 - \theta$ for any $\theta \in \mathbb{R}$.

Definition (*Convex functions*)

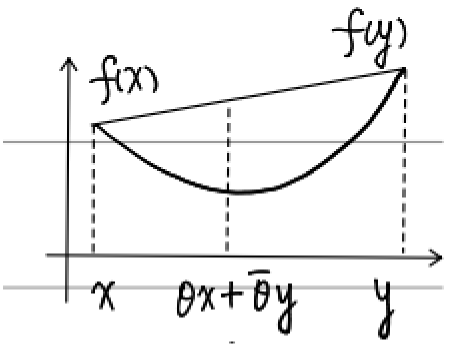
Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be a real-valued function. Then it is *convex* if

- the domain $\operatorname{dom} f$ is convex;
- f satisfies the *Jensen's inequality*, i.e., for all $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$ and $\theta \in [0, 1]$, it holds that

$$f(\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + \bar{\theta} f(\boldsymbol{y}) \,.$$

The function f is *concave* if $-f$ is convex.

Geometrically, the line segment between $(\boldsymbol{x}, f(\boldsymbol{x}))$ and $(\boldsymbol{y}, f(\boldsymbol{y}))$ lies above the graph of f.



Definition (*Strictly convex functions*)

Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a real-valued function. Then it is *strictly convex* if

- the domain $\operatorname{dom} f$ is convex;

- f satisfies the *strict Jensen's inequality*, i.e., for all $\boldsymbol{x} \neq \boldsymbol{y} \in \operatorname{dom} f$ and $\theta \in [0, 1]$, it holds that

$$f(\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}) < \theta f(\boldsymbol{x}) + \bar{\theta} f(\boldsymbol{y}) \,.$$

The function f is *strictly concave* if $-f$ is strictly convex.

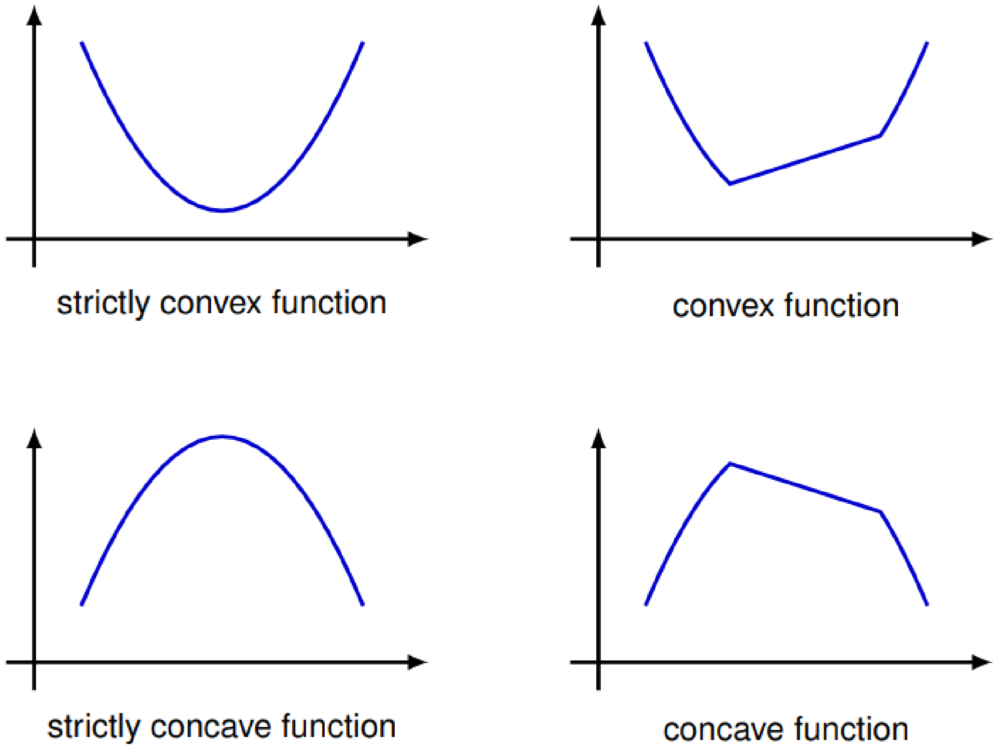
Note that an affine function $f(\boldsymbol{x}) = \boldsymbol{w}^\mathsf{T} \boldsymbol{x} + b$ is both convex and concave, but not strictly convex or strictly concave. The following proposition shows that if a function is both convex and concave, then it must be an affine function.

Proposition

Let f be convex. If $f(\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}) = \theta f(\boldsymbol{x}) + \bar{\theta} f(\boldsymbol{y})$ for some $\theta = \theta_0 \in (0, 1)$, then it holds for any $\theta \in [0, 1]$, i.e., $g(\theta) = f(\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y})$ is an affine function for $\theta \in [0, 1]$.

Why? Suppose there exists $\theta_1 \in (0, \theta_0)$ such that $f(\theta_1 \boldsymbol{x} + \bar{\theta}_1 \boldsymbol{y}) < \theta_1 f(\boldsymbol{x}) + \bar{\theta}_1 f(\boldsymbol{y})$. Then choose any $\theta_2 \in (0, \theta_0)$ and note that $(\theta_0 \boldsymbol{x} + \bar{\theta}_0 \boldsymbol{y})$ is a convex combination of $(\theta_1 \boldsymbol{x} + \bar{\theta}_1 \boldsymbol{y})$ and $(\theta_2 \boldsymbol{x} + \bar{\theta}_2 \boldsymbol{y})$. Applying the Jensen's inequality on them, we can obtain contradictions.

Why these functions are called *convex*? Someone may think their graphs are somehow *concave*. Actually, what we concern is the area above the graphs of convex functions.



Definition

Given a real-valued function $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$,

- the *graph* of f is defined as

$$\{(\boldsymbol{x}, f(\boldsymbol{x})) \in \mathbb{R}^{n+1} \mid \boldsymbol{x} \in D\};$$

- the *epigraph* of f is defined by

$$\operatorname{epi}(f) \triangleq \{(\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \mid \boldsymbol{x} \in D, y \geq f(\boldsymbol{x})\}\,;$$

- the *hypograph* of f is defined by

$$\operatorname{hyp}(f) \triangleq \{(\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \mid \boldsymbol{x} \in D, y \leq f(\boldsymbol{x})\}\,.$$

Theorem

Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a real-valued function. Then f is convex if and only if $\operatorname{epi}(f)$ is convex.

Proof

- " $\Longleftarrow$ ". Given any $\boldsymbol{x}, \boldsymbol{y} \in D$ and $\theta \in [0,1]$, since epi(f) is convex, and both $(\boldsymbol{x}, f(\boldsymbol{x}))$ and $(\boldsymbol{y}, f(\boldsymbol{y}))$ in epi(f), then

$$\left(\theta\boldsymbol{x} + \bar{\theta}\boldsymbol{y}, \theta f(\boldsymbol{x}) + \bar{\theta} f(\boldsymbol{y})\right) \in \text{epi}(f)\,,$$

which implies that $\theta\boldsymbol{x} + \bar{\theta}\boldsymbol{y} \in D$ and

$$f(\theta\boldsymbol{x} + \bar{\theta}\boldsymbol{y}) \le \theta f(\boldsymbol{x}) + \bar{\theta} f(\boldsymbol{y})\,.$$

Thus D is convex and the Jensen's inequality holds.

- " $\Longrightarrow$ ". Given $(\boldsymbol{x}_1, x_2), (\boldsymbol{y}_1, y_2) \in \text{epi}(f)$ where $\boldsymbol{x}_1, \boldsymbol{y}_1 \in D$ and $\boldsymbol{x}_2, \boldsymbol{y}_2 \in \mathbb{R}$, and $\theta \in [0,1]$, let

$$\boldsymbol{z}_1 = \theta\boldsymbol{x}_1 + \bar{\theta}\boldsymbol{y}_1, z_2 = \theta x_2 + \bar{\theta} y_2\,.$$

By the convexity of D, $\boldsymbol{z}_1 \in D$. By the definition of epi(f),

$$\begin{aligned} z_2 &= \theta x_2 + \bar{\theta} y_2 \\ &\ge \theta f(\boldsymbol{x}_1) + \bar{\theta} f(\boldsymbol{x}_2) \\ &\ge f\left(\theta\boldsymbol{x}_1 + \bar{\theta}\boldsymbol{x}_2\right) \\ &= f(\boldsymbol{z}_1) \end{aligned}$$

Then, $(\boldsymbol{z}_1, z_2) \in \text{epi}(f)$, which means epi$(f)$ is convex.

Example

- (*Univariate functions*) $f(x) = \frac{1}{x}$ where $x > 0$, $f(x) = e^x$ and $f(x) = -\log x$ where $x > 0$ are all strictly convex. (Or we say $f(x) = \frac{1}{x}$ and $f(x) = -\log x$ are convex over $(0, \infty)$, and $f(x) = e^x$ is convex over $\mathbb{R}$.)
- (*Norm functions*) Any norm $\|\cdot\|$ is a convex function, but can not be strictly convex (e.g., L^1-norm, not strictly convex by absolute homogeneity).

Why do we consider convex functions? One of the most important properties of convex functions is that local minimum points must be global minimum.

Theorem

If $f(x) : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is a convex function and has a local minimum point $x^* \in D$, then x^* is a global minimum point.

Proof

Assume not. Then there exists $y^* \in D$ such that $f(y^*) < f(x^*)$. Since x^* is a local minimum point, there exists $\varepsilon > 0$ such that for all $x \in D \cap \mathcal{B}(x^*, \varepsilon)$, $f(x) \geq f(x^*)$. Choose $\theta \in (0, 1)$ sufficiently small such that $\theta \|y^* - x^*\| < \varepsilon$. Then $z = x^* + \theta(y^* - x^*) = \bar{\theta}x^* + \theta y^* \in \mathcal{B}(x^*, \varepsilon)$. By the convexity of f, $z \in D$ and thus $f(z) \leq \bar{\theta}f(x^*) + \theta f(y^*) < f(x^*)$, which contradicts that x^* is local minimum.

Extended-value functions

Suppose the domain of function f is not $\mathbb{R}^n$. Then we can extend the value of f to $\mathbb{R} \cup \{\infty\}$ so that the domain of f can be extended to $\mathbb{R}^n$, namely, we can define $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ for a function $f : D \to \mathbb{R}$ as follows:

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

where we assume $\infty + x = \infty$, $\infty \cdot x = \infty$ for any $x > 0$ and $\infty \cdot 0 = 0$.
Note that the extended-value function of a convex function is still convex, since the epigraph remains the same.

Generalization of the Jensen's inequality

It is easy to show the following generalization of Jensen's inequality by induction.

Proposition

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function. Then for any $\theta_1, \theta_2, \ldots, \theta_m \in [0, 1]$ where $\sum_i \theta_i = 1$ and any $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$, it holds that

$$f(\theta_1 \boldsymbol{x}_1 + \cdots + \theta_m \boldsymbol{x}_m) \leq \theta_1 f(\boldsymbol{x}_1) + \cdots + \theta_m f(\boldsymbol{x}_m) .$$

Intuitively we can generalize the inequality to the convex combination of infinite many variables. We actually have the following generalized form of the Jensen's inequality but we should note that the proof of it is nontrivial since we cannot use induction!

Theorem

Let $(\Omega, \mathfrak{F}, \mu)$ be a probability space, X be an integrable real-valued random variable and f be a convex function. Then it holds that

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)] \,.$$

Equivalently we have the following measure-theoretic form

$$f\left(\int_\Omega X \,\mathrm{d}\mu\right) \leq \int_\Omega f(X) \,\mathrm{d}\mu \,.$$

5.2 Properties and conditions of convexity

Midpoint convexity

Now we would like to proof that $f(x) = -\ln x$ is a convex function over $(0, \infty)$. We verify the Jensen's inequality:

$$\theta \ln x + \bar{\theta} \ln y \leq \ln(\theta x + \bar{\theta} y) \,.$$

Taking the exponent on the both sides, it is equivalent to show that

$$x^\theta \cdot y^{\bar{\theta}} \leq \theta x + \bar{\theta} y \,.$$

If $\theta = 1/2$, it is trivial by the *AM-GM inequality*. However, how can we verify the inequality for $\theta \neq 1/2$, or is it sufficient for us to verify Jensen's inequality only for $\theta = 1/2$?

Remark

Technically we cannot use the weighted AM-GM inequality here, since the weighted version is usually proved by the Jensen's inequality and concavity of the logarithm, which is what we want to show!

We say a function $f : D \to \mathbb{R}$ is *midpoint convex* if the Jensen's inequality holds for $\theta = 1/2$ and every $x, y \in D$. Clearly, convex functions are midpoint convex. Conversely, it is not necessarily true. But luckily, if the function is also continuous, then it is convex.

Theorem (*Jensen, 1905*)

If f is a continuous midpoint convex function defined on a convex set D, then f is convex.

Proof

Prove by contradiction. Assume $\exists\, x, y \in D$ and $\theta \neq 1/2$ such that $f(\theta x + \bar{\theta} y) > \theta f(x) + \bar{\theta} f(y)$. Let

$$g(\alpha) = f(\alpha x + \bar{\alpha} y) - (\alpha f(x) + \bar{\alpha} f(y)), \qquad \alpha \in [0,1]\,.$$

Then we have $g(0) = g(1) = 0$ and $g(\theta) > 0$. By the compactness of $[0,1]$, there exists $M = \max_{\alpha \in [0,1]} g(\alpha) > 0$, and exists $\alpha \in [0,1]$ such that $g(\alpha) = M$. Now let $\alpha_0 = \inf\{\alpha \in [0,1] : g(\alpha) = M\}$. By the continuity of g, $g(\alpha_0) = M > 0$. Thus $\alpha_0 \neq 0, 1$. Select $\delta > 0$ sufficiently small such that $(\alpha_0 - \delta, \alpha_0 + \delta) \subseteq (0,1)$. Since f is midpoint convexity , we have

$$g(\alpha_0 - \delta) + g(\alpha_0 + \delta) \geq 2f(\alpha_0 x + \bar{\alpha}_0 y) - (2\alpha f(x) + 2\bar{\alpha}_0 f(y)) = 2g(\alpha_0)\,.$$

However, by the definition of α_0, we have $g(\alpha_0) < M$, $g(\alpha_0) = M$, and $g(\alpha_0 + \delta) \leq M$, which leads to a contradiction.

Warning

There exists midpoint convex but not convex functions if we admit the *axiom of choice*. Such a function would have to be non-measurable.

Now we use this theorem to verify a more complicated example: $f(\boldsymbol{X}) = -\log\det \boldsymbol{X}$ is convex for positive definite matrix $\boldsymbol{X} \in \mathcal{S}^n_{++}$.

We admit the fact that $f(\boldsymbol{X})$ is continuous. Then we verify the midpoint convexity:

$$\det \frac{\boldsymbol{X} + \boldsymbol{Y}}{2} \geq \sqrt{(\det \boldsymbol{X})(\det \boldsymbol{Y})}\,.$$

Since $\boldsymbol{X} \succ 0$, $\boldsymbol{X}^{-1}$ exists and $\det \boldsymbol{X}^{-1} > 0$. So our goal is equivalent to show that

$$\det \frac{\boldsymbol{I} + \boldsymbol{X}^{-1}\boldsymbol{Y}}{2} \geq \sqrt{\det \boldsymbol{X}^{-1}\boldsymbol{Y}}\,.$$

Note that
$\det(\lambda \boldsymbol{I}-\boldsymbol{X}^{-1}\boldsymbol{Y})=0 \iff \det\left(\frac{\lambda \boldsymbol{I}-\boldsymbol{X}^{-1}\boldsymbol{Y}}{2}\right)=0 \iff \det\left(\frac{1+\lambda}{2}\boldsymbol{I}-\frac{\boldsymbol{I}+\boldsymbol{X}^{-1}\boldsymbol{Y}}{2}\right)$, so
$\lambda_i\left(\frac{\boldsymbol{I}+\boldsymbol{X}^{-1}\boldsymbol{Y}}{2}\right)=\frac{1+\lambda_i(\boldsymbol{X}^{-1}\boldsymbol{Y})}{2}$, where $\lambda_i(A)$ is the i-th eigenvalue of matrix A. Thus we have

$$\det \frac{\boldsymbol{I}+\boldsymbol{X}^{-1}\boldsymbol{Y}}{2}=\prod \lambda_i(\frac{\boldsymbol{I}+\boldsymbol{X}^{-1}\boldsymbol{Y}}{2})=\prod \frac{1+\lambda_i(\boldsymbol{X}^{-1}\boldsymbol{Y})}{2}$$

and

$$\sqrt{\det \boldsymbol{X}^{-1}\boldsymbol{Y}}=\sqrt{\prod \lambda_i(\boldsymbol{X}^{-1}\boldsymbol{Y})}\,.$$

Now it suffices to show that $\lambda_i(\boldsymbol{X}^{-1}\boldsymbol{Y})\geq 0$ for all i. Since $\boldsymbol{X}\succ 0$, consider the eigen-decomposition $\boldsymbol{X}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}$. It implies that there exists
$\boldsymbol{X}^{1/2}=\boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^{-1}\succ 0$ and invertible (note that $(\boldsymbol{X}^{1/2})^2=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}=\boldsymbol{X}$).
If $\boldsymbol{X}^{-1}\boldsymbol{Y}v=\lambda v$, then

$$\boldsymbol{X}^{1/2}\boldsymbol{X}^{-1}\boldsymbol{Y}\boldsymbol{X}^{-1/2}(\boldsymbol{X}^{1/2}v)=\boldsymbol{X}^{1/2}(\boldsymbol{X}^{-1}\boldsymbol{Y}v)=\lambda\boldsymbol{X}^{1/2}v\,,$$

and vice versa. So

$$\lambda_i(\boldsymbol{X}^{-1}\boldsymbol{Y})=\lambda_i(\boldsymbol{X}^{1/2}\boldsymbol{X}^{-1}\boldsymbol{Y}\boldsymbol{X}^{1/2})=\lambda_i(\boldsymbol{X}^{-1/2}\boldsymbol{Y}\boldsymbol{X}^{-1/2})\,.$$

Note that $\boldsymbol{X}^{-1/2}\boldsymbol{Y}\boldsymbol{X}^{-1/2}$ is symmetric, and $\forall\, v\neq 0$, $v^\top\boldsymbol{X}^{-1/2}\boldsymbol{Y}\boldsymbol{X}^{-1/2}v=u^\top\boldsymbol{Y}u\geq 0$, where $u=\boldsymbol{X}^{-1/2}v$. Hence, $\boldsymbol{X}^{-1/2}\boldsymbol{Y}\boldsymbol{X}^{-1/2}\succeq 0$, which yields that all eigenvalues are nonnegative.
Combining all of above, we conclude that $f(\boldsymbol{X})$ is convex.

Zeroth order condition

We now consider some properties of convexity. Conversely, these properties also provide some criteria to verify convexity.

Let f be a single-variable function. Usually it is easy to verify the Jensen's inequality. So our first condition is that, f is convex if and only if its restriction to any line is convex.

Theorem

Suppose f is a function defined on a convex set $D\subseteq\mathbb{R}^n$. Then f is convex iff $\forall\,\boldsymbol{x}\in D,\boldsymbol{v}\in\mathbb{R}^n$, $g(t)=f(\boldsymbol{x}+t\boldsymbol{v})$ is convex.

Example

$f(x_1, x_2, \ldots, x_n) = e^{x_1+x_2+\cdots+x_n}$ is convex, since $\forall\, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,
$g(t) = f(\boldsymbol{u} + t\boldsymbol{v}) = e^{u_1+\cdots+u_n} \cdot e^{(v_1+\cdots+v_n)t}$ is convex.

Proof

- " $\Longrightarrow$ ". Assume f is convex. Fix $\boldsymbol{x} \in D$. For any $\boldsymbol{v} \in \mathbb{R}^n$, let $t_1, t_2 \in \operatorname{dom} g$. It suffices to show that $\forall\, \theta \in [0,1]$, (i) $\theta t_1 + \bar{\theta} t_2 \in \operatorname{dom} g$; (ii) $g(\theta t_1 + \bar{\theta} t_2) \le \theta g(t_1) + \bar{\theta} g(t_2)$.
  Let $\boldsymbol{x}_i = \boldsymbol{x} + t_i \cdot \boldsymbol{v}$. Since $t_1, t_2 \in \operatorname{dom} g$, $\boldsymbol{x}_1, \boldsymbol{x}_2 \in D$. Thus, $\theta \boldsymbol{x}_1 + \bar{\theta} \boldsymbol{x}_2 = \boldsymbol{x} + (\theta t_1 + \bar{\theta} t_2)\boldsymbol{v} \in D$, which indicates that $\theta t_1 + \bar{\theta} t_2 \in \operatorname{dom} g$. Furthermore,
  $$g(\theta t_1 + \bar{\theta} t_2) = f(\theta \boldsymbol{x}_1 + \bar{\theta} \boldsymbol{x}_2) \le \theta f(\boldsymbol{x}_1) + \bar{\theta} f(\boldsymbol{x}_2) = \theta g(t_1) + \bar{\theta} g(t_2).$$
- " $\Longleftarrow$ ". Given $\boldsymbol{x}, \boldsymbol{y} \in D$, let $\boldsymbol{v} = \boldsymbol{y} - \boldsymbol{x}$ and $g(t) = f(\boldsymbol{x} + t\boldsymbol{v})$. Since g is convex and $0, 1 \in \operatorname{dom} g$, we have that $\forall\, \theta \in [0,1]$, $\theta \in \operatorname{dom} g$. Thus $\boldsymbol{x} + \theta(\boldsymbol{y} - \boldsymbol{x}) = \bar{\theta}\boldsymbol{x} + \theta \boldsymbol{y} \in D$. Moreover, $g(\theta) = f(\bar{\theta}\boldsymbol{x} + \theta\boldsymbol{y}) \le \bar{\theta} g(0) + \theta g(1) = \bar{\theta} f(\boldsymbol{x}) + \theta f(\boldsymbol{y})$, which implies that f is convex.

First order condition

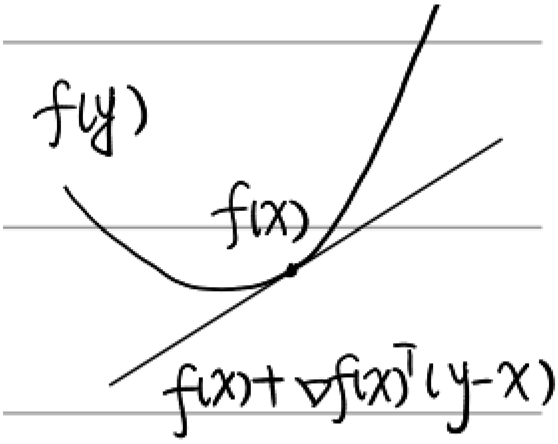
If f is further differentiable, we have the following important criterion (and an important property) for convex functions.

Theorem

Suppose f is differentiable in an open convex set $D = \operatorname{dom} f$. Then f is convex in D iff

$$\forall\, x, y \in D, \quad f(y) \ge f(x) + \nabla f(x)^{\mathsf{T}} (y - x)\,.$$

The first order condition shows that convex functions have linear lower bounds.



We usually use $\langle \cdot, \cdot \rangle$ to denote the inner product. So the first condition is also written as $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$.

Example (*Bernoulli's inequality*)

1. $(1+x)^r \geq 1 + rx$ if $r \geq 1$ and $x \geq -1$;
2. $e^x \geq 1 + x$.

Proof

- " $\Longrightarrow$ ". Fix any $x, y \in D$. Let $v = y - x$. By the Jensen's inequality,

  $$\forall t \in [0,1], \quad f(x+tv) \leq (1-t)f(x) + tf(y).$$

  Rearranging it, we have

  $$f(x+tv) - f(x) \leq t(f(y) - f(x)) \,.$$

  Recall that $\nabla f(x)^\top v = \lim_{t \to 0} \frac{f(x+tv)-f(x)}{t}$. Taking the limits $t \to 0$ on both sides, we have

  $$\nabla f(x)^\top v \leq f(y) - f(x) \,.$$

- " $\Longleftarrow$ ". For all $x, y \in D$ and $\theta \in [0,1]$, let $z = \theta x + \bar{\theta} y$. The first-order condition gives that

  $$\begin{cases} f(x) \geq f(z) + \nabla f(z)^\top (x - z) & (1) \\ f(y) \geq f(z) + \nabla f(z)^\top (y - z) & (2) \end{cases}.$$

  Then $\theta(1) + \bar{\theta}(2)$ immediately implies that $\theta f(x) + \bar{\theta} f(y) \geq f(z)$.

Corollary

1. In particular, if $\nabla f(x) = \mathbf{0}$, then $f(y) \geq f(x)$ for all $y \in D$. If f is further *strictly convex*, x is the unique global minimum point.
2. Given $\boldsymbol{x}_0 \in D$, $\{(\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \mid y = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^\mathsf{T}(\boldsymbol{x} - \boldsymbol{x}_0)\}$ is a supporting hyperplane of epi f at $\boldsymbol{x}_0$.

The first order condition also holds for the strict convexity if applying strict inequality. For the proof for strict convexity, the $\Longleftarrow$ direction remains the same. However how can we prove the $\Longrightarrow$ direction? Note that taking the limit cannot keep the strict inequality.

Theorem (*First order condition for strict convexity*)

Suppose f is differentiable in an open convex set $D = \operatorname{dom} f$. Then f is strictly convex in D iff

$$\forall x \neq y \in D, \quad f(y) > f(x) + \nabla f(x)^\mathsf{T}(y - x).$$

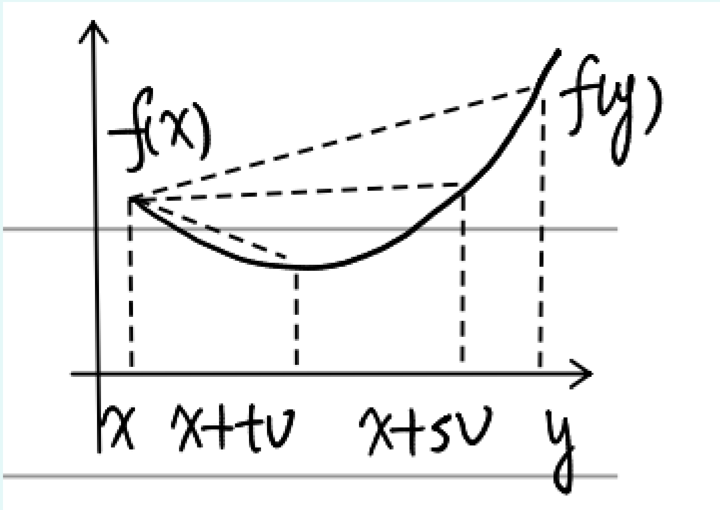
Proof

Let $v = y - x$, similar to the proof of non-strict version, we have

$$\frac{f(x + tv) - f(x)}{t} < f(y) - f(x)\,.$$

Consider another coefficient s such that $0 < t < s < 1$, then we also have

$$\frac{f(x + sv) - f(x)}{s} < f(y) - f(x)\,.$$



Applying Jensen inequality (writing $x + tv$ as a convex combination of x and $x + sv$), it's easy to verify that

$$\frac{f(x+tv)-f(x)}{t} < \frac{f(x+sv)-f(x)}{s}\,.$$

Taking $t \to 0$, we have

$$\begin{aligned}\nabla f(x)^\top (y-x) &= \lim_{t\to 0}\frac{f(x+tv)-f(x)}{t}\\ &\le \frac{f(x+sv)-f(x)}{s}\\ &< f(y)-f(x)\,,\end{aligned}$$

which prove the first order condition for strictly convex functions.

An important corollary of the first order condition is the property of monotone gradient.

Corollary (*Monotone gradient*)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Then, f is a convex function if and only if ∇f is monotone, i.e., $\langle \nabla f(x) - \nabla f(y), x - y\rangle \ge 0$.

Proof

- " $\Longrightarrow$ ". When f is convex, for all $x, y \in \mathbb{R}^n$, by the first order condition,

  $$\begin{aligned} f(y) &\ge f(x) + \langle \nabla f(x), y-x\rangle,\\ f(x) &\ge f(y) + \langle \nabla f(y), x-y\rangle. \end{aligned}$$

  Then,

  $$0 \ge \langle \nabla f(x) - \nabla f(y), y-x\rangle.$$

  Thus we obtain the monotone gradient.
- " $\Longleftarrow$ ". When ∇f is monotone, for all $x, y \in \mathbb{R}^n$, define the function $g(t) = f(\bar{t}x + ty)$. Then

  $$g'(t) = \langle \nabla f(\bar{t}x + ty), y-x\rangle.$$

  For $t_1, t_2 \in [0,1]$, by elementary calculation

  $$(g'(t_1) - g'(t_2))(t_1 - t_2) = \langle \nabla f(\bar{t}_1 x + t_1 y) - \nabla f(\bar{t}_2 x + t_2 y), (t_1 - t_2)(y-x\rangle$$

Note that $(\bar{t}_1 x + t_1 y) - (\bar{t}_2 x + t_2 y) = (t_1 - t_2)(y - x)$. Then

$$(g'(t_1) - g'(t_2))(t_1 - t_2) \geq 0.$$

This means, we can assume the dimension $n = 1$. For all $x, y \in \mathbb{R}$ (without loss of generality, assuming $x < y$), by the mean value theorem, there exists $c \in [x, y]$ such that

$$f(y) - f(x) = f'(c)(y - x)\,.$$

Since $(f'(c) - f'(x))(c - x) \geq 0$, we have $f'(c) \geq f'(x)$, which leads to the first order condition $f(y) \geq f(x) + f'(x)(y - x)$.

Second order condition

The property of monotone gradients indicates that the second order derivative is somehow nonnegative. Assume $f : \mathbb{R} \to \mathbb{R}$ is a univariate function. Then

$$(f'(x) - f'(y))(x - y) \geq 0$$

implies that $f'(x)$ is increasing and thus $f''(x)$ is nonnegative. If $f : \mathbb{R}^n \to \mathbb{R}$ be a multivariate function, the second order derivative is $\nabla^2 f(x) \in \mathbb{R}^{n \times n}$. A generalized notion of $f''(x) \geq 0$ is $\nabla^2 f(x) \succeq 0$ in this case.

Theorem

Suppose f is twice differentiable in an open convex set $D = \operatorname{dom} f$. Then f is convex over D iff $\forall x \in D$, $\nabla^2 f(x) \succeq 0$.
Furthermore, if $\forall\, x \in D$, $\nabla^2 f(x) \succ 0$, then f is strictly convex over D (but not vice versa).

Example

- $f(x) = -x \log x$ is strictly concave over $\mathbb{R}_{>0}$, since $f'(x) = -(1 + \log x)$ and $f''(x) = -\frac{1}{x}$.
- $f(x) = e^{ax}$ is strictly convex for all $a \in \mathbb{R} \setminus \{0\}$, since $f''(x) = a^2 e^{ax}$.
- $f(x) = x^a$ is convex over $(0, \infty)$ for $a \geq 1$ or $a < 0$, and concave otherwise.
- The *log-sum-exp* function $f(x) = \log\left(e^{x_1} + e^{x_2} + \cdots + e^{x_n}\right)$ is convex over $\mathbb{R}^n$. (Exercise. Hint: $\frac{\partial f}{\partial x_i} = \frac{e^{x_i}}{\sum e^{x_k}}$, $\frac{\partial^2 f}{\partial x_i^2} = \frac{e^{x_i}(\sum e^{x_k}) - (e^{x_i})^2}{(\sum e^{x_k})^2}$, $\frac{\partial^2 f}{\partial x_i x_j} = -\frac{e^{x_i + x_j}}{(\sum e^{x_k})^2}$,

so $\nabla^2 f(x) = \frac{1}{S(x)^2}\left(S(x)\,\mathrm{diag}\{e^{x_1},\dots,e^{x_n}\} - s(x)s(x)^\top\right)$, where $s(x) = (e^{x_1},\dots,e^{x_n})^\top$ and $S(x) = \sum e^{x_k}$, thus $v^\top \nabla^2 f(x) v = ?)$

Proof

- " $\Longrightarrow$ ". For any $x \in \mathrm{dom}\, f$, we define $g(y) = f(y) - f(x) - \nabla f(x)^\top (y-x)$, where $y \in \mathrm{dom}\, f$. Hence $g(x) = 0$. By the first order condition, $g(y) \geq 0$ for all $y \in \mathrm{dom}\, f$, which implies $\nabla^2 g(x) \succeq 0$ by the second-order condition for optimality. So we have $\nabla^2 f(x) = \nabla^2 g(x) \succeq 0$.
- " $\Longleftarrow$ ". Given two arbitrary points $x, y \in \mathrm{dom}\, f$, let $v = y - x$, and $g(t) = f(x + tv)$. Applying Taylor series with Lagrange remainder to g, there exists $\theta \in [0,1]$ such that

  $$f(y) = f(x+v) = f(x) + \nabla f(x)^\top v + \frac{1}{2} v^\top \nabla^2 f(x + \theta v) v\,.$$

  Since, $\nabla^2 f(x) \succeq 0$ for all $x \in \mathrm{dom}\, f$, it follows that

  $$f(y) \geq f(x) + \nabla f(x)^\top v\,.$$

  Therefore, f is convex by the first order condition.

For strict convexity, we can replace $\succeq$ by $\succ$ and $\geq$ by $>$ in the " $\Longleftarrow$ " direction, and apply the first order condition for strict convexity. However, for the " $\Longrightarrow$ " direction, similar argument cannot be true, since strictly optimal point cannot imply $\nabla^2 f \succ 0$.

Example

Consider the function $f(x) = x^4$. It is strictly convex, but $f''(0) = 0$ is not strictly greater than zero.
Similarly, consider the function $f(x_1, x_2) = x_1^2 + x_2^4$. It is strictly convex, but $\nabla^2 f(x_1, x_2) = \begin{pmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{pmatrix}$, which is not positive definite for $x_2 = 0$.

However, for a series of special functions, the equivalent relation of strict convexity holds. Consider quadratic functions,

$$f(x) = \frac{1}{2} x^\top Q x + w^\top x + b\,.$$

Without loss of generality, we can assume Q is symmetric. This is because

$$x^\top Qx = \frac{1}{2}x^\top Qx + \frac{1}{2}\left(x^\top Qx\right)^\top = x^\top\left(\frac{Q+Q^\top}{2}\right)x\,.$$

It is easy to compute that $\nabla^2 f(x) = Q$. Then the following propositions are true:

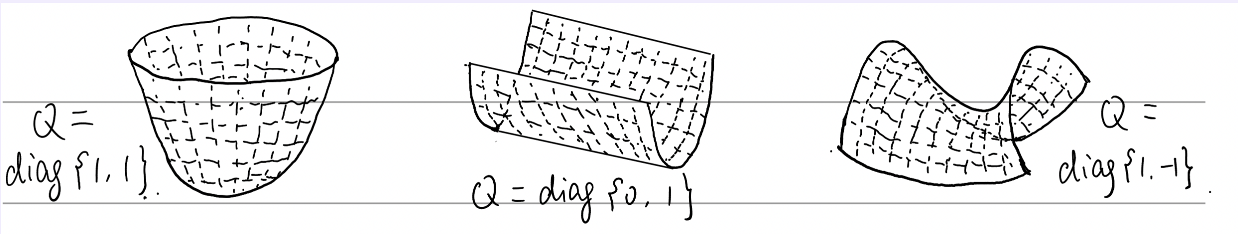
- f is convex iff $Q \succeq 0$. It can be implied by the above theorem.
- f is strictly convex iff $Q \succ 0$. The " $\Longleftarrow$ " direction is easy to verify. So we only need to prove the " $\Longrightarrow$ " direction. Note that

$$\begin{aligned} f(x+v) &= \frac{1}{2}(x+v)^\top Q(x+v) + w^\top(x+v) + b \\ &= \frac{1}{2}(x^\top Qx + 2x^\top Qv + v^\top Qv) + w^\top x + w^\top v + b \\ &= f(x) + \nabla f(x)^\top v + \frac{1}{2}v^\top \nabla^2 f(x) v\,. \end{aligned}$$

  Since f is strictly convex, we have $f(x+v) > f(x) + \nabla f(x)^\top v$ for all $v \neq 0$ (applying the first order condition). That is, $v^\top \nabla^2 f(x) v > 0$, which implies that $Q \succ 0$.

Example

The following figures show the different convexity of f when Q takes different values.



- The first one is strict convex since $Q \succ 0$.
- The second one is convex since $Q \succeq 0$.
- The third one is not convex since $Q \not\succeq 0$.