Strong convexity criteria. Optimality conditions. Lagrange function.

Daniil Merkulov

Optimization for ML. Faculty of Computer Science. HSE University

First-order differential criterion of convexity

The differentiable function $f(x)$ defined on the convex set $S \subseteq \mathbb{R}^n$ is convex if and only if $\forall x, y \in S$:

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

Let $y = x + \Delta x$, then the criterion will become more tractable:

$$f(x + \Delta x) \geq f(x) + \nabla f^T(x)\Delta x$$

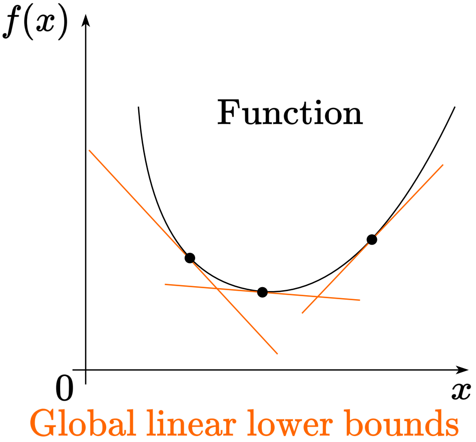

Figure 1: Convex function is greater or equal than Taylor linear approximation at any point

Second-order differential criterion of convexity

Twice differentiable function $f(x)$ defined on the convex set $S \subseteq \mathbb{R}^n$ is convex if and only if $\forall x \in \mathbf{int}(S) \neq \emptyset$:

$$\nabla^2 f(x) \succeq 0$$

In other words, $\forall y \in \mathbb{R}^n$:

$$\langle y, \nabla^2 f(x) y \rangle \geq 0$$

Connection with epigraph

The function is convex if and only if its epigraph is a convex set.

Example

Let a norm $\|\cdot\|$ be defined in the space U. Consider the set:

$$K := \{(x,t) \in U \times \mathbb{R}^+ : \|x\| \leq t\}$$

which represents the epigraph of the function $x \mapsto \|x\|$. This set is called the cone norm. According to the statement above, the set K is convex.
In the case where $U = \mathbb{R}^n$ and $\|x\| = \|x\|_2$ (Euclidean norm), the abstract set K transitions into the set:

$$\{(x,t) \in \mathbb{R}^n \times \mathbb{R}^+ : \|x\|_2 \leq t\}$$

Connection with sublevel set

If $f(x)$ - is a convex function defined on the convex set $S \subseteq \mathbb{R}^n$, then for any β sublevel set $\mathcal{L}_\beta$ is convex.

The function $f(x)$ defined on the convex set $S \subseteq \mathbb{R}^n$ is closed if and only if for any β sublevel set $\mathcal{L}_\beta$ is closed.

Reduction to a line

$f : S \to \mathbb{R}$ is convex if and only if S is a convex set and the function $g(t) = f(x + tv)$ defined on $\{t \mid x + tv \in S\}$ is convex for any $x \in S, v \in \mathbb{R}^n$, which allows checking convexity of the scalar function to establish convexity of the vector function.

Strong convexity

$f(x)$, **defined on the convex set** $S \subseteq \mathbb{R}^n$, is called μ-strongly convex (strongly convex) on S, if:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) - \frac{\mu}{2}\lambda(1-\lambda)\|x_1 - x_2\|^2$$

for any $x_1, x_2 \in S$ and $0 \leq \lambda \leq 1$ for some $\mu > 0$.

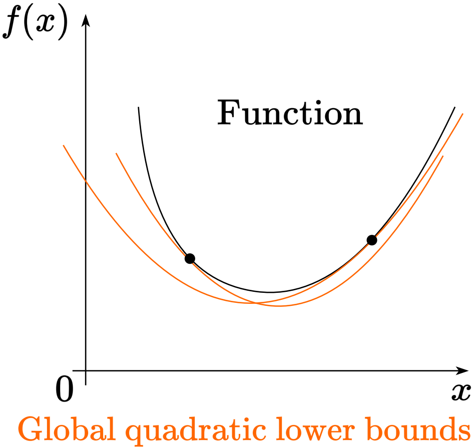

Figure 2: Strongly convex function is greater or equal than Taylor quadratic approximation at any point

First-order differential criterion of strong convexity

Differentiable $f(x)$ defined on the convex set $S \subseteq \mathbb{R}^n$ is μ-strongly convex if and only if $\forall x, y \in S$:

$$
f(y) \geq f(x) + \nabla f^T(x)(y - x) + \frac{\mu}{2}\|y - x\|^2
$$

Let $y = x + \Delta x$, then the criterion will become more tractable:

$$
f(x + \Delta x) \geq f(x) + \nabla f^T(x)\Delta x + \frac{\mu}{2}\|\Delta x\|^2
$$

Theorem

Let $f(x)$ be a differentiable function on a convex set $X \subseteq \mathbb{R}^n$. Then $f(x)$ is strongly convex on X with a constant $\mu > 0$ if and only if

$$
f(x) - f(x_0) \geq \langle \nabla f(x_0), x - x_0 \rangle + \frac{\mu}{2}\|x - x_0\|^2
$$

for all $x, x_0 \in X$.

Proof of first-order differential criterion of strong convexity

Necessity: Let $0 < \lambda \leq 1$. According to the definition of a strongly convex function,

$$
f(\lambda x + (1-\lambda)x_0) \leq \lambda f(x) + (1-\lambda)f(x_0) - \frac{\mu}{2}\lambda(1-\lambda)\|x - x_0\|^2
$$

or equivalently,

$$
f(x) - f(x_0) - \frac{\mu}{2}(1-\lambda)\|x - x_0\|^2 \geq \frac{1}{\lambda}[f(\lambda x + (1-\lambda)x_0) - f(x_0)] =
$$

$$
= \frac{1}{\lambda}[f(x_0 + \lambda(x - x_0)) - f(x_0)] = \frac{1}{\lambda}[\lambda\langle \nabla f(x_0), x - x_0\rangle + o(\lambda)] =
$$

$$
= \langle \nabla f(x_0), x - x_0\rangle + \frac{o(\lambda)}{\lambda}.
$$

Thus, taking the limit as $\lambda \downarrow 0$, we arrive at the initial statement.

Proof of first-order differential criterion of strong convexity

Sufficiency: Assume the inequality in the theorem is satisfied for all $x, x_0 \in X$. Take $x_0 = \lambda x_1 + (1-\lambda)x_2$, where $x_1, x_2 \in X$, $0 \leq \lambda \leq 1$. According to the inequality, the following inequalities hold:

$$f(x_1) - f(x_0) \geq \langle \nabla f(x_0), x_1 - x_0 \rangle + \frac{\mu}{2}\|x_1 - x_0\|^2,$$

$$f(x_2) - f(x_0) \geq \langle \nabla f(x_0), x_2 - x_0 \rangle + \frac{\mu}{2}\|x_2 - x_0\|^2.$$

Multiplying the first inequality by λ and the second by $1-\lambda$ and adding them, considering that

$$x_1 - x_0 = (1-\lambda)(x_1 - x_2), \quad x_2 - x_0 = \lambda(x_2 - x_1),$$

and $\lambda(1-\lambda)^2 + \lambda^2(1-\lambda) = \lambda(1-\lambda)$, we get

$$\lambda f(x_1) + (1-\lambda) f(x_2) - f(x_0) - \frac{\mu}{2}\lambda(1-\lambda)\|x_1 - x_2\|^2 \geq \\ \langle \nabla f(x_0), \lambda x_1 + (1-\lambda)x_2 - x_0 \rangle = 0.$$

Thus, inequality from the definition of a strongly convex function is satisfied. It is important to mention, that $\mu = 0$ stands for the convex case and corresponding differential criterion.

Second-order differential criterion of strong convexity

Twice differentiable function $f(x)$ defined on the convex set $S \subseteq \mathbb{R}^n$ is called μ-strongly convex if and only if $\forall x \in \mathbf{int}(S) \neq \emptyset$:

$$\nabla^2 f(x) \succeq \mu I$$

In other words:

$$\langle y, \nabla^2 f(x) y \rangle \geq \mu \|y\|^2$$

Theorem

Let $X \subseteq \mathbb{R}^n$ be a convex set, with int$X \neq \emptyset$. Furthermore, let $f(x)$ be a twice continuously differentiable function on X. Then $f(x)$ is strongly convex on X with a constant $\mu > 0$ if and only if

$$\langle y, \nabla^2 f(x) y \rangle \geq \mu \|y\|^2$$

for all $x \in X$ and $y \in \mathbb{R}^n$.

Proof of second-order differential criterion of strong convexity

The target inequality is trivial when $y = \mathbf{0}_n$, hence we assume $y \neq \mathbf{0}_n$.

Necessity: Assume initially that x is an interior point of X. Then $x + \alpha y \in X$ for all $y \in \mathbb{R}^n$ and sufficiently small α. Since $f(x)$ is twice differentiable,

$$f(x + \alpha y) = f(x) + \alpha \langle \nabla f(x), y \rangle + \frac{\alpha^2}{2} \langle y, \nabla^2 f(x) y \rangle + o(\alpha^2).$$

Based on the first order criterion of strong convexity, we have

$$\frac{\alpha^2}{2} \langle y, \nabla^2 f(x) y \rangle + o(\alpha^2) = f(x + \alpha y) - f(x) - \alpha \langle \nabla f(x), y \rangle \geq \frac{\mu}{2} \alpha^2 \|y\|^2.$$

This inequality reduces to the target inequality after dividing both sides by α^2 and taking the limit as $\alpha \downarrow 0$.

If $x \in X$ but $x \notin \text{int}X$, consider a sequence $\{x_k\}$ such that $x_k \in \text{int}X$ and $x_k \to x$ as $k \to \infty$. Then, we arrive at the target inequality after taking the limit.

Proof of second-order differential criterion of strong convexity

Sufficiency: Using Taylor's formula with the Lagrange remainder and the target inequality, we obtain for $x + y \in X$:

$$f(x+y) - f(x) - \langle \nabla f(x), y \rangle = \frac{1}{2} \langle y, \nabla^2 f(x + \alpha y) y \rangle \geq \frac{\mu}{2} \|y\|^2,$$

where $0 \leq \alpha \leq 1$. Therefore,

$$f(x+y) - f(x) \geq \langle \nabla f(x), y \rangle + \frac{\mu}{2} \|y\|^2.$$

Consequently, by the first order criterion of strong convexity, the function $f(x)$ is strongly convex with a constant μ. It is important to mention, that $\mu = 0$ stands for the convex case and corresponding differential criterion.

Facts

- $f(x)$ is called (strictly) concave, if the function $-f(x)$ - is (strictly) convex.

Facts

- $f(x)$ is called (strictly) concave, if the function $-f(x)$ - is (strictly) convex.
- Jensen's inequality for the convex functions:

$$
f\left(\sum_{i=1}^n \alpha_i x_i\right) \leq \sum_{i=1}^n \alpha_i f(x_i)
$$

for $\alpha_i \geq 0; \quad \sum\limits_{i=1}^n \alpha_i = 1$ (probability simplex)

For the infinite dimension case:

$$
f\left(\int\limits_S x p(x) dx\right) \leq \int\limits_S f(x) p(x) dx
$$

If the integrals exist and $p(x) \geq 0, \quad \int\limits_S p(x) dx = 1.$

Facts

- $f(x)$ is called (strictly) concave, if the function $-f(x)$ - is (strictly) convex.
- Jensen's inequality for the convex functions:

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i)$$

for $\alpha_i \geq 0; \quad \sum_{i=1}^{n} \alpha_i = 1$ (probability simplex)
For the infinite dimension case:

$$f\left(\int_{S} x p(x) dx\right) \leq \int_{S} f(x) p(x) dx$$

If the integrals exist and $p(x) \geq 0, \quad \int_{S} p(x) dx = 1.$

- If the function $f(x)$ and the set S are convex, then any local minimum $x^* = \text{arg}\min_{x \in S} f(x)$ will be the global one. Strong convexity guarantees the uniqueness of the solution.

Facts

- $f(x)$ is called (strictly) concave, if the function $-f(x)$ - is (strictly) convex.
- Jensen's inequality for the convex functions:

$$f\left(\sum_{i=1}^n \alpha_i x_i\right) \leq \sum_{i=1}^n \alpha_i f(x_i)$$

for $\alpha_i \geq 0; \quad \sum\limits_{i=1}^n \alpha_i = 1$ (probability simplex)
For the infinite dimension case:

$$f\left(\int\limits_S x p(x)dx\right) \leq \int\limits_S f(x)p(x)dx$$

If the integrals exist and $p(x) \geq 0, \quad \int\limits_S p(x)dx = 1$.

- If the function $f(x)$ and the set S are convex, then any local minimum $x^* = \text{arg}\min\limits_{x \in S} f(x)$ will be the global one. Strong convexity guarantees the uniqueness of the solution.
- Let $f(x)$ - be a convex function on a convex set $S \subseteq \mathbb{R}^n$. Then $f(x)$ is continuous $\forall x \in \mathbf{ri}(S)$.

Operations that preserve convexity

- Non-negative sum of the convex functions: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$.

Operations that preserve convexity

- Non-negative sum of the convex functions: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$.
- Composition with affine function $f(Ax + b)$ is convex, if $f(x)$ is convex.

Operations that preserve convexity

- Non-negative sum of the convex functions: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$.
- Composition with affine function $f(Ax + b)$ is convex, if $f(x)$ is convex.
- Pointwise maximum (supremum) of any number of functions: If $f_1(x), \ldots, f_m(x)$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex.

Operations that preserve convexity

- Non-negative sum of the convex functions: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$.
- Composition with affine function $f(Ax + b)$ is convex, if $f(x)$ is convex.
- Pointwise maximum (supremum) of any number of functions: If $f_1(x), \ldots, f_m(x)$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex.
- If $f(x, y)$ is convex on x for any $y \in Y$: $g(x) = \sup\limits_{y \in Y} f(x, y)$ is convex.

Operations that preserve convexity

- Non-negative sum of the convex functions: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$.
- Composition with affine function $f(Ax + b)$ is convex, if $f(x)$ is convex.
- Pointwise maximum (supremum) of any number of functions: If $f_1(x), \ldots, f_m(x)$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex.
- If $f(x, y)$ is convex on x for any $y \in Y$: $g(x) = \sup\limits_{y \in Y} f(x, y)$ is convex.
- If $f(x)$ is convex on S, then $g(x, t) = tf(x/t)$ - is convex with $x/t \in S, t > 0$.

Operations that preserve convexity

- Non-negative sum of the convex functions: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$.
- Composition with affine function $f(Ax+b)$ is convex, if $f(x)$ is convex.
- Pointwise maximum (supremum) of any number of functions: If $f_1(x), \ldots, f_m(x)$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex.
- If $f(x, y)$ is convex on x for any $y \in Y$: $g(x) = \sup\limits_{y \in Y} f(x, y)$ is convex.
- If $f(x)$ is convex on S, then $g(x, t) = tf(x/t)$ - is convex with $x/t \in S, t > 0$.
- Let $f_1 : S_1 \to \mathbb{R}$ and $f_2 : S_2 \to \mathbb{R}$, where range$(f_1) \subseteq S_2$. If f_1 and f_2 are convex, and f_2 is increasing, then $f_2 \circ f_1$ is convex on S_1.

Other forms of convexity

- Log-convexity: $\log f$ is convex; Log convexity implies convexity.

Other forms of convexity

- Log-convexity: $\log f$ is convex; Log convexity implies convexity.
- Log-concavity: $\log f$ concave; **not** closed under addition!

Other forms of convexity

- Log-convexity: $\log f$ is convex; Log convexity implies convexity.
- Log-concavity: $\log f$ concave; **not** closed under addition!
- Exponential convexity: $[f(x_i + x_j)] \succeq 0$, for $x_1, \ldots, x_n$

Other forms of convexity

- Log-convexity: $\log f$ is convex; Log convexity implies convexity.
- Log-concavity: $\log f$ concave; **not** closed under addition!
- Exponential convexity: $[f(x_i + x_j)] \succeq 0$, for $x_1, \dots, x_n$
- Operator convexity: $f(\lambda X + (1-\lambda)Y)$

Other forms of convexity

- Log-convexity: $\log f$ is convex; Log convexity implies convexity.
- Log-concavity: $\log f$ concave; **not** closed under addition!
- Exponential convexity: $[f(x_i + x_j)] \succeq 0$, for $x_1, \ldots, x_n$
- Operator convexity: $f(\lambda X + (1 - \lambda)Y)$
- Quasiconvexity: $f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}$

Other forms of convexity

- Log-convexity: $\log f$ is convex; Log convexity implies convexity.
- Log-concavity: $\log f$ concave; **not** closed under addition!
- Exponential convexity: $[f(x_i + x_j)] \succeq 0$, for $x_1, \ldots, x_n$
- Operator convexity: $f(\lambda X + (1-\lambda)Y)$
- Quasiconvexity: $f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$
- Pseudoconvexity: $\langle \nabla f(y), x - y \rangle \geq 0 \longrightarrow f(x) \geq f(y)$

Other forms of convexity

- Log-convexity: $\log f$ is convex; Log convexity implies convexity.
- Log-concavity: $\log f$ concave; **not** closed under addition!
- Exponential convexity: $[f(x_i + x_j)] \succeq 0$, for $x_1, \ldots, x_n$
- Operator convexity: $f(\lambda X + (1 - \lambda)Y)$
- Quasiconvexity: $f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}$
- Pseudoconvexity: $\langle \nabla f(y), x - y \rangle \geq 0 \longrightarrow f(x) \geq f(y)$
- Discrete convexity: $f : \mathbb{Z}^n \to \mathbb{Z}$; “convexity + matroid theory.”

Examples

Example

Show, that $f(x) = c^\top x + b$ is convex and concave.

Examples

Example

Show, that $f(x) = x^\top Ax$, where $A \succeq 0$ - is convex on $\mathbb{R}^n$.

Examples

Example

Show, that $f(A) = \lambda_{max}(A)$ - is convex, if $A \in S^n$.

PL inequality holds if the following condition is satisfied for some $\mu > 0$,

$$\|\nabla f(x)\|^2 \geq \mu(f(x) - f^*) \forall x$$

The example of a function, that satisfies the PL-condition, but is not convex.

$$f(x, y) = \frac{(y - \sin x)^2}{2}$$

Convex optimization problem

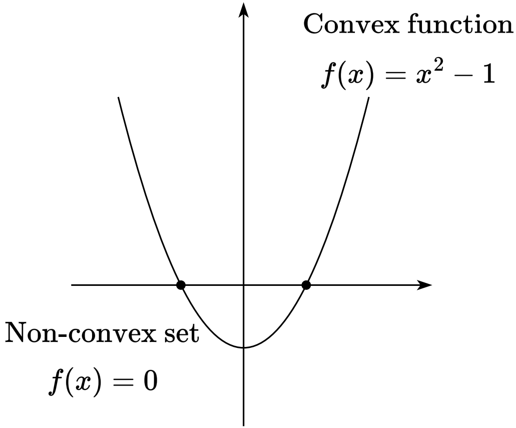

Figure 3: The idea behind the definition of a convex optimization problem

Note, that there is an agreement in notation of mathematical programming. The problems of the following type are called **Convex optimization problem**:

$$
\begin{aligned}
& f_0(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & f_i(x) \leq 0, \; i = 1, \ldots, m \\
& Ax = b,
\end{aligned}
\tag{COP}
$$

where all the functions $f_0(x), f_1(x), \ldots, f_m(x)$ are convex and all the equality constraints are affine. It sounds a bit strange, but not all convex problems are convex optimization problems.

$$
f_0(x) \to \min_{x \in S}, \tag{CP}
$$

where $f_0(x)$ is a convex function, defined on the convex set S. The necessity of affine equality constraint is essential.

Linear Least Squares

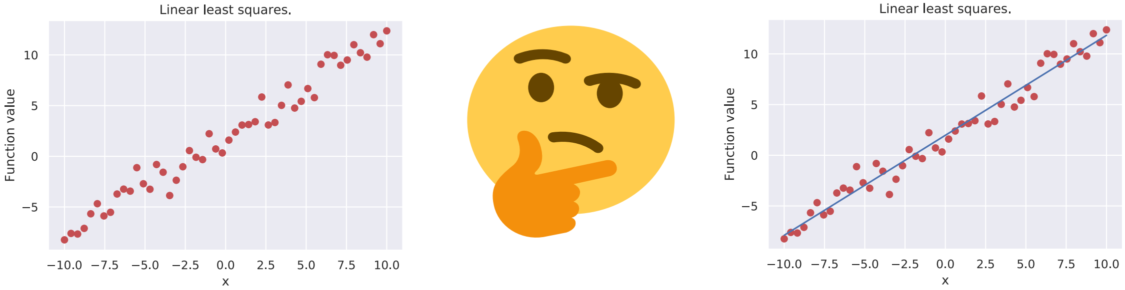

Figure 4: Illustration

Neural networks?