CS 2601 Linear and Convex Optimization

4. Convex functions (part 1)

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

Fall 2022

Outline

- Convex functions
- First-order condition for convexity
- Second-order condition for convexity

Convex functions

A function $f : S \subset \mathbb{R}^n \to \mathbb{R}$ is convex if

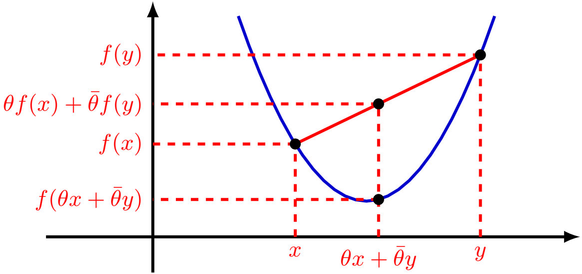
1. its domain $\operatorname{dom} f = S$ is a convex set
2. for any $\boldsymbol{x}, \boldsymbol{y} \in S$ and $\theta \in [0, 1]$, Jensen's inequality holds,

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

Note. Condition 1 guarantees $\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}$ is in the domain.

Note. We only need to check Condition 2 for $\boldsymbol{x} \neq \boldsymbol{y}$ and $\theta \in (0, 1)$.

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



Convex functions (cont'd)

A function $f : S \subset \mathbb{R}^n \to \mathbb{R}$ is strictly convex if

1. its domain $\operatorname{dom} f = S$ is a convex set
2. for any $\boldsymbol{x} \neq \boldsymbol{y} \in S$ and $\theta \in (0, 1)$,

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

Proposition. Let f be convex and $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$. Exactly one of the following holds,

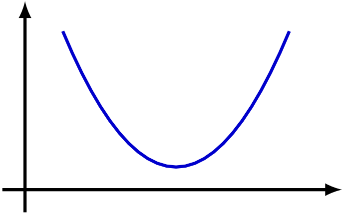
1. $f(\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}) = \theta f(\boldsymbol{x}) + \bar{\theta} f(\boldsymbol{y})$ for all $\theta \in [0, 1]$
2. $f(\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}) < \theta f(\boldsymbol{x}) + \bar{\theta} f(\boldsymbol{y})$ for all $\theta \in (0, 1)$

Strict convexity says the restriction of f to any line segment in S is not an affine function.

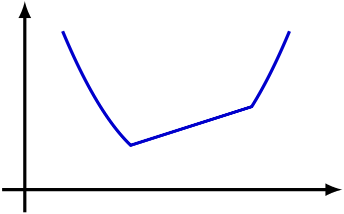
A function f is (strictly) concave if $-f$ is (strictly) convex.

An affine function $f(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + b$ is both convex and concave, but not strictly convex or strictly concave.

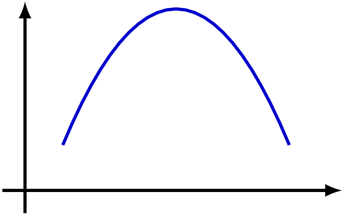
Convex functions (cont'd)



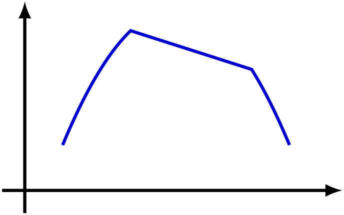
strictly convex function



convex function



strictly concave function



concave function

Examples

Example. Univariate functions

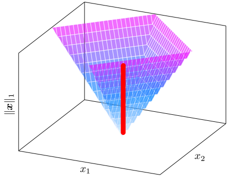
- $f(x) = e^{ax}$ ($a \in \mathbb{R}$) is convex, and strictly convex for $a \neq 0$
- $f(x) = \log x$ is strictly concave over $(0, \infty)$
- $f(x) = x^a$ is convex over $(0, \infty)$ for $a \geq 1$ or $a \leq 0$
- $f(x) = x^a$ is concave over $(0, \infty)$ for $0 \leq a \leq 1$

Example. Any norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is convex,

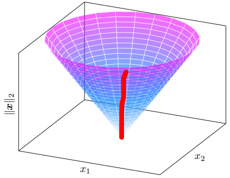
$$\|\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}\| \leq \|\theta \boldsymbol{x}\| + \|\bar{\theta} \boldsymbol{y}\| = \theta \|\boldsymbol{x}\| + \bar{\theta} \|\boldsymbol{y}\|$$

But not strictly convex (why?)

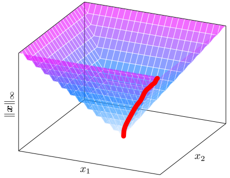
generator母线是直的，所以都不是严格的



1-norm



2-norm



∞-norm

Restriction to lines

Proposition. f is convex iff for any $\boldsymbol{x} \in \operatorname{dom} f$ and any direction $\boldsymbol{d}$, the function $g(t) = f(\boldsymbol{x} + t\boldsymbol{d})$ is convex on $\operatorname{dom} g = \{t : \boldsymbol{x} + t\boldsymbol{d} \in \operatorname{dom} f\}$.

Proof. "$\Rightarrow$". Assume f is convex. Fix an arbitrary $\boldsymbol{x} \in \operatorname{dom} f$ and direction $\boldsymbol{d}$. Need to show $g(t) = f(\boldsymbol{x} + t\boldsymbol{d})$ is convex.

Let $t_1, t_2 \in \operatorname{dom} g$, $\theta \in [0, 1]$.

1. Note $\boldsymbol{x} + (\theta t_1 + \bar{\theta} t_2)\boldsymbol{d} = \theta(\boldsymbol{x} + t_1\boldsymbol{d}) + \bar{\theta}(\boldsymbol{x} + t_2\boldsymbol{d})$. Let $\boldsymbol{x}_i = \boldsymbol{x} + t_i\boldsymbol{d}$.
2. $t_i \in \operatorname{dom} g \implies \boldsymbol{x}_i \in \operatorname{dom} f$
3. $\operatorname{dom} f$ is convex $\implies \boldsymbol{x} + (\theta t_1 + \bar{\theta} t_2)\boldsymbol{d} = \theta \boldsymbol{x}_1 + \bar{\theta}\boldsymbol{x}_2 \in \operatorname{dom} f$
   $\implies \theta t_1 + \bar{\theta} t_2 \in \operatorname{dom} g \implies \operatorname{dom} g$ is convex
4. Since f is convex,

$$g(\theta t_1 + \bar{\theta} t_2) = f(\theta \boldsymbol{x}_1 + \bar{\theta}\boldsymbol{x}_2) \leq \theta f(\boldsymbol{x}_1) + \bar{\theta} f(\boldsymbol{x}_2) = \theta g(t_1) + \bar{\theta} g(t_2)$$

   so g is convex.

Restriction to lines (cont'd)

Proof (cont'd). "$\Leftarrow$". Assume $g(t) = f(\boldsymbol{x} + t\boldsymbol{d})$ is convex for any $\boldsymbol{x} \in \operatorname{dom} f$ and any direction $\boldsymbol{d}$. Need to show f is convex.

Fix $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$, $\theta \in [0, 1]$.

1. Note $\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y} = \boldsymbol{y} + \theta(\boldsymbol{x} - \boldsymbol{y})$. Let $\boldsymbol{d} = \boldsymbol{x} - \boldsymbol{y}$, and $g(t) = f(\boldsymbol{y} + t\boldsymbol{d})$
2. $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f \implies 1, 0 \in \operatorname{dom} g$ 用这一步构造theta
3. $\operatorname{dom} g$ is convex $\implies \theta \in \operatorname{dom} g \implies \theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y} = \boldsymbol{y} + \theta \boldsymbol{d} \in \operatorname{dom} f$ $\implies \operatorname{dom} f$ is convex.
4. Since g is convex and $\theta = \theta \times 1 + \bar{\theta} \times 0$,

$$f(\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}) = g(\theta) \leq \theta g(1) + \bar{\theta} g(0) = \theta f(\boldsymbol{x}) + \bar{\theta} f(\boldsymbol{y})$$

   so f is convex.

Note. The same proof shows that f is strictly convex iff for any $\boldsymbol{x} \in \operatorname{dom} f$ and any direction $\boldsymbol{d} \neq \boldsymbol{0}$, the function $g(t) = f(\boldsymbol{x} + t\boldsymbol{d})$ is strictly convex on $\operatorname{dom} g = \{t : \boldsymbol{x} + t\boldsymbol{d} \in \operatorname{dom} f\}$.

Extended-value Extension

Given $f : S \subset \mathbb{R}^n \to \mathbb{R}$, its extended-value extension $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is defined by

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

如果要证明concave，就把正无穷变成负无穷

The (effective) domain of $\tilde{f}$, also the domain of f, is

$$\operatorname{dom} \tilde{f} = \operatorname{dom} f = S = \{\boldsymbol{x} : \tilde{f}(\boldsymbol{x}) < \infty\}$$

Proposition. f is convex iff $\tilde{f}$ is convex, i.e. $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\theta \in [0, 1]$,

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

with the usual extended arithmetic and ordering

$$a + \infty = \infty \text{ for } a > -\infty; \quad a \cdot \infty = \infty \text{ for } a > 0; \quad 0 \cdot \infty = 0$$

Note. We can similarly extend a concave function by $f(\boldsymbol{x}) = -\infty$ for $\boldsymbol{x} \notin \operatorname{dom} f$.

Outline

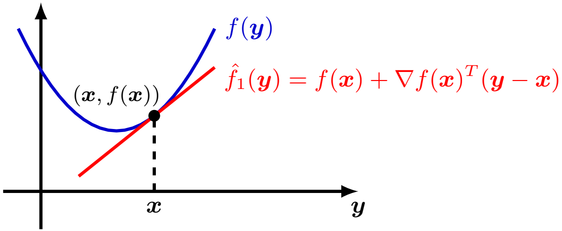
- Convex functions
- First-order condition for convexity
- Second-order condition for convexity

First-order condition for convexity

Theorem. A differentiable f with an open convex domain $\mathrm{dom}\, f$ is convex iff

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}), \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}\, f$$

Note. First-order Taylor approximation underestimates a convex function. Geometrically, all tangent "planes" lie below the graph.



Example. $e^x \geq e^0 + e^0(x - 0) = 1 + x$.

First-order condition for convexity (cont'd)

Proof. "$\Rightarrow$". Assume f is convex. Let $\boldsymbol{d} = \boldsymbol{y} - \boldsymbol{x}$. By convexity of f,

$$f(\boldsymbol{x} + t\boldsymbol{d}) = f(t\boldsymbol{y} + \bar{t}\boldsymbol{x}) \leq tf(\boldsymbol{y}) + \bar{t}f(\boldsymbol{x}), \quad t \in (0, 1)$$

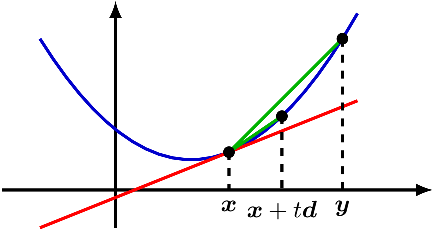
Rearranging,

$$\frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t} \leq f(\boldsymbol{y}) - f(\boldsymbol{x})$$

方向导数

Letting $t \to 0$,

$$\nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) = \nabla f(\boldsymbol{x})^T\boldsymbol{d} \leq f(\boldsymbol{y}) - f(\boldsymbol{x})$$



Note. $\frac{f(\boldsymbol{x}+t\boldsymbol{d})-f(\boldsymbol{x})}{t\|\boldsymbol{d}\|}$ is the slope of the secant line through $\boldsymbol{x}$ and $\boldsymbol{x} + t\boldsymbol{d}$.

First-order condition for convexity (cont'd)

Proof (cont'd). "$\Leftarrow$". Assume the first-order condition holds.

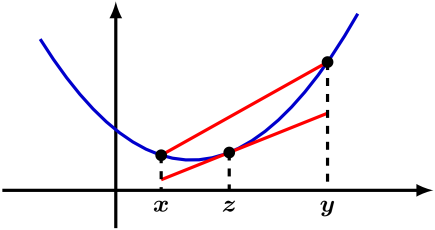
Let $z = \theta x + \bar{\theta} y$. The first-order condition implies

$$f(x) \geq f(z) + \nabla f(z)^T (x - z) \quad (1)$$

$$f(y) \geq f(z) + \nabla f(z)^T (y - z) \quad (2)$$

$\theta \times (1) + \bar{\theta} \times (2)$ yields

$$\theta f(x) + \bar{\theta} f(y) \geq f(z) = f(\theta x + \bar{\theta} y)$$



Note. Alternatively, using $g(t) = f(x + td)$, we can reduce the entire proof to the 1D case $g(t) \geq g(0) + g'(0)t$.

First-order condition for strict convexity

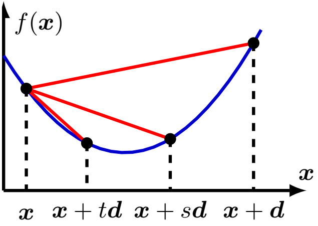
Theorem. A differentiable f with an open convex domain $\operatorname{dom} f$ is strictly convex iff

$$f(\boldsymbol{y}) > f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}), \quad \forall \boldsymbol{x} \neq \boldsymbol{y} \in \operatorname{dom} f$$

Proof. Essentially the same proof with inequalities being strict. The proof of "$\Rightarrow$" requires a further modification. Fix $\boldsymbol{x}$ and $\boldsymbol{d} = \boldsymbol{y} - \boldsymbol{x}$.

Add an intermediate point $\boldsymbol{x} + s\boldsymbol{d}$ between $\boldsymbol{x} + t\boldsymbol{d}$ and $\boldsymbol{x} + \boldsymbol{d}$. For $0 < t < s < 1$,

$$\begin{aligned} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t} &< \frac{f(\boldsymbol{x} + s\boldsymbol{d}) - f(\boldsymbol{x})}{s} \\ &< f(\boldsymbol{x} + \boldsymbol{d}) - f(\boldsymbol{x}), \end{aligned}$$



Now letting $t \to 0$ yield

$$\nabla f(\boldsymbol{x})^T\boldsymbol{d} \leq \frac{f(\boldsymbol{x} + s\boldsymbol{d}) - f(\boldsymbol{x})}{s} < f(\boldsymbol{x}+\boldsymbol{d}) - f(\boldsymbol{x}) = f(\boldsymbol{y}) - f(\boldsymbol{x})$$

First-order condition for univariate convex functions

Corollary. A differentiable $f : I \to \mathbb{R}$ defined on an open interval $I \subset \mathbb{R}$ is (strictly) convex iff f' is (strictly) increasing on I. 大于等于就行

Proof. We prove the convex case. The strictly convex case can be proved by replacing the inequalities by strict inequalities.

"$\Leftarrow$". Assume f' is increasing. Let $x, y \in I$ and $x < y$.

- By Mean Value Theorem, $\exists c \in (x, y)$ s.t. $f(y) - f(x) = f'(c)(y - x)$
- f' is increasing $\implies f'(c) \geq f'(x) \implies f(y) - f(x) \geq f'(x)(y - x)$
- The case $x > y$ is similar. By the first-order condition, f is convex.

"$\Rightarrow$". Assume f is convex. Let $x, y \in I$ and $x < y$. By the first-order condition,

$$f(y) \geq f(x) + f'(x)(y - x), \quad f(x) \geq f(y) + f'(y)(x - y)$$

Rearranging,

$$f'(x) \leq \frac{f(y) - f(x)}{y - x} \leq f'(y)$$

Optimality of stationary points

理解了一阶条件后这个就很好理解

Corollary. If $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ for a convex function f, then $\boldsymbol{x}^*$ is a global minimum. If f is strictly convex, then $\boldsymbol{x}^*$ is the unique global minimum.

Proof. By the first-order condition and the assumption $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^T (\boldsymbol{x} - \boldsymbol{x}^*) = f(\boldsymbol{x}^*), \quad \forall \boldsymbol{x} \in \operatorname{dom} f$$

so $\boldsymbol{x}^*$ is a global minimum.

Similarly, if f is strictly convex,

$$f(\boldsymbol{x}) > f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^T (\boldsymbol{x} - \boldsymbol{x}^*) = f(\boldsymbol{x}^*), \quad \forall \boldsymbol{x}^* \neq \boldsymbol{x} \in \operatorname{dom} f$$

so $\boldsymbol{x}^*$ is the unique global minimum.

Note. For concave functions, similar results hold with all inequalities reversed, and min replaced by max.

Outline

- Convex functions
- First-order condition for convexity
- Second-order condition for convexity

Second-order condition for convexity

Theorem. A twice differentiable $f : I \to \mathbb{R}$ defined on an open interval $I \subset \mathbb{R}$ is convex iff $f''(x) \geq 0$ for all $x \in I$.

Proof. f convex $\iff$ f' increasing on I $\iff$ $f''(x) \geq 0$ for all $x \in I$.

Theorem. A twice continuously differentiable f with an open convex domain $\operatorname{dom} f$ is convex iff $\nabla^2 f(\boldsymbol{x}) \succeq \boldsymbol{O}$ is positive semidefinite at every $\boldsymbol{x} \in \operatorname{dom} f$.

Proof. "$\Rightarrow$". Assume f is convex.

- $g(t) = f(\boldsymbol{x} + t\boldsymbol{d})$ is convex $\implies$ $g''(0) \geq 0$
- $\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{d} = g''(0) \geq 0$ for every $\boldsymbol{d}$ $\implies$ $\nabla^2 f(\boldsymbol{x}) \succeq \boldsymbol{O}$.

"$\Leftarrow$". Assume $\nabla^2 f(\boldsymbol{x}) \succeq \boldsymbol{O}$ for every $\boldsymbol{x} \in \operatorname{dom} f$.

- Let $g(t) = f(\boldsymbol{x} + t\boldsymbol{d})$, with $g''(t) = \boldsymbol{d}^T \nabla^2 f(\boldsymbol{x} + t\boldsymbol{d}) \boldsymbol{d}$
- $\nabla^2 f(\boldsymbol{x} + t\boldsymbol{d}) \succeq \boldsymbol{O}$ $\implies$ $g''(t) \geq 0$ for all $t \in \operatorname{dom} g$ $\implies$ $g(t)$ convex
- $g(t) = f(\boldsymbol{x} + t\boldsymbol{d})$ is convex for every $\boldsymbol{x} \in \operatorname{dom} f$ and $\boldsymbol{d}$ $\implies$ f convex

Second-order condition for convexity (cont'd)

Theorem. A twice continuously differentiable f with an open convex domain $\mathrm{dom} f$ is strictly convex if $\nabla^2 f(\boldsymbol{x})$ is positive definite at every $\boldsymbol{x} \in \mathrm{dom} f$.

Proof. Replace $\succeq$ and $\geq$ by $\succ$ and $>$ respectively in "$\Leftarrow$" part.

Note. Positive definiteness is sufficient but not necessary.

Example. $f(x) = x^4$ is strictly convex, but $f''(x) = 0$ at $x = 0$

Example. $f(\boldsymbol{x}) = f(x_1, x_2) = x_1^2 + x_2^4$ is strictly convex, but $\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{pmatrix}$ is not positive definite for $x_2 = 0$.

Note. For concave functions, replace "positive (semi)definite" by "negative (semi)definite" in previous theorems.

Examples

Example. Exponential $f(x) = e^{ax}$ is convex for $a \in \mathbb{R}$

Proof. $f''(x) = a^2 e^{ax} \geq 0$

Example. Logarithm $f(x) = \log x$ is strictly concave over $(0, \infty)$

Proof. $f''(x) = -x^{-2} < 0$

Example. Power $f(x) = x^a$ is convex over $(0, \infty)$ for $a \geq 1$ or $a \leq 0$, and concave over $(0, \infty)$ for $0 \leq a \leq 1$

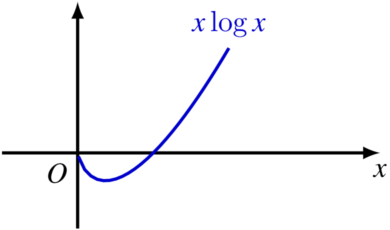
Proof. $f''(x) = a(a-1)x^{a-2} \gtreqless 0$ depending on a

Note. Domain is important. $f(x) = x^{-2}$ is concave over $(-\infty, 0)$ or $(0, \infty)$, but neither convex nor concave over $(-1, 0) \cup (0, 1)$.

Example: Negative entropy

The negative entropy $f(x) = x \log x$ is strictly convex over $(0, \infty)$.

Proof. $f'(x) = \log x + 1, f''(x) = x^{-1} > 0$



Note. We typically extend the definition of f to $x = 0$ by continuity, i.e.

$$f(0) \triangleq \lim_{x \to 0_+} f(x) = 0$$

f is still strictly convex with this extension.

Example: Quadratic functions

A quadratic function

$$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{x} + \boldsymbol{c}$$

with symmetric $\boldsymbol{Q}$ is convex iff $\boldsymbol{Q} \succeq \boldsymbol{O}$, and strictly convex iff $\boldsymbol{Q} \succ \boldsymbol{O}$.

Proof. For convexity, $\nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{Q}$ and use second-order condition.

For strict convexity, note $\nabla f(\boldsymbol{x}) = 2\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}$ and

$$f(\boldsymbol{x} + \boldsymbol{d}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{d} + \boldsymbol{d}^T \boldsymbol{Q} \boldsymbol{d}$$

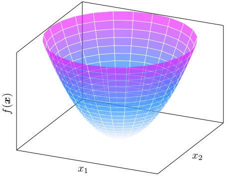
By first-order condition, f is strictly convex iff

$$\begin{aligned} f(\boldsymbol{x} + \boldsymbol{d}) > f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{d}, \quad \forall \boldsymbol{d} \neq \boldsymbol{0} &\iff \boldsymbol{d}^T \boldsymbol{Q} \boldsymbol{d} > 0, \quad \forall \boldsymbol{d} \neq \boldsymbol{0} \\ &\iff \boldsymbol{Q} \succ \boldsymbol{O} \end{aligned}$$

Note. Recall in general $\nabla^2 f(\boldsymbol{x}) \succ \boldsymbol{O}$ is not a necessary condition for strict convexity, but it is necessary when f is quadratic.

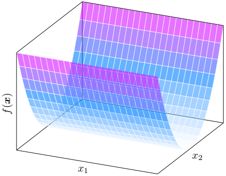
Example: Quadratic functions (cont'd)

Quadratic function $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$ in $\mathbb{R}^2$



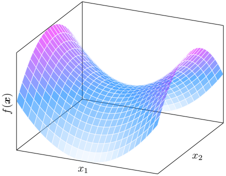
$\boldsymbol{Q} = \text{diag}\{1, 1\}$

strictly convex



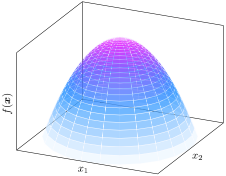
$\boldsymbol{Q} = \text{diag}\{0, 1\}$

convex



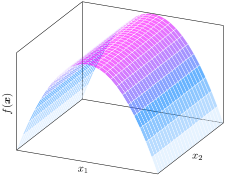
$\boldsymbol{Q} = \text{diag}\{1, -1\}$

neither convex nor concave



$\boldsymbol{Q} = \text{diag}\{-1, -1\}$

strictly concave



$\boldsymbol{Q} = \text{diag}\{-1, 0\}$

concave

Example: Least squares loss

The least squares loss

$$f(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$$

is always convex.

Proof. f is a quadratic function,

$$f(\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{y}^T\boldsymbol{y}.$$

$\boldsymbol{A}^T\boldsymbol{A} \succeq \boldsymbol{O}$ since

$$\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{A}\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \geq 0$$

Question. When is it strictly convex? 与线性方程组联系

Answer. When $\boldsymbol{A}^T\boldsymbol{A} \succ \boldsymbol{O}$, which is true iff $\boldsymbol{A}$ has full column rank.

$$\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = 0 \iff \|\boldsymbol{A}\boldsymbol{x}\|_2 = 0 \iff \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$$

Example: Log-sum-exp function

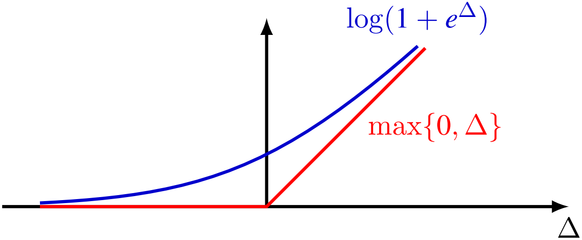
Log-sum-exp function defined below is convex

$$f(\boldsymbol{x}) = \log\left(\sum_{i=1}^{n} e^{x_i}\right)$$

Note. Also called “soft max”, as it smoothly approximates $\max\limits_{1\le i\le n} x_i$.

For $n = 2$, $\Delta = x_2 - x_1$,

$$\begin{aligned} f(x_1, x_2) &= \log(e^{x_1} + e^{x_2}) = \log[e^{x_1}(1 + e^{\Delta})] = x_1 + \log(1 + e^{\Delta}) \\ &\approx x_1 + \max\{0, \Delta\} = \max\{x_1, x_2\} \end{aligned}$$



Example: Log-sum-exp function (cont'd)

Proof. Show $\nabla^2 f(\boldsymbol{x}) \succeq \boldsymbol{O}, \forall \boldsymbol{x}$. Let $s(\boldsymbol{x}) = \sum_{k=1}^n e^{x_k}$, so $f(\boldsymbol{x}) = \log s(\boldsymbol{x})$.

$$g_i \triangleq \frac{\partial f(\boldsymbol{x})}{\partial x_i} = \frac{1}{s}\frac{\partial s}{\partial x_i} = \frac{e^{x_i}}{s}$$

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = \frac{\partial g_i}{\partial x_j} = \frac{e^{x_i}}{s}\delta_{ij} - \frac{e^{x_i}e^{x_j}}{s^2} = g_i\delta_{ij} - g_i g_j, \quad \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

For any $\boldsymbol{d} \in \mathbb{R}^n$,

$$\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{d} = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} d_i d_j = \sum_{i=1}^n g_i d_i^2 - \left(\sum_{i=1}^n g_i d_i \right)^2 \geq 0$$

where the last inequality follows from the fact $\sum_{i=1}^n g_i = 1$, and Cauchy-Schwarz inequality (or the convexity of $x \mapsto x^2$)