MA 5037: Optimization Methods and Applications
Convex Functions

Suh-Yuh Yang (楊肅煜)

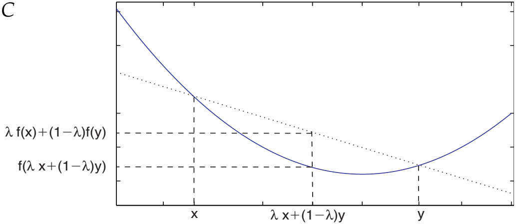
Department of Mathematics, National Central University
Jhongli District, Taoyuan City 320317, Taiwan

First version: July 21, 2018/Last updated: June16, 2025

Convex function

- **Definition:** *A function $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ defined on a convex set C is called convex over C if*

$$f(\underbrace{\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}}_{\in C}) \le \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}), \ \forall\, \boldsymbol{x}, \boldsymbol{y} \in C, \ \lambda \in [0,1].$$



- **Definition:** *A function $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ defined on a convex set C is called strictly convex over C if*

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) < \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}), \ \forall\, \boldsymbol{x} \neq \boldsymbol{y} \in C, \ \lambda \in (0,1).$$

- *A function is called concave if $-f$ is convex. Similarly, f is called strictly concave if $-f$ is strictly convex.*

Convexity of affine functions and norms

- **Convexity of affine functions:** Let $f(\boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x} + b$, where $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Taking $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\lambda \in [0, 1]$, we have

$$
\begin{aligned}
f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= \boldsymbol{a}^\top(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) + b \\
&= \lambda(\boldsymbol{a}^\top \boldsymbol{x}) + (1-\lambda)(\boldsymbol{a}^\top \boldsymbol{y}) + \lambda b + (1-\lambda) b \\
&= \lambda(\boldsymbol{a}^\top \boldsymbol{x} + b) + (1-\lambda)(\boldsymbol{a}^\top \boldsymbol{y} + b) \\
&= \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}).
\end{aligned}
$$

  Therefore, f is both convex and concave.

- **Convexity of norms:** Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Then $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ is convex. Indeed, $\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\lambda \in [0, 1]$, we have

$$
\begin{aligned}
f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= \|\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\| \\
&\leq \|\lambda \boldsymbol{x}\| + \|(1-\lambda)\boldsymbol{y}\| \\
&= \lambda\|\boldsymbol{x}\| + (1-\lambda)\|\boldsymbol{y}\| \\
&= \lambda f(\boldsymbol{x}) + (1-\lambda) f\boldsymbol{y}).
\end{aligned}
$$

Jensen's inequality

Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a convex function, where C is a convex set. Then for any $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k \in C$ and $\boldsymbol{\lambda} \in \Delta_k$, the following inequality holds:

$$f\Big(\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i\Big) \leq \sum_{i=1}^{k} \lambda_i f(\boldsymbol{x}_i).$$

Proof: We will prove the inequality by induction on k. The inequality holds for the simplest case $k = 1$. Assume that for any k vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k \in C$ and $\boldsymbol{\lambda} \in \Delta_k$, the inequality holds. We will now prove the inequality for $k+1$ vectors. Assume that $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{k+1} \in C$ and $\boldsymbol{\lambda} \in \Delta_{k+1}$. If $\lambda_{k+1} = 1$ then $\lambda_i = 0$ for $1 \leq i \leq k$ and the inequality holds. If $\lambda_{k+1} < 1$ then

$$\begin{aligned} f\Big(\sum_{i=1}^{k+1} \lambda_i \boldsymbol{x}_i\Big) &= f\Big(\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i + \lambda_{k+1} \boldsymbol{x}_{k+1}\Big) = f\Big((1-\lambda_{k+1}) \underbrace{\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} \boldsymbol{x}_i}_{\boldsymbol{v}} + \lambda_{k+1} \boldsymbol{x}_{k+1}\Big) \\ &\leq (1-\lambda_{k+1}) f(\boldsymbol{v}) + \lambda_{k+1} f(\boldsymbol{x}_{k+1}), \end{aligned}$$

where $\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} = 1$ and then $\boldsymbol{v} \in C$. Since $f(\boldsymbol{v}) \leq \sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} f(\boldsymbol{x}_i)$, we obtain

$$f\Big(\sum_{i=1}^{k+1} \lambda_i \boldsymbol{x}_i\Big) \leq \sum_{i=1}^{k+1} \lambda_i f(\boldsymbol{x}_i). \quad \square$$

The gradient inequality

Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function defined on the convex set C. Then f is convex over C if and only if

$$f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) \le f(\boldsymbol{y}),\ \forall\, \boldsymbol{x}, \boldsymbol{y} \in C.$$

Proof:
($\Rightarrow$): Let $\boldsymbol{x} \neq \boldsymbol{y} \in C$ and $\lambda \in (0,1)$. We have $f(\lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{x}) \le \lambda f(\boldsymbol{y}) + (1-\lambda) f(\boldsymbol{x})$. Hence,

$$\frac{f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{\lambda} = \frac{f(\lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{x}) - f(\boldsymbol{x})}{\lambda} \le f(\boldsymbol{y}) - f(\boldsymbol{x}).$$

Taking $\lambda \to 0^+$, the left-hand side converges to the directional derivative of f at $\boldsymbol{x}$ in the direction $\boldsymbol{y} - \boldsymbol{x}$, so that $\nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) = f'(\boldsymbol{x}; \boldsymbol{y} - \boldsymbol{x}) \le f(\boldsymbol{y}) - f(\boldsymbol{x})$.

($\Leftarrow$): Let $\boldsymbol{z}, \boldsymbol{w} \in C$, $\lambda \in (0,1)$. We will show $f(\lambda \boldsymbol{z} + (1-\lambda)\boldsymbol{w}) \le \lambda f(\boldsymbol{z}) + (1-\lambda) f(\boldsymbol{w})$. Let $\boldsymbol{u} := \lambda \boldsymbol{z} + (1-\lambda)\boldsymbol{w} \in C$. Then $\boldsymbol{z} - \boldsymbol{u} = -\frac{1-\lambda}{\lambda}(\boldsymbol{w} - \boldsymbol{u})$. Invoking the gradient inequality on the pairs $\boldsymbol{z}, \boldsymbol{u}$ and $\boldsymbol{w}, \boldsymbol{u}$, we have

$$f(\boldsymbol{u}) + \nabla f(\boldsymbol{u})^\top (\boldsymbol{z} - \boldsymbol{u}) \le f(\boldsymbol{z}) \quad \text{and} \quad f(\boldsymbol{u}) - \frac{\lambda}{1-\lambda} \nabla f(\boldsymbol{u})^\top (\boldsymbol{z} - \boldsymbol{u}) \le f(\boldsymbol{w}).$$

Multiplying the first inequality by $\frac{\lambda}{1-\lambda}$ and adding it to the second one, we obtain

$$\frac{1}{1-\lambda} f(\boldsymbol{u}) \le \frac{\lambda}{1-\lambda} f(\boldsymbol{z}) + f(\boldsymbol{w}).$$

That is, $f(\lambda \boldsymbol{z} + (1-\lambda)\boldsymbol{w}) \le \lambda f(\boldsymbol{z}) + (1-\lambda) f(\boldsymbol{w})$. $\square$

The gradient inequality for strictly convex function

- *Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function defined on the convex set C. Then f is strictly convex over C if and only if*

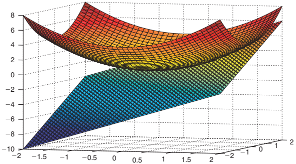
$$f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) < f(\boldsymbol{y}), \ \forall\, \boldsymbol{x} \neq \boldsymbol{y} \in C.$$

Proof:

($\Leftarrow$): It is similar to the proof on the previous page.

($\Rightarrow$): Suppose not, then $\exists\, \boldsymbol{x} \neq \boldsymbol{y} \in C$ such that $f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) = f(\boldsymbol{y})$. Let $\lambda \in (0,1)$ and $\boldsymbol{z} := \lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}$. Then, since f is strictly convex, we have $f(\boldsymbol{z}) < \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) = \lambda f(\boldsymbol{x}) + (1-\lambda)\big(f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x})\big) =$
$f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top \big(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} - \boldsymbol{x}\big) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top \big(\boldsymbol{z} - \boldsymbol{x}\big)$, a contradiction!

- Geometrically, the gradient inequality essentially states that for convex functions, the tangent plane is below the surface of f.



$f(x,y) = x^2 + y^2$ and its tangent plane at $(1,1)$.

A global minimizer

- *Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable convex function over the convex set C. Assume that $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ for some $\boldsymbol{x}^* \in C$. Then $\boldsymbol{x}^*$ is a global minimum point of f over C.*

  Proof: Let $\boldsymbol{z} \in C$. Then from the gradient inequality, we have

  $$f(\boldsymbol{z}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{z} - \boldsymbol{x}^*) = f(\boldsymbol{x}^*).$$

  That is, $\boldsymbol{x}^*$ is a global minimum point of f over C. $\square$

- *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable convex function. Then $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ if and only if $\boldsymbol{x}^*$ is a global minimum point of f over $\mathbb{R}^n$.*

  Proof: Using the above theorem and Theorem 2.6, we reach the conclusion. $\square$

Theorem 2.6: *Let $f : U \to \mathbb{R}$ be a function defined on a set $\varnothing \neq U \subseteq \mathbb{R}^n$. Assume that $\boldsymbol{x}^* \in int(U)$ is a local optimum point and that all the partial derivatives of f exist at $\boldsymbol{x}^*$. Then $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.* (Fermat's theorem in 1D)

Convexity and strict convexity of quadratic functions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be the quadratic function, $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + 2\boldsymbol{b}^\top \boldsymbol{x} + c$, where $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is symmetric, $\boldsymbol{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Then f is (strictly) convex if and only if $\boldsymbol{A} \succeq \boldsymbol{0}$ ($\boldsymbol{A} \succ \boldsymbol{0}$).

Proof:

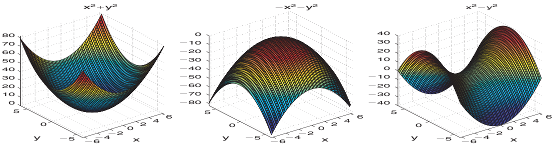
- The convexity of f is equivalent to the validity of the gradient inequality (see page 5): $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}), \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. That is,

$$\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y} + 2\boldsymbol{b}^\top \boldsymbol{y} + c \geq \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + 2\boldsymbol{b}^\top \boldsymbol{x} + c + 2(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^\top(\boldsymbol{y}-\boldsymbol{x}), \quad \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,$$

  which is equivalent to

$$(\boldsymbol{y}-\boldsymbol{x})^\top \boldsymbol{A}(\boldsymbol{y}-\boldsymbol{x}) \geq 0,\ \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \iff \boldsymbol{d}^\top \boldsymbol{A}\boldsymbol{d} \geq 0,\ \forall\, \boldsymbol{d} \in \mathbb{R}^n \iff \boldsymbol{A} \succeq \boldsymbol{0}.$$

- To prove the strict convexity variant, note that strict convexity of f is the same as $f(\boldsymbol{y}) > f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}), \forall\, \boldsymbol{x} \neq \boldsymbol{y} \in \mathbb{R}^n$. The same arguments as above imply that this is equivalent to $\boldsymbol{A} \succ \boldsymbol{0}$. $\square$



The left quadratic function $f(x,y) = x^2 + y^2$ is convex, while the middle $f(x,y) = -(x^2+y^2)$ and right $f(x,y) = x^2 - y^2$ are nonconvex.

Monotonicity of the gradient

Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function defined on the convex set C. Then f is convex over C if and only if

$$(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))^\top (\boldsymbol{x} - \boldsymbol{y}) \geq 0, \ \forall \ \boldsymbol{x}, \boldsymbol{y} \in C.$$

Proof:
($\Rightarrow$): Since f is convex over C, by the gradient inequality, we have $\forall \ \boldsymbol{x}, \boldsymbol{y} \in C$,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y}), \quad f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}),$$

By summing the two inequalities, we have $(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))^\top (\boldsymbol{x} - \boldsymbol{y}) \geq \mathbf{0}$.
($\Leftarrow$): Let g be the one-dimensional function defined by $g(t) = f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))$, $t \in [0, 1]$.
By the fundamental theorem of calculus, we have the gradient inequality,

$$\begin{aligned} f(\mathbf{y}) = g(1) = g(0) + \int_0^1 g'(t)\,dt \\ = f(\mathbf{x}) + \int_0^1 (\mathbf{y} - \mathbf{x})^T \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))\,dt \\ = f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \int_0^1 (\mathbf{y} - \mathbf{x})^T (\nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}))\,dt \\ \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}), \end{aligned}$$

where the last inequality follows from the monotonicity of ∇f that

$$t(\boldsymbol{y} - \boldsymbol{x})^\top (\nabla f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})) - \nabla f(\boldsymbol{x})) = (\nabla f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})) - \nabla f(\boldsymbol{x}))^\top (\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}) - \boldsymbol{x}) \geq 0. \quad \square$$

Second order characterization of convexity

Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function over an open convex set C. Then f is convex over $C \Leftrightarrow \nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0}, \forall\, \boldsymbol{x} \in C$.

Proof:

($\Rightarrow$): Let $\boldsymbol{x} \in C$ and $\mathbf{0} \neq \boldsymbol{y} \in \mathbb{R}^n$. Since C is open, $\exists\, \varepsilon > 0$ s.t. $\boldsymbol{x} + \lambda \boldsymbol{y} \in C$ for $0 < \lambda < \varepsilon$. Invoking the gradient inequality, we have

$$f(\boldsymbol{x} + \lambda \boldsymbol{y}) \geq f(\boldsymbol{x}) + \lambda \nabla f(\boldsymbol{x})^\top \boldsymbol{y}.$$

By the quadratic approximation theorem, we have

$$f(\boldsymbol{x} + \lambda \boldsymbol{y}) = f(\boldsymbol{x}) + \lambda \nabla f(\boldsymbol{x})^\top \boldsymbol{y} + \frac{\lambda^2}{2} \boldsymbol{y}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{y} + o(\lambda^2 \|\boldsymbol{y}\|^2),$$

which implies that

$$\frac{\lambda^2}{2} \boldsymbol{y}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{y} + o(\lambda^2 \|\boldsymbol{y}\|^2) \geq 0 \quad \Longrightarrow \quad \frac{1}{2} \boldsymbol{y}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{y} + \frac{o(\lambda^2 \|\boldsymbol{y}\|^2)}{\lambda^2} \geq 0$$

for any $\lambda \in (0, \varepsilon)$. Taking $\lambda \to 0^+$, we can conclude that $\boldsymbol{y}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{y} \geq 0$, for any $\boldsymbol{y} \in \mathbb{R}^n$. Therefore, $\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0}$ for any $\boldsymbol{x} \in C$.

($\Leftarrow$): Suppose that $\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0}$ for any $\boldsymbol{x} \in C$. We will prove the gradient inequality. Let $\boldsymbol{x}, \boldsymbol{y} \in C$. By the linear approximation theorem, $\exists\, \boldsymbol{z} \in [\boldsymbol{x}, \boldsymbol{y}] \subseteq C$ such that

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2} (\boldsymbol{y} - \boldsymbol{x})^\top \nabla^2 f(\boldsymbol{z}) (\boldsymbol{y} - \boldsymbol{x}).$$

Since $\nabla^2 f(\boldsymbol{z}) \succeq \mathbf{0}$, the gradient inequality $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x})$ holds. $\square$

Sufficient second order condition for strict convexity

- *Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function over a convex set C. Assume that $\nabla^2 f(\boldsymbol{x}) \succ \mathbf{0}, \forall\, \boldsymbol{x} \in C$, then f is strictly convex over C.*
- **Note:** *The positive definiteness of the Hessian is only a sufficient condition for strict convexity and is not necessary. Indeed, the function $f(x) = x^4$ is strictly convex, but its second order derivative $f''(x) = 12x^2$ is equal to zero for $x = 0$.*

Example: convexity of the log-sum-exp function

Consider the log-sum-exp function $f(\boldsymbol{x}) = \ln(e^{x_1} + e^{x_2} + \cdots + e^{x_n})$. The partial derivatives of f are given by

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}},\ i = 1, 2, \cdots, n, \quad \frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}) = \begin{cases} -\dfrac{e^{x_i} e^{x_j}}{(\sum_{k=1}^n e^{x_k})^2}, & i \neq j, \\ -\dfrac{e^{x_i} e^{x_i}}{(\sum_{k=1}^n e^{x_k})^2} + \dfrac{e^{x_i}}{\sum_{k=1}^n e^{x_k}}, & i = j. \end{cases}$$

The Hessian matrix is

$$\nabla^2 f(\boldsymbol{x}) = \operatorname{diag}(\boldsymbol{w}) - \boldsymbol{w}\boldsymbol{w}^\top, \quad w_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}, \quad \boldsymbol{w} \in \Delta_n.$$

Consider the expression for any $\mathbf{0} \neq \boldsymbol{v} \in \mathbb{R}^n$,

$$\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{v} = \sum_{i=1}^n w_i v_i^2 - (\boldsymbol{v}^\top \boldsymbol{w})^2.$$

Define vectors $\boldsymbol{s}, \boldsymbol{t}$ by $s_i = \sqrt{w_i} v_i$, $t_i = \sqrt{w_i}$, $i = 1, 2, \cdots, n$. Then by the Cauchy-Schwarz inequality, we have

$$(\boldsymbol{v}^\top \boldsymbol{w})^2 = (\boldsymbol{s}^\top \boldsymbol{t})^2 \leq \|\boldsymbol{s}\|^2 \|\boldsymbol{t}\|^2 = \left(\sum_{i=1}^n w_i v_i^2\right)\left(\sum_{i=1}^n w_i\right) = \left(\sum_{i=1}^n w_i v_i^2\right).$$

Therefore, $\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{v} \geq \mathbf{0}$ for any $\mathbf{0} \neq \boldsymbol{v} \in \mathbb{R}^n$.

Convexity under summation and multiplication

- *Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a convex function defined over a convex set C and let $\alpha \geq 0$. Then αf is a convex function over C.*

  Proof: Let $\boldsymbol{x}, \boldsymbol{y} \in C$ and $\lambda \in [0,1]$. Then

$$\begin{aligned}(\alpha f)(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= \alpha f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \\ &\leq \alpha \lambda f(\boldsymbol{x}) + \alpha(1-\lambda) f(\boldsymbol{y}) \\ &= \lambda(\alpha f)(\boldsymbol{x}) + (1-\lambda)(\alpha f)(\boldsymbol{y}). \quad \square\end{aligned}$$

- *Let $f_1, f_2, \cdots, f_p : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be convex functions over a convex set C. Then the sum function $f_1 + f_2 + \cdots + f_p$ is a convex over C.*

  Proof: Let $\boldsymbol{x}, \boldsymbol{y} \in C$ and $\lambda \in [0,1]$. Then

$$\begin{aligned}&(f_1 + f_2 + \cdots + f_p)(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \\ &= f_1(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) + \cdots + f_p(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \\ &\leq \lambda f_1(\boldsymbol{x}) + (1-\lambda) f_1(\boldsymbol{y}) + \cdots + \lambda f_p(\boldsymbol{x}) + (1-\lambda) f_p(\boldsymbol{y}) \\ &= \lambda(f_1(\boldsymbol{x}) + \cdots + f_p(\boldsymbol{x})) + (1-\lambda)(f_1(\boldsymbol{y}) + \cdots + f_p(\boldsymbol{y})) \\ &= \lambda(f_1 + f_2 + \cdots + f_p)(\boldsymbol{x}) + (1-\lambda)(f_1 + f_2 + \cdots + f_p)(\boldsymbol{y}). \quad \square\end{aligned}$$

Convexity under linear change of variables

Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a convex function defined on a convex set C. Let $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ and $\boldsymbol{b} \in \mathbb{R}^n$. Then the function g defined by $g(\boldsymbol{y}) := f(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b})$ is convex over the convex set $D := \{\boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{A}\boldsymbol{y} + \boldsymbol{b} \in C\}$.

Proof: First, D is convex because it is the *linear inverse image* of the convex set $C - \boldsymbol{b}$, i.e., $D = \boldsymbol{A}^{-1}(C - \boldsymbol{b})$ (cf. Chapter 6, page 7).

Let $\boldsymbol{y}_1, \boldsymbol{y}_2 \in D$. Define $\boldsymbol{x}_1 = \boldsymbol{A}\boldsymbol{y}_1 + \boldsymbol{b}$ and $\boldsymbol{x}_2 = \boldsymbol{A}\boldsymbol{y}_2 + \boldsymbol{b}$. Then $\boldsymbol{x}_1, \boldsymbol{x}_2 \in C$. Let $\lambda \in [0, 1]$. By the convexity of f we have

$$f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \le \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2).$$

That is,

$$f(\lambda(\boldsymbol{A}\boldsymbol{y}_1 + \boldsymbol{b}) + (1-\lambda)(\boldsymbol{A}\boldsymbol{y}_2 + \boldsymbol{b})) \le \lambda f(\boldsymbol{A}\boldsymbol{y}_1 + \boldsymbol{b}) + (1-\lambda) f(\boldsymbol{A}\boldsymbol{y}_2 + \boldsymbol{b}),$$

or equivalently,

$$f(\boldsymbol{A}(\lambda \boldsymbol{y}_1 + (1-\lambda)\boldsymbol{y}_2) + \boldsymbol{b}) \le \lambda f(\boldsymbol{A}\boldsymbol{y}_1 + \boldsymbol{b}) + (1-\lambda) f(\boldsymbol{A}\boldsymbol{y}_2 + \boldsymbol{b}).$$

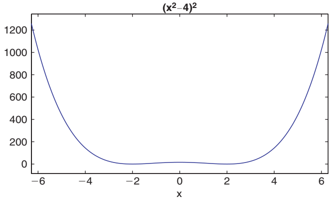
In other words, we have

$$g(\lambda \boldsymbol{y}_1 + (1-\lambda)\boldsymbol{y}_2) \le \lambda g(\boldsymbol{y}_1) + (1-\lambda) g(\boldsymbol{y}_2),$$

establishing the convexity of g. □

Composition of convex functions

- *Convexity is not preserved under composition of convex functions:* $g(t) = t^2$ and $h(t) = t^2 - 4$ are strictly convex. However, their composition $s(t) := g(h(t)) = (t^2 - 4)^2$ is not convex.



- *Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a convex function over a convex set C. Let $g : I \to \mathbb{R}$ be a one-dimensional nondecreasing convex function over the interval $I \subseteq \mathbb{R}$. Assume that $f(C) \subseteq I$. Then the composition of g with f defined by $h(\boldsymbol{x}) := g(f(\boldsymbol{x}))$, $\boldsymbol{x} \in C$, is a convex function over C.*

  Proof: Let $\boldsymbol{x}, \boldsymbol{y} \in C$ and $\lambda \in [0, 1]$. Then we have

$$\begin{aligned} h(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= g(f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y})) \leq g(\lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})) \\ &\leq \lambda g(f(\boldsymbol{x})) + (1-\lambda) g(f(\boldsymbol{y})) = \lambda h(\boldsymbol{x}) + (1-\lambda) h(\boldsymbol{y}). \quad \square \end{aligned}$$

Pointwise maximum of convex functions

- **Theorem:** *Let $f_1, f_2, \cdots, f_p : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be convex functions over the convex set C. Then the maximum function*

$$f(\boldsymbol{x}) := \max_{i=1,2,\cdots,p} f_i(\boldsymbol{x})$$

is a convex function over C.

Proof: Let $\boldsymbol{x}, \boldsymbol{y} \in C$ and $\lambda \in [0,1]$. Then we have

$$\begin{aligned} f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= \max_{i=1,2,\cdots,p} f_i(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \\ &\leq \max_{i=1,2,\cdots,p} \Big(\lambda f_i(\boldsymbol{x}) + (1-\lambda) f_i(\boldsymbol{y})\Big) \\ &\leq \lambda \max_{i=1,2,\cdots,p} f_i(\boldsymbol{x}) + (1-\lambda) \max_{i=1,2,\cdots,p} f_i(\boldsymbol{y}) \\ &= \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}). \quad \square \end{aligned}$$

- **Example:** $f(\boldsymbol{x}) := \max\{x_1, x_2, \cdots, x_n\}$ is a convex function $\because f$ is the maximum of n linear functions, which are convex.

Partial minimization

- **Theorem:** *Let $f : C \times D \subseteq \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ be a convex function over the set $C \times D$, where $C \subseteq \mathbb{R}^m$ and $D \subseteq \mathbb{R}^n$ are convex. Let*
$$g(\boldsymbol{x}) := \min_{\boldsymbol{y} \in D} f(\boldsymbol{x}, \boldsymbol{y}), \quad \boldsymbol{x} \in C,$$
where we assume that the minimum in the above definition is finite. Then g is convex over C.

  Proof: Let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in C$, $\lambda \in [0,1]$. Taking $\varepsilon > 0$, then $\exists\, \boldsymbol{y}_1, \boldsymbol{y}_2 \in D$ such that
$$f(\boldsymbol{x}_1, \boldsymbol{y}_1) \le g(\boldsymbol{x}_1) + \varepsilon, \quad f(\boldsymbol{x}_2, \boldsymbol{y}_2) \le g(\boldsymbol{x}_2) + \varepsilon.$$
  By the convexity of f we have
$$\begin{aligned} f(\lambda(\boldsymbol{x}_1, \boldsymbol{y}_1) + (1-\lambda)(\boldsymbol{x}_2, \boldsymbol{y}_2)) &\le \lambda f(\boldsymbol{x}_1, \boldsymbol{y}_1) + (1-\lambda) f(\boldsymbol{x}_2, \boldsymbol{y}_2) \\ &\le \lambda g(\boldsymbol{x}_1) + (1-\lambda) g(\boldsymbol{x}_2) + \varepsilon. \end{aligned}$$
  By the definition of g we can conclude that
$$g(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \le \lambda g(\boldsymbol{x}_1) + (1-\lambda) g(\boldsymbol{x}_2) + \varepsilon.$$
  Since the above inequality holds for any $\varepsilon > 0$, it follows that g is convex. □

- **Example:** *Let $C \subseteq \mathbb{R}^n$ be a convex set. The distance function defined by $d(\boldsymbol{x}, C) := \min\{\|\boldsymbol{x} - \boldsymbol{y}\| : \boldsymbol{y} \in C\}$ is convex since the function $f(\boldsymbol{x}, \boldsymbol{y}) := \|\boldsymbol{x} - \boldsymbol{y}\|$ is convex over $\mathbb{R}^n \times C$.*

Level sets of convex functions

- **Definition:** *Let $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function defined over the set S. Then the level set of f with level α is given by*

$$\mathrm{Lev}(f, \alpha) := \{\boldsymbol{x} \in S : f(\boldsymbol{x}) \leq \alpha\}.$$

- **Theorem:** *Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a convex function defined over a convex set C. Then for any $\alpha \in \mathbb{R}$ the level set $\mathrm{Lev}(f, \alpha)$ is convex.*

  Proof: Let $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{Lev}(f, \alpha)$ and $\lambda \in [0, 1]$. Then $f(\boldsymbol{x}), f(\boldsymbol{y}) \leq \alpha$. By the convexity of C, we have $\lambda\boldsymbol{x} + (1 - \lambda)\boldsymbol{y} \in C$ and then

$$f(\lambda\boldsymbol{x} + (1 - \lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1 - \lambda)f(\boldsymbol{y}) \leq \lambda\alpha + (1 - \lambda)\alpha = \alpha,$$

  which implies that $\lambda\boldsymbol{x} + (1 - \lambda)\boldsymbol{y} \in \mathrm{Lev}(f, \alpha)$. Therefore, the level set $\mathrm{Lev}(f, \alpha)$ is convex. $\square$

Example

Consider the following set:

$$D := \Big\{ \boldsymbol{x} \in \mathbb{R}^n : \overbrace{(\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + 1)^2 + \ln\Big(\sum_{i=1}^{n} e^{x_i}\Big)}^{f(\boldsymbol{x})} \le 3 \Big\},$$

where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{Q} \succeq \mathbf{0}$. The set D is convex as a level set of a convex function f. Specifically, $D = \mathrm{Lev}(f, 3)$.

The function f is indeed convex as the sum of two convex functions:

- *the log-sum-exp function, which was shown to be convex;*
- *and the function* $g(\boldsymbol{x}) := (\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + 1)^2$, *which is convex as a composition of the nondecreasing convex function* $\varphi(t) = (t+1)^2$ *defined on* $\mathbb{R}_+$ *with the convex quadratic function* $\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$.

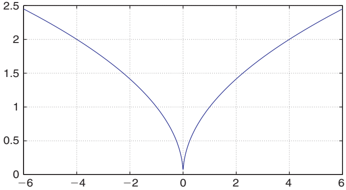
Quasi-convex functions

- *All convex functions defined over convex sets have convex level sets, but the reverse claim is not true. That is, there do exist nonconvex functions whose level sets are all convex.*
- **Definition:** *A function $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ defined over the convex set C is called quasi-convex if for any $\alpha \in \mathbb{R}$ the set $\text{Lev}(f, \alpha)$ is convex.*
- **Example:** The one-dimensional function $f(x) = \sqrt{|x|}$ is not convex, but its level sets are convex.

  For any $\alpha < 0$, we have $\text{Lev}(f, \alpha) = \varnothing$, and for any $\alpha \geq 0$,

  $$\text{Lev}(f, \alpha) = \{x \in \mathbb{R} : \sqrt{|x|} \leq \alpha\} = \{x \in \mathbb{R} : |x| \leq \alpha^2\} = [-\alpha^2, \alpha^2].$$

  Therefore, the nonconvex function f is quasi-convex.

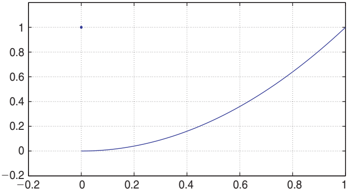


Continuity of convex functions

- Convex functions are not necessarily continuous when defined on non-open sets. Let us consider, for example, the function defined over the interval $[0, 1]$,

$$f(x) = \begin{cases} 1, & x = 0, \\ x^2, & 0 \leq x \leq 1, \end{cases}$$

It is easy to see that this is a convex function, and obviously it is not a continuous function.



- *We will prove that convex functions are always local Lipschitz continuous at interior points of their domain.*

Local Lipschitz continuity of convex functions

Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a convex function over the convex set C. Let $\boldsymbol{x}_0 \in \text{int}(C)$. Then there exist $\varepsilon > 0$ and $L > 0$ such that $B[\boldsymbol{x}_0, \varepsilon] \subseteq C$ and

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)| \leq L\|\boldsymbol{x} - \boldsymbol{x}_0\|, \quad \forall\, \boldsymbol{x} \in B[\boldsymbol{x}_0, \varepsilon].$$

Proof: Since $\boldsymbol{x}_0 \in \text{int}(C)$, $\exists\, \varepsilon > 0$ such that $B_\infty[\boldsymbol{x}_0, \varepsilon] := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{x}_0\|_\infty \leq \varepsilon\} \subseteq C$.

- **Claim:** f is upper bounded over $B_\infty[\boldsymbol{x}_0, \varepsilon]$.
  Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{2^n}$ be the 2^n extreme points of $B_\infty[\boldsymbol{x}_0, \varepsilon]$. These are the vectors $\boldsymbol{v}_i = \boldsymbol{x}_0 + \varepsilon \boldsymbol{w}_i$, where $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_{2^n}$ are the vectors in $\{-1, 1\}^n$. Then by the Krein-Milman theorem, for any $\boldsymbol{x} \in B_\infty[\boldsymbol{x}_0, \varepsilon]$, $\exists\, \boldsymbol{\lambda} \in \Delta_{2^n}$ such that $\boldsymbol{x} = \sum_{i=1}^{2^n} \lambda_i \boldsymbol{v}_i$. By Jensen's inequality, we have

$$f(\boldsymbol{x}) = f\Big(\sum_{i=1}^{2^n} \lambda_i \boldsymbol{v}_i\Big) \leq \sum_{i=1}^{2^n} \lambda_i f(\boldsymbol{v}_i) \leq M,$$

  where $M := \max_{i=1,\cdots,2^n} f(\boldsymbol{v}_i)$.

- **Claim:** $f(\boldsymbol{x}) \leq M$ for any $\boldsymbol{x} \in B[\boldsymbol{x}_0, \varepsilon]$.
  Since $\|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\|_2$ for any $\boldsymbol{x} \in \mathbb{R}^n$, it holds that

$$B_2[\boldsymbol{x}_0, \varepsilon] = B[\boldsymbol{x}_0, \varepsilon] = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{x}_0\|_2 \leq \varepsilon\} \subseteq B_\infty[\boldsymbol{x}_0, \varepsilon].$$

  Therefore, $f(\boldsymbol{x}) \leq M$ for any $\boldsymbol{x} \in B[\boldsymbol{x}_0, \varepsilon]$.

Continuing the proof

- **Claim:** $f(x) - f(x_0) \le L\|x - x_0\|$.
  Let $x \in B[x_0, \varepsilon]$ and $x \ne x_0$. Define $z := x_0 + \frac{1}{\alpha}(x - x_0)$, where $\alpha = \|x - x_0\|/\varepsilon$. Then $\alpha \le 1$, $z \in B[x_0, \varepsilon]$, and $f(z) \le M$. In addition, $x = \alpha z + (1-\alpha)x_0$. By the convexity of f, we have

$$f(x) \le \alpha f(z) + (1-\alpha)f(x_0) \le f(x_0) + \alpha(M - f(x_0)) = f(x_0) + \underbrace{\frac{M - f(x_0)}{\varepsilon}}_{:=L}\|x - x_0\|.$$

  Therefore, $f(x) - f(x_0) \le L\|x - x_0\|$.
- **Claim:** $f(x) - f(x_0) \ge -L\|x - x_0\|$.
  Define $u := x_0 + \frac{1}{\alpha}(x_0 - x)$. Then $\|u - x_0\| = \varepsilon$, $u \in B[x_0, \varepsilon]$, $f(u) \le M$, $x = x_0 + \alpha(x_0 - u)$, and

$$x_0 = \frac{1}{1+\alpha}(x_0 + \alpha(x_0 - u)) + \frac{\alpha}{1+\alpha}u.$$

  Therefore,

$$f(x_0) \le \frac{1}{1+\alpha}f\big(\overbrace{x_0 + \alpha(x_0 - u)}^{x}\big) + \frac{\alpha}{1+\alpha}f(u),$$

  which implies that

$$\begin{aligned} f(x) = f(x_0 + \alpha(x_0 - u)) &\ge f(x_0) + \alpha(f(x_0) - f(u)) \ge f(x_0) - \alpha(M - f(x_0)) \\ &= f(x_0) - \frac{M - f(x_0)}{\varepsilon}\|x - x_0\| = f(x_0) - L\|x - x_0\|, \end{aligned}$$

  and the desired result is established. □

Existence of directional derivatives for convex functions

Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a convex function over the convex set C and let $\boldsymbol{x} \in \mathrm{int}(C)$. Then for any $\mathbf{0} \neq \boldsymbol{d} \in \mathbb{R}^n$, $\displaystyle\lim_{t\to 0^+} \frac{f(\boldsymbol{x}+t\boldsymbol{d}) - f(\boldsymbol{x})}{t}$ exists.

Proof: Let $g(t) := f(\boldsymbol{x}+t\boldsymbol{d})$ and let $h(t) := \dfrac{g(t)-g(0)}{t}$. Then we wish to find $\displaystyle\lim_{t\to 0^+} h(t)$. Since $\boldsymbol{x} \in \mathrm{int}(C)$, $\exists\, \varepsilon > 0$ s.t. $\boldsymbol{x}+t\boldsymbol{d}, \boldsymbol{x}-t\boldsymbol{d} \in C$ for all $t \in [0,\varepsilon]$. Let $0 < t_1 < t_2 \leq \varepsilon$. Then

$$\boldsymbol{x}+t_1\boldsymbol{d} = (1-\frac{t_1}{t_2})\boldsymbol{x} + \frac{t_1}{t_2}(\boldsymbol{x}+t_2\boldsymbol{d}).$$

By the convexity of f we have

$$f(\boldsymbol{x}+t_1\boldsymbol{d}) \leq (1-\frac{t_1}{t_2})f(\boldsymbol{x}) + \frac{t_1}{t_2}f(\boldsymbol{x}+t_2\boldsymbol{d}) \quad\Longrightarrow\quad \frac{f(\boldsymbol{x}+t_1\boldsymbol{d})-f(\boldsymbol{x})}{t_1} \leq \frac{f(\boldsymbol{x}+t_2\boldsymbol{d})-f(\boldsymbol{x})}{t_2}.$$

Therefore, $h(t)$ is monotone nondecreasing over $(0,\varepsilon]$. Taking $0 < t \leq \varepsilon$, we have

$$\boldsymbol{x} = \frac{\varepsilon}{\varepsilon+t}(\boldsymbol{x}+t\boldsymbol{d}) + \frac{t}{\varepsilon+t}(\boldsymbol{x}-\varepsilon\boldsymbol{d}) \quad\Longrightarrow\quad f(\boldsymbol{x}) \leq \frac{\varepsilon}{\varepsilon+t}f(\boldsymbol{x}+t\boldsymbol{d}) + \frac{t}{\varepsilon+t}f(\boldsymbol{x}-\varepsilon\boldsymbol{d}).$$

After some rearrangement of terms, we obtain

$$h(t) = \frac{f(\boldsymbol{x}+t\boldsymbol{d})-f(\boldsymbol{x})}{t} \geq \frac{f(\boldsymbol{x})-f(\boldsymbol{x}-\varepsilon\boldsymbol{d})}{\varepsilon},$$

showing that h is bounded below over $(0,\varepsilon]$. Since h is nondecreasing and bounded below over $(0,\varepsilon]$, it follows that the limit $\displaystyle\lim_{t\to 0^+} h(t)$ exists. □

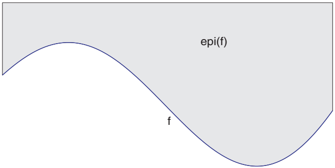
Extended real-valued functions

- It is natural to consider functions that are defined over the entire space $\mathbb{R}^n$ that take values in $\mathbb{R} \cup \{\infty\} = (-\infty, \infty]$. *Such a function is called an extended real-valued function.*
- The *indicator function $\delta_S : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$:* given a set $S \subseteq \mathbb{R}^n$,
$$\delta_S(\boldsymbol{x}) := \begin{cases} 0 & \text{if } \boldsymbol{x} \in S, \\ \infty & \text{if } \boldsymbol{x} \notin S. \end{cases}$$
- The *effective domain* of an extended real-valued function is
$$\text{dom}(f) := \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) < \infty\}.$$
- An extended real-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called *proper* if it is not always equal to ∞, meaning that there exists $\boldsymbol{x}_0 \in \mathbb{R}^n$ such that $f(\boldsymbol{x}_0) < \infty$.
- An extended real-valued function is *convex* if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\lambda \in [0, 1]$, we have $f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$, where we use the usual arithmetic with ∞:
$$a + \infty := \infty, \ \forall\, a \in \mathbb{R}, \quad a \cdot \infty := \infty, \ \forall\, a \in \mathbb{R}_{++}, \quad 0 \cdot \infty := 0.$$

Convexity of extended real-valued functions

- The above definition of convexity of extended real-valued functions is equivalent to saying that *(i)* $\mathrm{dom}(f)$ *is convex and (ii) the function* $g : \mathrm{dom}(f) \to \mathbb{R}$ *defined by* $g(\boldsymbol{x}) = f(\boldsymbol{x})$ *for any* $\boldsymbol{x} \in \mathrm{dom}(f)$ *is a convex finite-valued function over* $\mathrm{dom}(f)$*.*
- As an example, the indicator function $\delta_C(\cdot)$ of a set $C \subseteq \mathbb{R}^n$ is convex if and only if C is a convex set.
- Assume that $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. The epigraph set $\mathrm{epi}(f) \subseteq \mathbb{R}^{n+1}$ is defined by $\mathrm{epi}(f) := \left\{(\boldsymbol{x}, t)^\top : f(\boldsymbol{x}) \leq t\right\}$.

  Below is the epigraph of a one-dimensional function f.



- *An extended real-valued (or a real-valued) function* f *is convex if and only if its epigraph set* $\mathrm{epi}(f)$ *is convex.*

Convexity under maximum

- **Theorem:** *Let $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be an extended real-valued convex function for any $i \in I$, where I is an arbitrary index set. Then $f(\boldsymbol{x}) := \max_{i \in I} f_i(\boldsymbol{x})$ is an extended real-valued convex function.*

  Proof: $\mathrm{epi}(f_i)$ is convex for any $i \in I$ because f_i is convex. $\mathrm{epi}(f)$ is convex because $\mathrm{epi}(f) = \cap_{i \in I} \mathrm{epi}(f_i)$, intersection of convex sets. Hence, the convexity of f is established. $\square$

- **Definition of support function:** *Let $S \subseteq \mathbb{R}^n$. The support function of S is the function $\sigma_S(\boldsymbol{x}) := \max_{\boldsymbol{y} \in S} \boldsymbol{x}^\top \boldsymbol{y}$, $\forall\, \boldsymbol{x} \in \mathbb{R}^n$.*

  Since for each $\boldsymbol{y} \in S$, the function $f_{\boldsymbol{y}}(\boldsymbol{x}) := \boldsymbol{y}^\top \boldsymbol{x}$ is a convex function over $\mathbb{R}^n$, by the above theorem, it follows that the support function $\sigma_S(\boldsymbol{x})$ is an extended real-valued convex function.

- **Example:** Let $S = B[\mathbf{0}, 1] := \{\boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{y}\| \leq 1\}$. Let $\boldsymbol{x} \in \mathbb{R}^n$. We will show that $\sigma_S(\boldsymbol{x}) = \|\boldsymbol{x}\|$.

  Proof: If $\boldsymbol{x} = \mathbf{0}$ then $\sigma_S(\boldsymbol{x}) = 0 = \|\boldsymbol{x}\|$. If $\boldsymbol{x} \neq \mathbf{0}$ then $\forall\, \boldsymbol{y} \in S$ and $\boldsymbol{x} \in \mathbb{R}^n$, we have by the C-S inequality that $\boldsymbol{x}^\top \boldsymbol{y} \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\| \leq \|\boldsymbol{x}\|$. Taking $\widetilde{\boldsymbol{y}} = \boldsymbol{x} / \|\boldsymbol{x}\| \in S$, we have $\boldsymbol{x}^\top \widetilde{\boldsymbol{y}} = \|\boldsymbol{x}\|$. The desired formula follows. $\square$

Maxima of convex functions

- **Theorem:** *Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a convex function which is not constant over the convex set C. Then f does not attain a maximum at a point in $\text{int}(C)$.*
  Proof: Assume in contradiction that $\boldsymbol{x}^* \in \text{int}(C)$ is a global maximizer of f over C. Since f is not constant, $\exists\, \boldsymbol{y} \in C$ s.t. $f(\boldsymbol{y}) < f(\boldsymbol{x}^*)$. Since $\boldsymbol{x}^* \in \text{int}(C)$, $\exists\, \varepsilon > 0$ s.t. $\boldsymbol{z} = \boldsymbol{x}^* + \varepsilon(\boldsymbol{x}^* - \boldsymbol{y}) \in C$. Since $\boldsymbol{x}^* = \frac{\varepsilon}{\varepsilon+1}\boldsymbol{y} + \frac{1}{\varepsilon+1}\boldsymbol{z}$, $f(\boldsymbol{x}^*) \leq \frac{\varepsilon}{\varepsilon+1} f(\boldsymbol{y}) + \frac{1}{\varepsilon+1} f(\boldsymbol{z})$. Hence, $f(\boldsymbol{z}) \geq \varepsilon(f(\boldsymbol{x}^*) - f(\boldsymbol{y})) + f(\boldsymbol{x}^*) > f(\boldsymbol{x}^*)$, which is a contradiction. $\square$
- **Theorem:** *Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a convex and continuous function over the convex and compact set C. Then there exists at least one maximizer of f over C that is an extreme point of C.*
  Proof: By the Weierstrass theorem, the existence of maximizer of f over C is guaranteed. Let $\boldsymbol{x}^*$ be a maximizer of f over C. If $\boldsymbol{x}^*$ is an extreme point of C. Then the result is established. Otherwise, if $\boldsymbol{x}^*$ is not an extreme point, then by the Krein-Milman theorem, $C = \text{conv}(\text{ext}(C))$. $\exists\, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_k \in \text{ext}(C)$ and $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_k)^\top \in \Delta_k$, $\lambda_i > 0\ \forall\, i$, s.t. $\boldsymbol{x}^* = \sum_{i=1}^k \lambda_i \boldsymbol{x}_i$. By the convexity of f we have $f(\boldsymbol{x}^*) \leq \sum_{i=1}^k \lambda_i f(\boldsymbol{x}_i)$, or equivalently, $\sum_{i=1}^k \lambda_i (f(\boldsymbol{x}_i) - f(\boldsymbol{x}^*)) \geq 0$. Since $f(\boldsymbol{x}_i) \leq f(\boldsymbol{x}^*)$, $\forall\, i$, we have $f(\boldsymbol{x}_i) = f(\boldsymbol{x}^*)$, $\forall\, i$. Thus, the extreme points $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$ are all maximizers of f over C. $\square$

Computation of $\|A\|_{1,1}$

- Let $A \in \mathbb{R}^{m\times n}$. Recall that $\|A\|_{1,1} := \max\{\|\boldsymbol{A}\boldsymbol{x}\|_1 : \|\boldsymbol{x}\|_1 \leq 1\}$.
- Since the optimization problem consists of maximizing a convex and continuous function (composition of a norm function with a linear function) over a compact convex set, there exists a maximizer which is an extreme point of the ℓ_1 ball.
- Note that there are exactly $2n$ extreme points to the ℓ_1 ball, $\boldsymbol{e}_1$, $-\boldsymbol{e}_1, \boldsymbol{e}_2, -\boldsymbol{e}_2, \cdots, \boldsymbol{e}_n, -\boldsymbol{e}_n$.
- Since

$$\|\boldsymbol{A}\boldsymbol{e}_j\|_1 = \|\boldsymbol{A}(-\boldsymbol{e}_j)\|_1 = \sum_{i=1}^{m} |A_{ij}|,$$

we have

$$\|\boldsymbol{A}\|_{1,1} := \max_{1\leq j\leq n} \|\boldsymbol{A}\boldsymbol{e}_j\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |A_{ij}|.$$

The arithmetic geometric mean (AGM) inequality

AGM inequality: *For any $x_1, x_2, \cdots, x_n \geq 0$ the inequality holds:*

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n}.$$

More generally, for any $\boldsymbol{\lambda} \in \Delta_n$, one has $\sum_{i=1}^{n} \lambda_i x_i \geq \prod_{i=1}^{n} x_i^{\lambda_i}$.

Proof: Let $f(x) = -\ln(x)$. Then f is convex over $(0, \infty)$. For any $x_1, x_2, \cdots, x_n > 0$ and $\boldsymbol{\lambda} \in \Delta_n$, we have from Jensen's inequality that

$$f(\sum_{i=1}^{n} \lambda_i x_i) \leq \sum_{i=1}^{n} \lambda_i f(x_i) \Longrightarrow -\ln(\sum_{i=1}^{n} \lambda_i x_i) \leq -\sum_{i=1}^{n} \lambda_i \ln(x_i) \Longrightarrow \ln(\sum_{i=1}^{n} \lambda_i x_i) \geq \sum_{i=1}^{n} \lambda_i \ln(x_i).$$

Taking the exponent of both sides of the last inequality, we have

$$\sum_{i=1}^{n} \lambda_i x_i \geq \exp(\sum_{i=1}^{n} \lambda_i \ln(x_i)) = \prod_{i=1}^{n} x_i^{\lambda_i}.$$

Plugging in $\lambda_i = \frac{1}{n}$ for all i yields the special case. $\square$

Young's inequality and Hölder's inequality

- **Young's inequality:** *For any $s, t \geq 0$ and $p, q > 1$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$ it hold that $st \leq \frac{s^p}{p} + \frac{t^q}{q}$.*

  Proof: By the generalized AGM inequality we have for any $x, y \geq 0$, $x^{\frac{1}{p}} y^{\frac{1}{q}} \leq \frac{x}{p} + \frac{y}{q}$. Setting $x = s^p$ and $y = t^q$, the result follows. $\square$

- **Hölder's inequality:** *For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $p, q \geq 1$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$ it holds that $|\boldsymbol{x}^\top \boldsymbol{y}| \leq \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q$.*

  Proof: If $\boldsymbol{x} = \boldsymbol{0}$ or $\boldsymbol{y} = \boldsymbol{0}$ then the inequality is trivial. Assume that $\boldsymbol{x} \neq \boldsymbol{0}$ and $\boldsymbol{y} \neq \boldsymbol{0}$. The inequality is trivial for the cases of $(p, q) = (1, \infty)$ and $(p, q) = (\infty, 1)$. We assume that $1 < p, q < \infty$. For $1 \leq i \leq n$, setting $s = \frac{|x_i|}{\|\boldsymbol{x}\|_p}$ and $t = \frac{|y_i|}{\|\boldsymbol{y}\|_q}$ in Young's inequality yields

  $$\frac{|x_i y_i|}{\|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q} \leq \frac{1}{p} \frac{|x_i|^p}{\|\boldsymbol{x}\|_p^p} + \frac{1}{q} \frac{|y_i|^q}{\|\boldsymbol{y}\|_q^q}.$$

  Summing the above inequality over i, we have

  $$\frac{\sum_{i=1}^n |x_i y_i|}{\|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q} \leq \frac{1}{p} \frac{\sum_{i=1}^n |x_i|^p}{\|\boldsymbol{x}\|_p^p} + \frac{1}{q} \frac{\sum_{i=1}^n |y_i|^q}{\|\boldsymbol{y}\|_q^q} = \frac{1}{p} + \frac{1}{q} = 1.$$

  By the triangle inequality we have

  $$|\boldsymbol{x}^\top \boldsymbol{y}| \leq \sum_{i=1}^n |x_i y_i| \leq \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q. \quad \square$$

Minkowski's inequality

The p-norm (for $p \geq 1$) satisfies the triangle inequality: Let $p \geq 1$. Then for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\|\boldsymbol{x}+\boldsymbol{y}\|_p \leq \|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p$ holds.

Proof: The case $p = 1$ is trivial. We assume that $p > 1$, $\boldsymbol{x} \neq \mathbf{0}$, $\boldsymbol{y} \neq \mathbf{0}$, and $\boldsymbol{x}+\boldsymbol{y} \neq \mathbf{0}$. The function $\varphi(t) := t^p$ is convex over $\mathbb{R}_+$ since $\varphi''(t) = p(p-1)t^{p-2} > 0$ for $t > 0$. By the convexity, for any $\lambda_1, \lambda_2 \geq 0$, $\lambda_1 + \lambda_2 = 1$, we have

$$(\lambda_1 t + \lambda_2 s)^p \leq \lambda_1 t^p + \lambda_2 s^p.$$

Plugging

$$\lambda_1 = \frac{\|\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p}, \quad \lambda_2 = \frac{\|\boldsymbol{y}\|_p}{\|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p}, \quad t = \frac{|x_i|}{\|\boldsymbol{x}\|_p}, \quad s = \frac{|y_i|}{\|\boldsymbol{y}\|_p},$$

in the above inequality yields

$$\frac{1}{(\|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p)^p}(|x_i| + |y_i|)^p \leq \frac{\|\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p}\frac{|x_i|^p}{\|\boldsymbol{x}\|_p^p} + \frac{\|\boldsymbol{y}\|_p}{\|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p}\frac{|y_i|^p}{\|\boldsymbol{y}\|_p^p}.$$

Summing the above inequality over i, we obtain

$$\frac{1}{(\|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p)^p}\sum_{i=1}^{n}(|x_i| + |y_i|)^p \leq \frac{\|\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p} + \frac{\|\boldsymbol{y}\|_p}{\|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p} = 1.$$

Hence, $\|\boldsymbol{x}+\boldsymbol{y}\|_p^p = \sum_{i=1}^n |x_i + y_i|^p \leq \sum_{i=1}^n (|x_i| + |y_i|)^p \leq (\|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p)^p$. □