Strongly convex functions. Polyak - Lojasiewicz Condition. Conjugate sets

Daniil Merkulov

Optimization methods. MIPT

Strong convexity criteria

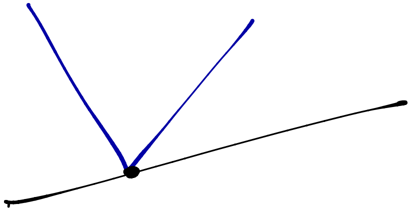
First-order differential criterion of convexity

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:

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



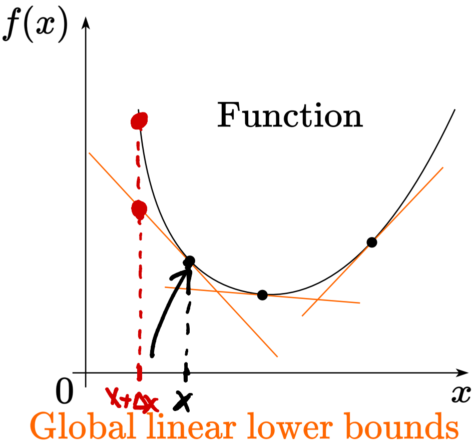


Figure 1: Convex function is greater or equal than Taylor linear approximation at any point

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 $\forall x \in \mathbf{int}(S) \neq \emptyset$:

$$\boxed{\nabla^2 f(x) \succeq 0}$$

$f(x) = \frac{1}{2} x^T x$

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

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

$\nabla^2 f = I$

$f(x) = \frac{1}{2} x^T A x$

$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $\quad f: \mathbb{R}^3 \to \mathbb{R}$

$0, 1, 2$

Tools for discovering convexity

- **Definition (Jensen's inequality)**

Tools for discovering convexity

- **Definition (Jensen's inequality)**
- **Differential criteria of convexity**

Tools for discovering convexity

- **Definition (Jensen's inequality)**
- **Differential criteria of convexity**
- **Operations, that preserve convexity**

Tools for discovering convexity

- **Definition (Jensen's inequality)**
- **Differential criteria of convexity**
- **Operations, that preserve convexity**
- **Connection with epigraph**

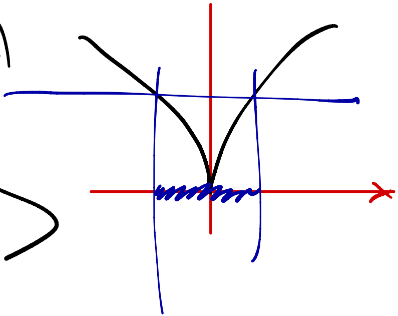
  The function is convex if and only if its epigraph is a convex set.

Tools for discovering convexity

- **Definition (Jensen's inequality)**
- **Differential criteria of convexity**
- **Operations, that preserve convexity**
- **Connection with epigraph**

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

$f(x)$ $\Longrightarrow$



- **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.

f - вып. $\Longleftarrow$ если $\forall \beta$ L_β - выпук

квази выпуклые функции

Tools for discovering convexity

- **Definition (Jensen's inequality)**
- **Differential criteria of convexity**
- **Operations, that preserve convexity**
- **Connection with epigraph**

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

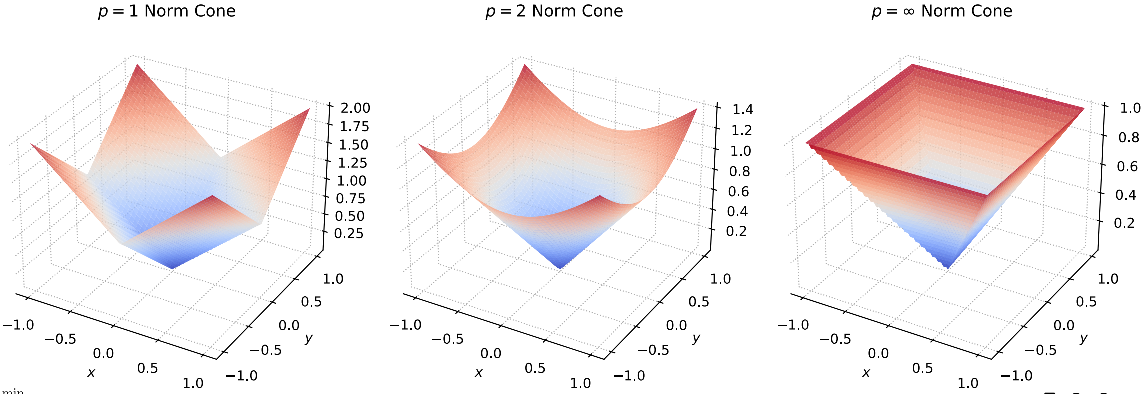
  $f : S \to \mathbb{R}$ is convex if and only if S is a convex set and the function $g(t) = f(x + tv)$ defined on $\{t \mid x + tv \in S\}$ is convex for any $x \in S, v \in \mathbb{R}^n$, which allows checking convexity of the scalar function to establish convexity of the vector function.

Example: norm cone

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

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

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. Code for the figures



Strong convexity

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

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

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

$\mu > 0$ сильная выпуклость

$\mu = 0$ – выпуклость

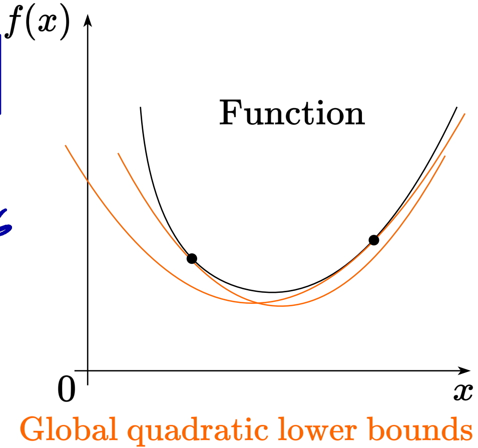


Figure 3: Strongly convex function is greater or equal than Taylor quadratic approximation at any point

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 $\forall x, y \in S$:

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

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:

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

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

$f(x) - f(x_0) \geq$

если f - μ сильно вып., то вып. утв. теорем:

$\geq \langle \nabla f(x_0), x - x_0 \rangle +$

$+ \frac{\mu}{2} \|x - x_0\|^2$

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

По опр.

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

or equivalently,

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

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

?

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

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

$\lim_{\lambda \to 0^+}$

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 $x_1, x_2 \in X$, $0 \le \lambda \le 1$. According to the inequality, the following inequalities hold:

Пусть $\forall x_1, x_2$: $f(x_1) - f(x_2) \geq \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{\mu}{2}\|x_1 - x_2\|^2 \Rightarrow f(x) - \mu$

сильно вып. функц

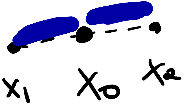
x_1, x_0 / x_2, x_0

$$
\begin{cases}
f(x_1) - f(x_0) \ge \langle \nabla f(x_0), x_1 - x_0 \rangle + \frac{\mu}{2}\|x_1 - x_0\|^2, & \cdot\, \lambda \\
f(x_2) - f(x_0) \ge \langle \nabla f(x_0), x_2 - x_0 \rangle + \frac{\mu}{2}\|x_2 - x_0\|^2. & (1-\lambda)
\end{cases}
\quad +
$$

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

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

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



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

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 $\forall x \in \textbf{int}(S) \neq \emptyset$:

$\mu = \lambda_{min}(\nabla^2 f)$

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

для выпуклости

$\nabla^2 f \succeq 0$

$\nabla^2 f \succeq \mu \cdot I$

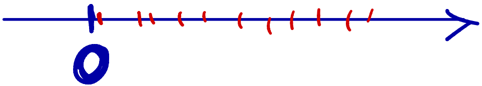
$\nabla^2 f - \mu I \succeq 0$

In other words:

$\mu > 0$ сильн. вып.

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

$\mu = 0$ вып.

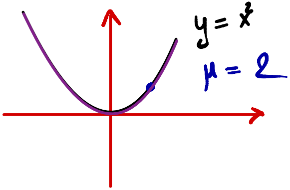


спектр $\nabla^2 f$

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 $\text{int}X \neq \emptyset$. 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

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

$\nabla^2 f(x) - \mu I \succeq 0$

$\mu > 0$

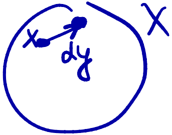
for all $x \in X$ and $y \in \mathbb{R}^n$.

Proof of second-order differential criterion of strong convexity

если f - μ сильно вып, то $\nabla^2 f \succeq \mu I$, $\mu > 0$

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,



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

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

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

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

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.

$\forall y \quad \langle y, \nabla^2 f(x) y \rangle \geq \frac{\mu}{2} \|y\|^2$

Proof of second-order differential criterion of strong convexity

по усл. лем $\nabla^2 f \succeq \mu I \Rightarrow$

x — $x+\alpha y$ — y

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,

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

по дифф. кр. I

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.

Convex and concave function

$\nabla^2 f = 0 \succeq 0$?

f - вып.

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

по опр f - вогн. (concave) выпукл, если $-f(x)$ вып.

Simplest strongly convex function

$$\nabla^2 f = 2A \succeq 0$$

$n > 1$

$A \succ 0$

Example

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

$$\nabla^2 f = a a^T$$

$\forall y \in \mathbb{R}^n$

$$y^T a a^T y = (a^T y)^T \cdot a^T y =$$

$$= p^T \cdot p \geq 0$$

Convexity and continuity

Let $f(x)$ - be a convex function on a convex set $S \subseteq \mathbb{R}^n$. Then $f(x)$ is continuous $\forall x \in \mathbf{ri}(S)$. [a]

Proper convex function

Function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be **proper convex function** if it never takes on the value $-\infty$ and not identically equal to ∞.

Indicator function

$$\delta_S(x) = \begin{cases} \infty, & x \in S, \\ 0, & x \notin S, \end{cases}$$

is a proper convex function.

[a] Please, read here about difference between interior and relative interior.

Convexity and continuity

Let $f(x)$ - be a convex function on a convex set $S \subseteq \mathbb{R}^n$. Then $f(x)$ is continuous $\forall x \in \textbf{ri}(S)$. [a]

Proper convex function

Function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be **proper convex function** if it never takes on the value $-\infty$ and not identically equal to ∞.

Indicator function

$$\delta_S(x) = \begin{cases} \infty, & x \in S, \\ 0, & x \notin S, \end{cases}$$

is a proper convex function.

[a] Please, read here about difference between interior and relative interior.

Closed function

Function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be **closed** if for each $\alpha \in \mathbb{R}$, the sublevel set is a closed set. Equivalently, if the epigraph is closed, then the function f is closed.

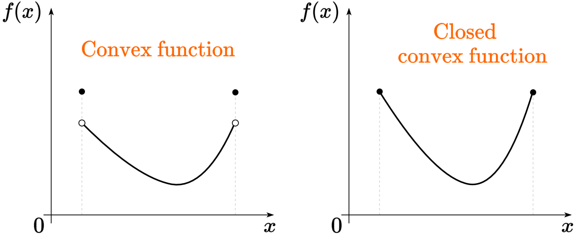


Figure 4: The concept of a closed function is introduced to avoid such breaches at the border.

Facts about convexity

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

Facts about convexity

- $f(x)$ is called (strictly, strongly) concave, if the function $-f(x)$ - is (strictly, strongly) 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 about convexity

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

Operations that preserve convexity

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

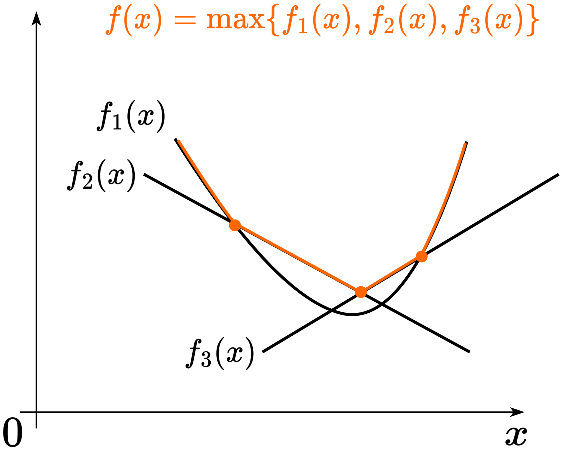


Figure 5: Pointwise maximum (supremum) of convex functions is convex

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.

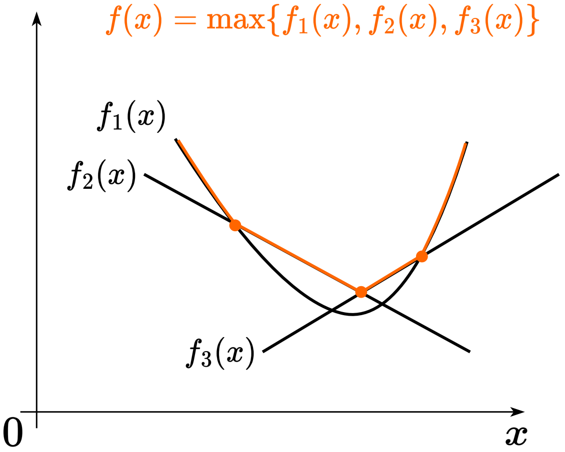


Figure 5: Pointwise maximum (supremum) of convex functions is convex

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.

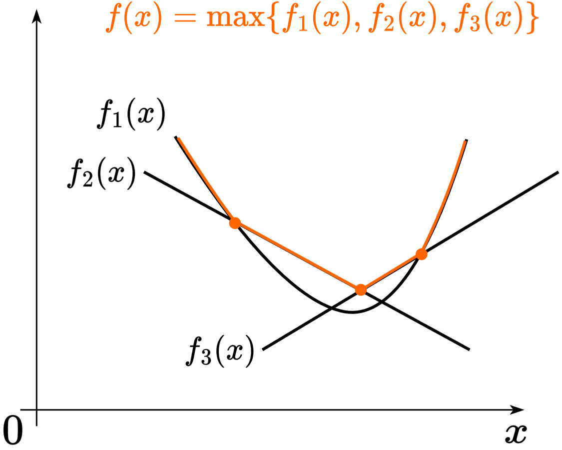


Figure 5: Pointwise maximum (supremum) of convex functions is convex

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.

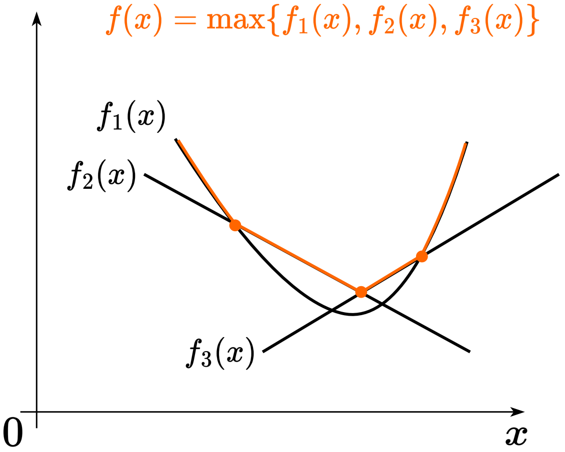


Figure 5: Pointwise maximum (supremum) of convex functions is convex

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 $x/t \in S, t > 0$.

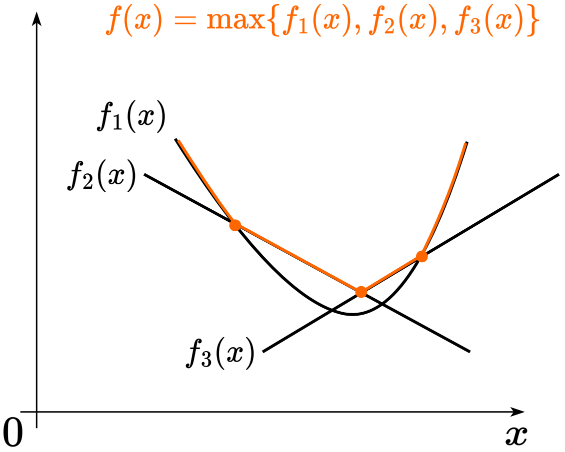


Figure 5: Pointwise maximum (supremum) of convex functions is convex

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 $g(x, t) = t f(x/t)$ - 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 $\text{range}(f_1) \subseteq S_2$. If f_1 and f_2 are convex, and f_2 is increasing, then $f_2 \circ f_1$ is convex on S_1.

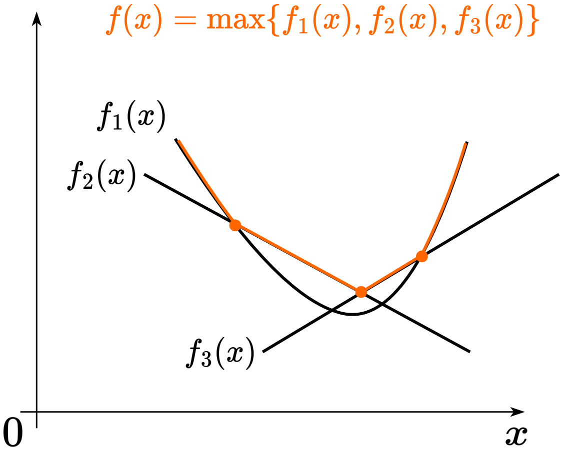


Figure 5: Pointwise maximum (supremum) of convex functions is convex

Maximum eigenvalue of a matrix is a convex function

соотн. Рэлея

$\forall \lambda$: — СЗ

X — СВ

$$Ax = \lambda x$$

$$x^T A x = \lambda x^T x \quad | \; x \neq 0$$

$$\lambda_{min} \leq \frac{x^T A x}{x^T x} \leq \lambda_{max}$$

$\forall x$

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

$$\boxed{\lambda_{max} = \max_x \frac{x^T A x}{x^T x}}$$

$$\lambda = \frac{x^T A x}{x^T x}$$

$$\lambda^x(A) = \frac{x^T A x}{x^T x}$$

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."

Polyak- Lojasiewicz condition. Linear convergence of gradient descent without convexity

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

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

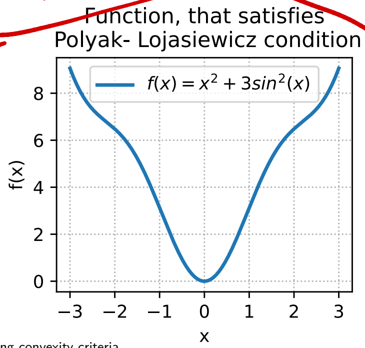
It is interesting, that Gradient Descent converges linearly under this condition (weaker, then strong convexity).

The following functions satisfy the PL-condition, but are not convex. Link to the code

$$f(x) = x^2 + 3\sin^2(x)$$

работает для PL

ЗА ЧЕМ?



если f – вып.

- лок. минимумов м.б. много
- градиентный спуск сх-ся медленно

если f – μ сильновып.

- лок. мин. единств.
- град. спуск сходится быстро

Polyak- Lojasiewicz condition. Linear convergence of gradient descent without convexity

f^* – мин. знач. функции $= \min_{x \in \text{dom} f} f$

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

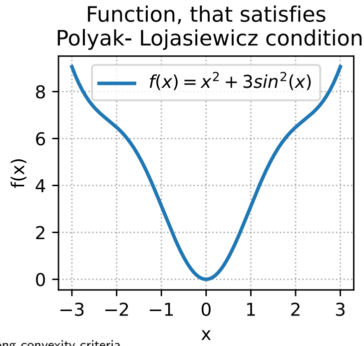
$$\boxed{\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*) \forall x}$$

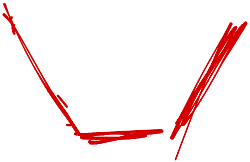
It is interesting, that Gradient Descent converges linearly under this condition (weaker, then strong convexity).

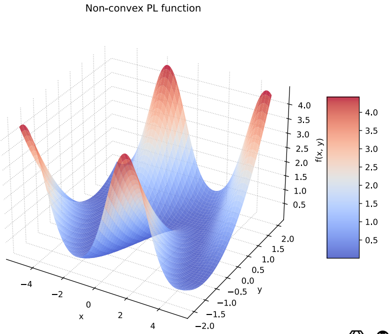
The following functions satisfy the PL-condition, but are not convex. Link to the code

$$f(x) = x^2 + 3\sin^2(x)$$

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







Convexity in ML

Linear Least Squares aka Linear Regression

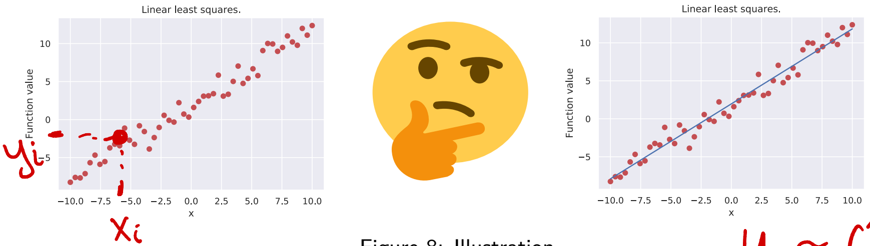


Figure 8: Illustration

$y_i \approx \theta^T x_i$

In a least-squares, or linear regression, problem, we have measurements $X \in \mathbb{R}^{m \times n}$ and $y \in \mathbb{R}^m$ and seek a vector $\theta \in \mathbb{R}^n$ such that $X\theta$ is close to y. Closeness is defined as the sum of the squared differences:

y m X m n размерность

$$\boxed{\sum_{i=1}^{m} (x_i^\top \theta - y_i)^2 =} \|X\theta - y\|_2^2 \to \min_{\theta \in \mathbb{R}^n}$$

$X\theta - y$

For example, we might have a dataset of m users, each represented by n features. Each row $x_i^\top$ of X is the features for user i, while the corresponding entry y_i of y is the measurement we want to predict from $x_i^\top$, such as ad spending. The prediction is given by $x_i^\top \theta$.

Linear Least Squares aka Linear Regression [1]

$$f(\theta) = \|X\theta - y\|_2^2$$

$$f(x) = \|Ax - b\|_2^2$$

1. Is this problem convex? Strongly convex?

$$df = d(\langle Ax-b, Ax-b\rangle) = 2\langle Ax-b, d(Ax-b)\rangle$$

$$= 2\langle Ax-b, Ad(x)\rangle =$$

$$= 2\langle A^T(Ax-b), dx\rangle$$

Linear Least Squares aka Linear Regression [1]

$$\Rightarrow \nabla f = 2A^T(Ax - b) \quad \Rightarrow \nabla^2 f = 2\underbrace{A^T A}_{n \times n}$$

A^T: $n \times m$, A: $m \times n$

$$\nabla^2 f = 2A^T A \succeq 0 \, ?$$

1. Is this problem convex? Strongly convex?
2. What do you think about convergence of Gradient Descent for this problem?

$$y^T \cdot 2A^T A \cdot y = 2 \cdot p^T \cdot p \geq 0$$

$p = Ay$

$p^T = y^T A^T$

$m \, ? \, n$

$m < n$	$m = n$	$m > n$
$\det \nabla^2 f = 0$		
$\mu = 0$	$\mu > 0$	$\mu > 0$

[1] Take a look at the example of real-world data linear least squares problem

l_2-regularized Linear Least Squares

μ^0

In the underdetermined case, it is often desirable to restore strong convexity of the objective function by adding an l_2-penality, also known as Tikhonov regularization, l_2-regularization, or weight decay.

$$
\overbrace{\|X\theta - y\|_2^2}^{f(\theta)} + \underbrace{\frac{\mu}{2}\|\theta\|_2^2}_{\mu I} \to \min_{\theta \in \mathbb{R}^n}
$$

Note: With this modification the objective is μ-strongly convex again.

Take a look at the code

$$
\nabla^2 g(\theta) = \nabla^2 f(\theta) + \mu I \succeq \mu I
$$

Most important difference between convexity and strong convexity

$$f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \frac{\mu}{2}\|x\|_2^2 \to \min_{x \in \mathbb{R}^n}, \quad A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

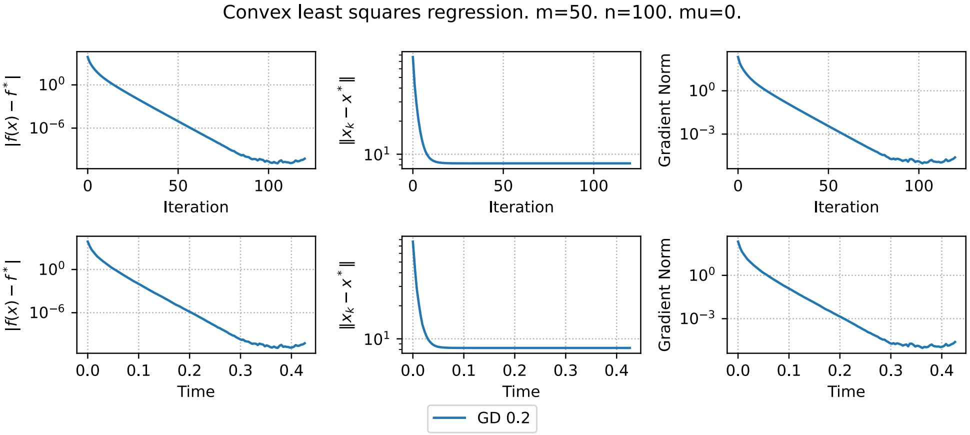


Figure 9: Convex problem does not have convergence in domain

Most important difference between convexity and strong convexity

$$f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \frac{\mu}{2}\|x\|_2^2 \to \min_{x \in \mathbb{R}^n}, \quad A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

TRAIN

TEST

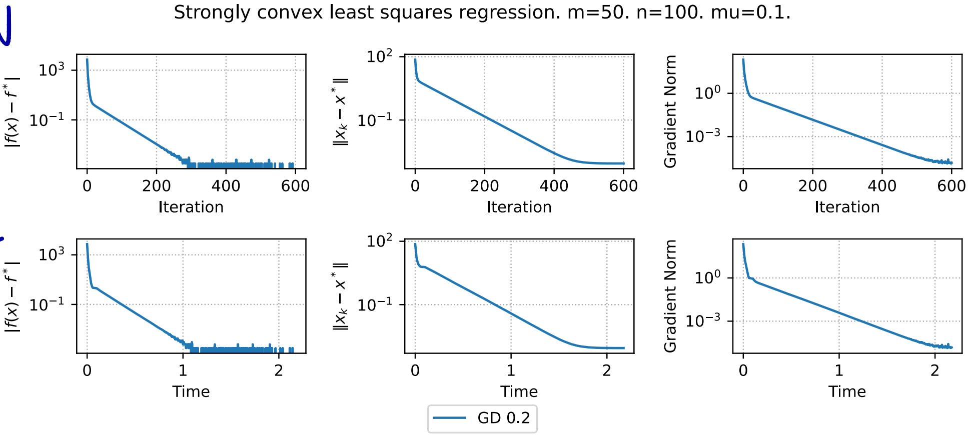


Figure 10: But if you add even small amount of regularization, you will ensure convergence in domain

Most important difference between convexity and strong convexity

$$
f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \frac{\mu}{2}\|x\|_2^2 \to \min_{x \in \mathbb{R}^n}, \quad A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m
$$

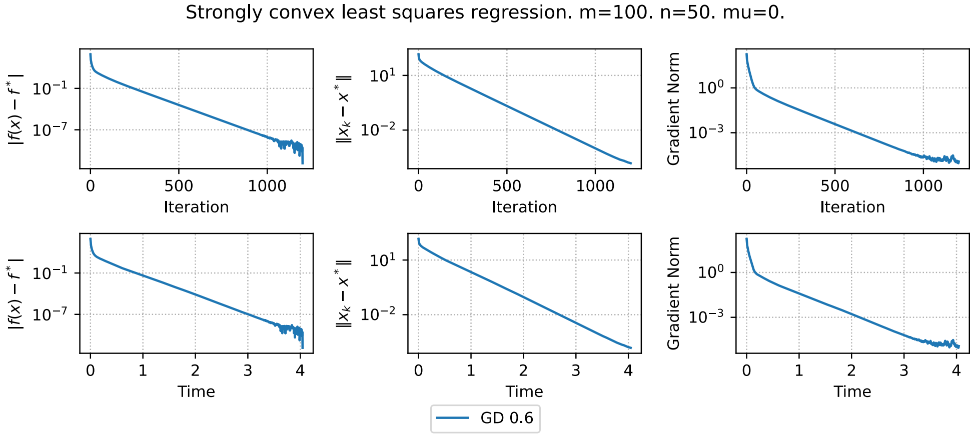


Figure 11: Another way to ensure convergence in the previous problem is to switch the dimension values

You have to have strong convexity (or PL) to ensure convergence with a high precision

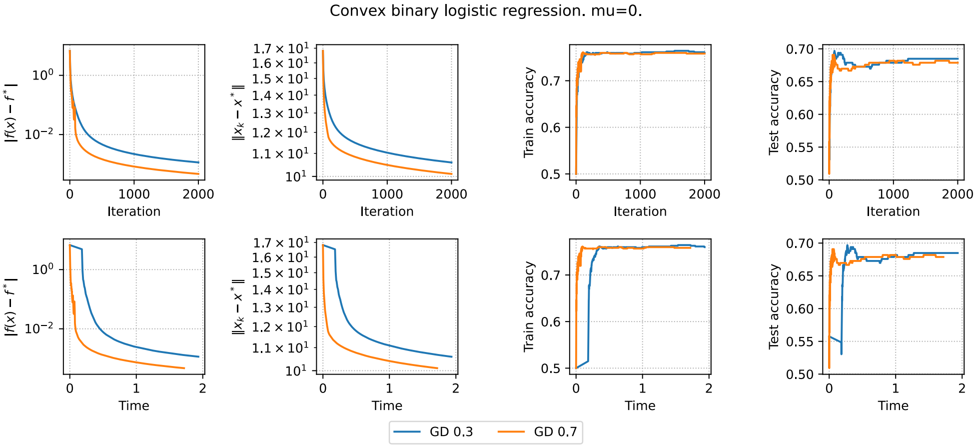


Figure 12: Only small precision is achievable with sublinear convergence

You have to have strong convexity (or PL) to ensure convergence with a high precision

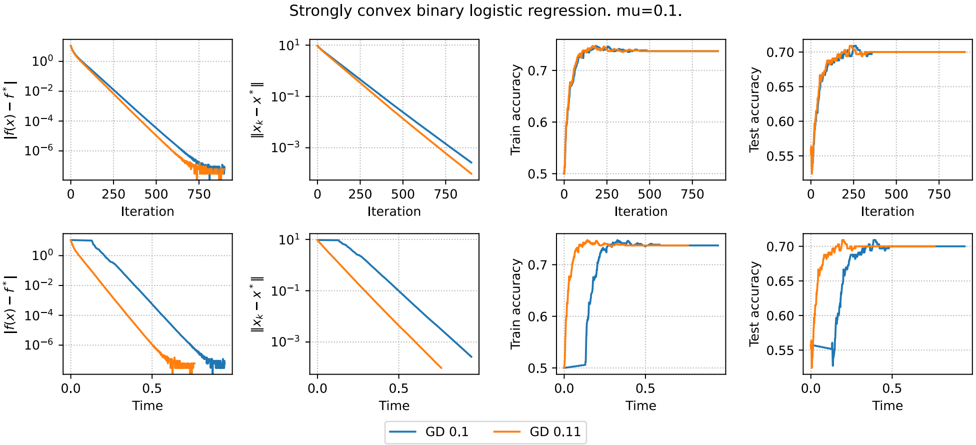


Figure 13: Strong convexity ensures linear convergence

Convex optimization problem

COP

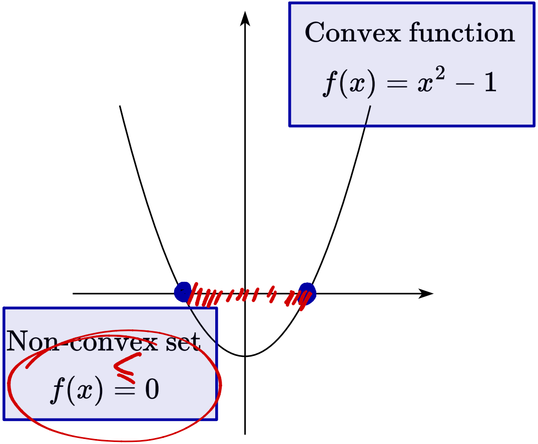


Figure 14: The idea behind the definition of a convex optimization problem

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

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

$- \; f_j(x) = 0 = a_j^T x - b_j$

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.

Conjugate sets

Conjugate set

Let $S \subseteq \mathbb{R}^n$ be an arbitrary non-empty set. Then its conjugate set is defined as:

$$S^* = \{y \in \mathbb{R}^n \mid \langle y, x\rangle \geq -1 \;\; \forall x \in S\}$$

A set S^{**} is called double conjugate to a set S if:

$$S^{**} = \{y \in \mathbb{R}^n \mid \langle y, x\rangle \geq -1 \;\; \forall x \in S^*\}$$

- The sets S_1 and S_2 are called **inter-conjugate** if $S_1^* = S_2, S_2^* = S_1$.

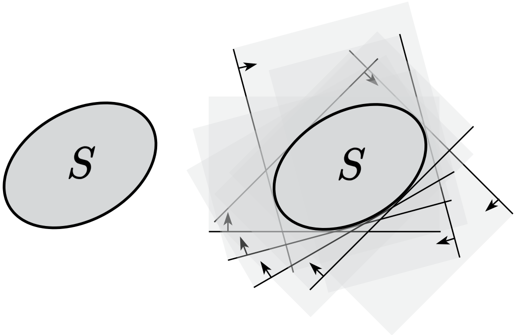


Figure 15: Convex sets may be described in a dual way - through the elements of the set and through the set of hyperplanes supporting it

Conjugate set

Let $S \subseteq \mathbb{R}^n$ be an arbitrary non-empty set. Then its conjugate set is defined as:

$$S^* = \{y \in \mathbb{R}^n \mid \langle y, x\rangle \geq -1 \;\; \forall x \in S\}$$

A set S^{**} is called double conjugate to a set S if:

$$S^{**} = \{y \in \mathbb{R}^n \mid \langle y, x\rangle \geq -1 \;\; \forall x \in S^*\}$$

- The sets S_1 and S_2 are called **inter-conjugate** if $S_1^* = S_2, S_2^* = S_1$.
- A set S is called **self-conjugate** if $S^* = S$.

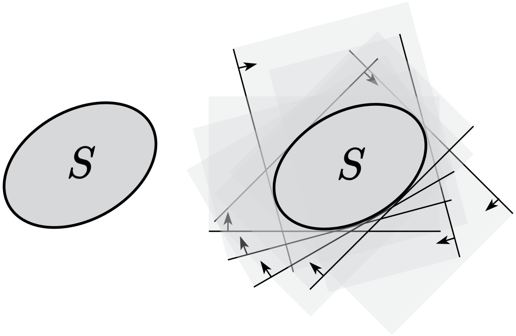


Figure 15: Convex sets may be described in a dual way - through the elements of the set and through the set of hyperplanes supporting it

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.
- For an arbitrary set $S \subseteq \mathbb{R}^n$:

$$S^{**} = \overline{\mathbf{conv}(S \cup \{0\})}$$

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.
- For an arbitrary set $S \subseteq \mathbb{R}^n$:

$$S^{**} = \overline{\mathbf{conv}(S \cup \{0\})}$$

- If $S_1 \subseteq S_2$, then $S_2^* \subseteq S_1^*$.

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.
- For an arbitrary set $S \subseteq \mathbb{R}^n$:

$$S^{**} = \overline{\mathbf{conv}(S \cup \{0\})}$$

- If $S_1 \subseteq S_2$, then $S_2^* \subseteq S_1^*$.
- $\left(\bigcup\limits_{i=1}^m S_i \right)^* = \bigcap\limits_{i=1}^m S_i^*$.

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.
- For an arbitrary set $S \subseteq \mathbb{R}^n$:

$$S^{**} = \overline{\mathbf{conv}(S \cup \{0\})}$$

- If $S_1 \subseteq S_2$, then $S_2^* \subseteq S_1^*$.
- $\left(\bigcup\limits_{i=1}^{m} S_i \right)^* = \bigcap\limits_{i=1}^{m} S_i^*$.
- If S is closed, convex, and includes 0, then $S^{**} = S$.

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.
- For an arbitrary set $S \subseteq \mathbb{R}^n$:

$$S^{**} = \overline{\mathbf{conv}(S \cup \{0\})}$$

- If $S_1 \subseteq S_2$, then $S_2^* \subseteq S_1^*$.
- $\left(\bigcup\limits_{i=1}^{m} S_i\right)^* = \bigcap\limits_{i=1}^{m} S_i^*$.
- If S is closed, convex, and includes 0, then $S^{**} = S$.
- $S^* = \left(\overline{S}\right)^*$.

Example 1

Example

Prove that $S^* = \left(\overline{S}\right)^*$.

Example 1

Example

Prove that $S^* = \left(\overline{S}\right)^*$.

- $S \subset \overline{S} \to \left(\overline{S}\right)^* \subset S^*$.

Example 1

Example

Prove that $S^* = \left(\overline{S}\right)^*$.

- $S \subset \overline{S} \to \left(\overline{S}\right)^* \subset S^*$.
- Let $p \in S^*$ and $x_0 \in \overline{S}, x_0 = \lim\limits_{k \to \infty} x_k$. Then by virtue of the continuity of the function $f(x) = p^T x$, we have: $p^T x_k \geq -1 \to p^T x_0 \geq -1$. So $p \in \left(\overline{S}\right)^*$, hence $S^* \subset \left(\overline{S}\right)^*$.

Example 2

Example

Prove that $(\mathbf{conv}(S))^* = S^*$.

Example 2

Example

Prove that $(\mathbf{conv}(S))^* = S^*$.

- $S \subset \mathbf{conv}(S) \to (\mathbf{conv}(S))^* \subset S^*$.

Example 2

Example

Prove that $(\mathbf{conv}(S))^* = S^*$.

- $S \subset \mathbf{conv}(S) \to (\mathbf{conv}(S))^* \subset S^*$.
- Let $p \in S^*$, $x_0 \in \mathbf{conv}(S)$, i.e., $x_0 = \sum\limits_{i=1}^k \theta_i x_i \mid x_i \in S, \sum\limits_{i=1}^k \theta_i = 1, \theta_i \geq 0.$

  So $p^T x_0 = \sum\limits_{i=1}^k \theta_i p^T x_i \geq \sum\limits_{i=1}^k \theta_i(-1) = 1 \cdot (-1) = -1.$ So $p \in (\mathbf{conv}(S))^*$, hence $S^* \subset (\mathbf{conv}(S))^*$.

Example 3

Example

Prove that if $B(0, r)$ is a ball of radius r by some norm centered at zero, then $(B(0, r))^* = B(0, 1/r)$.

Example 3

Example

Prove that if $B(0, r)$ is a ball of radius r by some norm centered at zero, then $(B(0, r))^* = B(0, 1/r)$.

- Let $B(0, r) = X, B(0, 1/r) = Y$. Take the normal vector $p \in X^*$, then for any $x \in X : p^T x \geq -1$.

Example 3

Example

Prove that if $B(0,r)$ is a ball of radius r by some norm centered at zero, then $(B(0,r))^* = B(0,1/r)$.

- Let $B(0,r) = X, B(0,1/r) = Y$. Take the normal vector $p \in X^*$, then for any $x \in X : p^T x \geq -1$.
- From all points of the ball X, take such a point $x \in X$ that its scalar product of p: $p^T x$ is minimal, then this is the point $x = -\frac{p}{\|p\|} r$.

$$p^T x = p^T \left(-\frac{p}{\|p\|} r \right) = -\|p\| r \geq -1$$

$$\|p\| \leq \frac{1}{r} \in Y$$

So $X^* \subset Y$.

Example 3

Example

Prove that if $B(0, r)$ is a ball of radius r by some norm centered at zero, then $(B(0, r))^* = B(0, 1/r)$.

- Let $B(0, r) = X, B(0, 1/r) = Y$. Take the normal vector $p \in X^*$, then for any $x \in X : p^T x \geq -1$.
- From all points of the ball X, take such a point $x \in X$ that its scalar product of p: $p^T x$ is minimal, then this is the point $x = -\frac{p}{\|p\|} r$.

$$p^T x = p^T \left(-\frac{p}{\|p\|} r \right) = -\|p\| r \geq -1$$

$$\|p\| \leq \frac{1}{r} \in Y$$

So $X^* \subset Y$.

- Now let $p \in Y$. We need to show that $p \in X^*$, i.e., $\langle p, x \rangle \geq -1$. It's enough to apply the Cauchy-Bunyakovsky inequality:

$$\|\langle p, x \rangle\| \leq \|p\| \|x\| \leq \frac{1}{r} \cdot r = 1$$

The latter comes from the fact that $p \in B(0, 1/r)$ and $x \in B(0, r)$.
So $Y \subset X^*$.

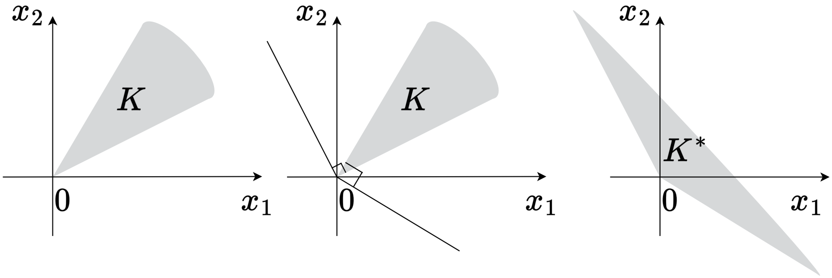
Dual cone

A conjugate cone to a cone K is a set K^* such that:

$$K^* = \{y \mid \langle x, y\rangle \geq 0 \quad \forall x \in K\}$$

To show that this definition follows directly from the definitions above, recall what a conjugate set is and what a cone $\forall \lambda > 0$ is.

$$\{y \in \mathbb{R}^n \mid \langle y, x\rangle \geq -1 \;\; \forall x \in S\} \to \{\lambda y \in \mathbb{R}^n \mid \langle y, x\rangle \geq -\frac{1}{\lambda} \;\; \forall x \in S\}$$



Dual cones properties

- Let K be a closed convex cone. Then $K^{**} = K$.

Dual cones properties

- Let K be a closed convex cone. Then $K^{**} = K$.
- For an arbitrary set $S \subseteq \mathbb{R}^n$ and a cone $K \subseteq \mathbb{R}^n$:

$$(S + K)^* = S^* \cap K^*$$

Dual cones properties

- Let K be a closed convex cone. Then $K^{**} = K$.
- For an arbitrary set $S \subseteq \mathbb{R}^n$ and a cone $K \subseteq \mathbb{R}^n$:

$$(S + K)^* = S^* \cap K^*$$

- Let $K_1, \ldots, K_m$ be cones in $\mathbb{R}^n$, then:

$$\left(\sum_{i=1}^m K_i \right)^* = \bigcap_{i=1}^m K_i^*$$

Dual cones properties

- Let K be a closed convex cone. Then $K^{**} = K$.
- For an arbitrary set $S \subseteq \mathbb{R}^n$ and a cone $K \subseteq \mathbb{R}^n$:

$$(S + K)^* = S^* \cap K^*$$

- Let $K_1, \ldots, K_m$ be cones in $\mathbb{R}^n$, then:

$$\left(\sum_{i=1}^m K_i\right)^* = \bigcap_{i=1}^m K_i^*$$

- Let $K_1, \ldots, K_m$ be cones in $\mathbb{R}^n$. Let also their intersection have an interior point, then:

$$\left(\bigcap_{i=1}^m K_i\right)^* = \sum_{i=1}^m K_i^*$$

Example

Example

Find the conjugate cone for a monotone nonnegative cone:

$$K = \{x \in \mathbb{R}^n \mid x_1 \geq x_2 \geq \ldots \geq x_n \geq 0\}$$

Example

Example

Find the conjugate cone for a monotone nonnegative cone:

$$K = \{x \in \mathbb{R}^n \mid x_1 \geq x_2 \geq \ldots \geq x_n \geq 0\}$$

Note that:

$$\sum_{i=1}^{n} x_i y_i = y_1(x_1 - x_2) + (y_1 + y_2)(x_2 - x_3) + \ldots + (y_1 + y_2 + \ldots + y_{n-1})(x_{n-1} - x_n) + (y_1 + \ldots + y_n)x_n$$

Since in the presented sum in each summand, the second multiplier in each summand is non-negative, then:

$$y_1 \geq 0, \;\; y_1 + y_2 \geq 0, \;\; \ldots, \;\; y_1 + \ldots + y_n \geq 0$$

So $K^* = \left\{ y \mid \sum\limits_{i=1}^{k} y_i \geq 0, k = \overline{1,n} \right\}$.

Polyhedra

The set of solutions to a system of linear inequalities and equalities is a polyhedron:

$$Ax \preceq b, \quad Cx = d$$

Here $A \in \mathbb{R}^{m\times n}, C \in \mathbb{R}^{p\times n}$, and the inequality is a piecewise inequality.

Theorem

Let $x_1, \ldots, x_m \in \mathbb{R}^n$. Conjugate to a polyhedral set:

$$S = \mathbf{conv}(x_1, \ldots, x_k) + \mathbf{cone}(x_{k+1}, \ldots, x_m)$$

is a polyhedron (polyhedron):

$$S^* = \left\{ p \in \mathbb{R}^n \mid \langle p, x_i \rangle \geq -1, i = \overline{1,k}; \langle p, x_i \rangle \geq 0, i = \overline{k+1,m} \right\}$$

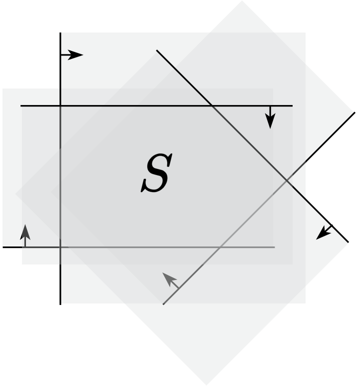


Figure 17: Polyhedra

Proof

- Let $S = X, S^* = Y$. Take some $p \in X^*$, then $\langle p, x_i \rangle \geq -1, i = \overline{1, k}$. At the same time, for any $\theta > 0, i = \overline{k+1, m}$:

$$\langle p, x_i \rangle \geq -1 \rightarrow \langle p, \theta x_i \rangle \geq -1$$

$$\langle p, x_i \rangle \geq -\frac{1}{\theta} \rightarrow \langle p, x_i \rangle \geq 0.$$

So $p \in Y \rightarrow X^* \subset Y$.

Proof

- Let $S = X, S^* = Y$. Take some $p \in X^*$, then $\langle p, x_i \rangle \geq -1, i = \overline{1, k}$. At the same time, for any $\theta > 0, i = \overline{k+1, m}$:

$$
\langle p, x_i \rangle \geq -1 \rightarrow \langle p, \theta x_i \rangle \geq -1
$$

$$
\langle p, x_i \rangle \geq -\frac{1}{\theta} \rightarrow \langle p, x_i \rangle \geq 0.
$$

  So $p \in Y \rightarrow X^* \subset Y$.

- Suppose, on the other hand, that $p \in Y$. For any point $x \in X$:

$$
x = \sum_{i=1}^m \theta_i x_i \qquad \sum_{i=1}^k \theta_i = 1, \theta_i \geq 0
$$

  So:

$$
\langle p, x \rangle = \sum_{i=1}^m \theta_i \langle p, x_i \rangle = \sum_{i=1}^k \theta_i \langle p, x_i \rangle + \sum_{i=k+1}^m \theta_i \langle p, x_i \rangle \geq \sum_{i=1}^k \theta_i (-1) + \sum_{i=1}^k \theta_i \cdot 0 = -1.
$$

Example