Example 3

Let $x \in \mathbb{R}$ is a random variable with a given probability distribution of $\mathbb{P}(x = a_i) = p_i$, where $i = 1, \dots, n$, and $a_1 < \dots < a_n$. It is said that the probability vector of outcomes of $p \in \mathbb{R}^n$ belongs to the probabilistic simplex, i.e.
$P = \{p \mid \mathbf{1}^T p = 1, p \succeq 0\} = \{p \mid p_1 + \dots + p_n = 1, p_i \geq 0\}$. Determine if the following sets of p are convex: 1. $\alpha < \mathbb{E}f(x) < \beta$, where $\mathbb{E}f(x)$ stands for expected value of $f(x) : \mathbb{R} \to \mathbb{R}$, i.e. $\mathbb{E}f(x) = \sum\limits_{i=1}^n p_i f(a_i)$ 1. $\mathbb{E}x^2 \leq \alpha$ 1. $\mathbb{V}x \leq \alpha$

Convex function

Convex function

~~$f: \mathbb{R}^n \to \mathbb{R}$~~ $f: S \subseteq \mathbb{R}^n \to \mathbb{R}$

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

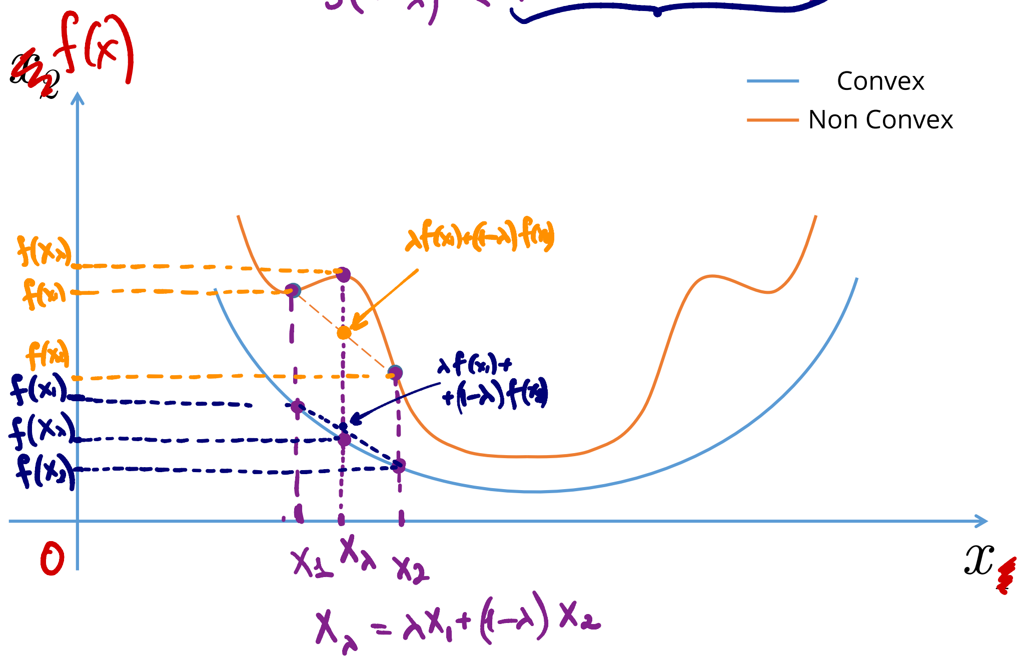
Нер-во Йенсена

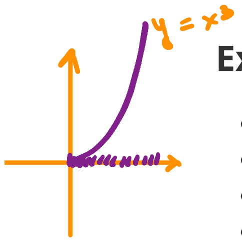
$$\boxed{f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda) f(x_2)}$$

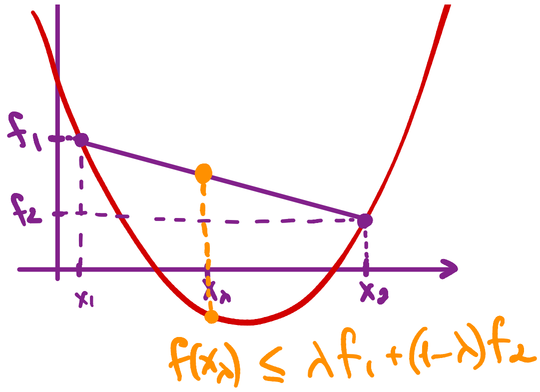
$\lambda x_1 + (1-\lambda) x_2$

for any $x_1, x_2 \in S$ and $0 \leq \lambda \leq 1$. $x_1 \in \mathbb{R}^n$ $x_2 \in \mathbb{R}^n$
If above inequality holds as strict inequality $x_1 \neq x_2$ and $0 < \lambda < 1$, then function is called strictly convex S

$$f(x_\lambda) \leq \lambda \cdot f(x_1) + (1-\lambda) f(x_2)$$







Examples

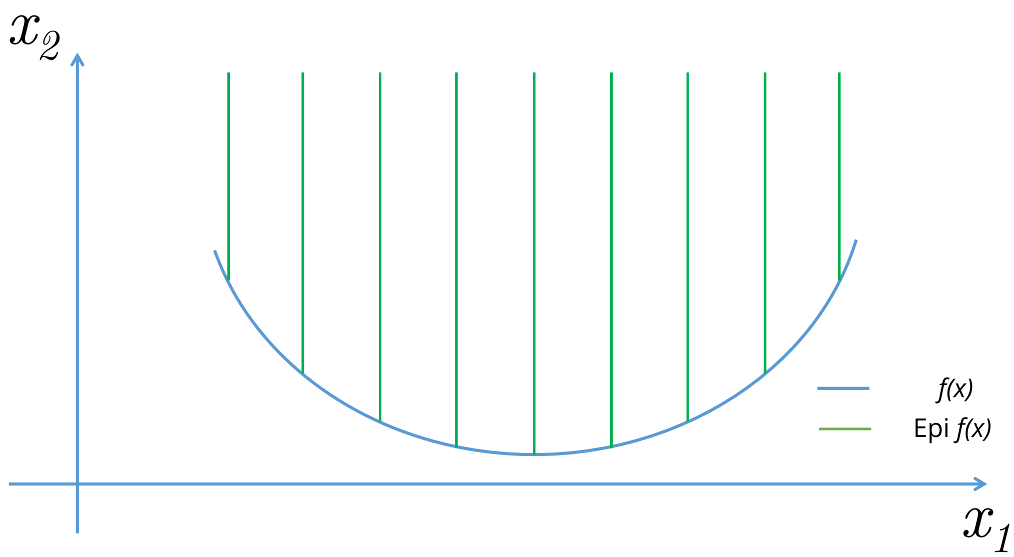
- $f(x) = x^p, p > 1, \quad S = \mathbb{R}_+$
- $f(x) = \|x\|^p, \quad p > 1, S = \mathbb{R}$
- $f(x) = e^{cx}, \quad c \in \mathbb{R}, S = \mathbb{R}$
- $f(x) = -\ln x, \quad S = \mathbb{R}_{++}$
- $f(x) = x \ln x, \quad S = \mathbb{R}_{++}$
- The sum of the largest k coordinates $f(x) = x_{(1)} + \ldots + x_{(k)}, \quad S = \mathbb{R}^n$
- $f(X) = \lambda_{max}(X), \quad X = X^T$
- $f(X) = -\log \det X, \quad S = S^n_{++}$

Epigraph

For the function $f(x)$, defined on $S \subseteq \mathbb{R}^n$, the following set:

$$\text{epi } f = \{[x, \mu] \in S \times \mathbb{R} : f(x) \leq \mu\}$$

is called **epigraph** of the function $f(x)$

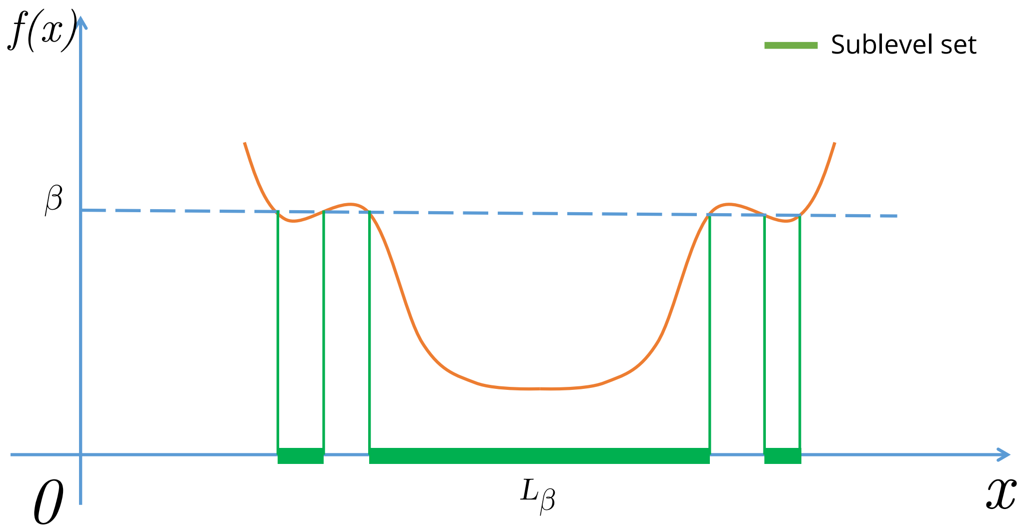


Sublevel set

For the function $f(x)$, defined on $S \subseteq \mathbb{R}^n$, the following set:

$$\mathcal{L}_\beta = \{x \in S : f(x) \leq \beta\}$$

is called **sublevel set** or Lebesgue set of the function $f(x)$



Criteria of convexity

First order differential criterion of convexity

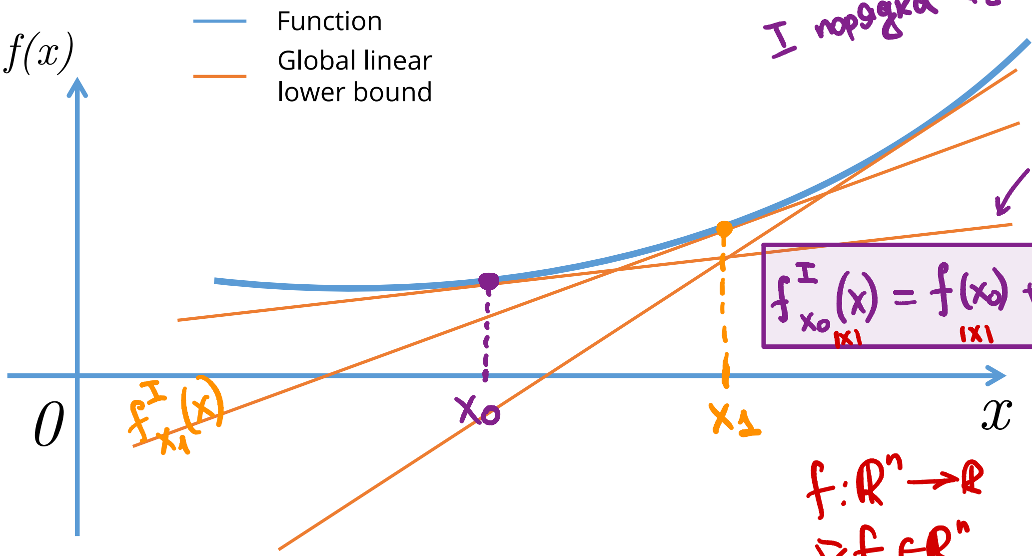
The differentiable function $f(x)$ defined on the convex set $S \subseteq \mathbb{R}^n$ is convex if and only if $\forall x, y \in S$:

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

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

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

Наилучш. аппроксимация I порядка функции $f(x)$ вблизи т. x_0



$$\boxed{f^I_{x_0}(x) = \underset{1\times1}{f(x_0)} + \langle \underset{n\times1}{\nabla f(x_0)}, \underset{n\times1}{x - x_0} \rangle}$$

$f : \mathbb{R}^n \to \mathbb{R}$

$\nabla f \in \mathbb{R}^n$

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

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

$\forall x_0 \in \mathbb{R}^n$

$x_0^T \nabla^2 f(x) \cdot x_0 \geq 0$

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 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.

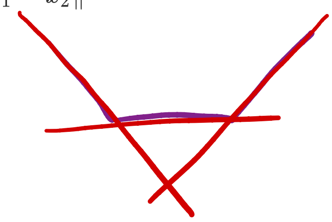
Reduction to a line

$f : S \to \mathbb{R}$ is convex if and only if S is 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 to check convexity of the scalar function in order to establish covexity of the vector function.

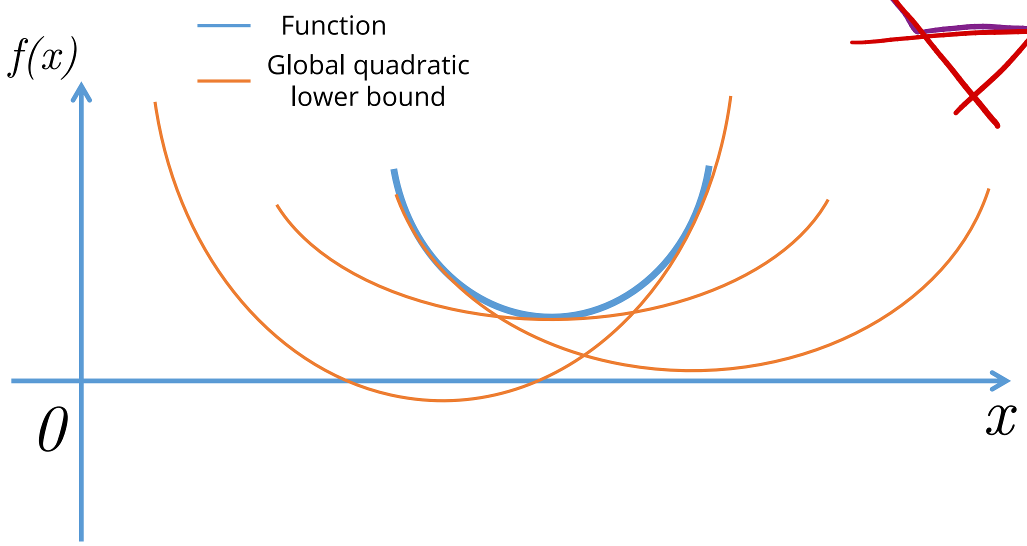
Strong convexity

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

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



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



Criteria of strong convexity

First order differential criterion of strong convexity

Differentiable $f(x)$ defined on the convex set $S \subseteq \mathbb{R}^n$ 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:

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

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:

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

Facts

- $f(x)$ is called (strictly) concave, if the function $-f(x)$ - (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 $x^* = \arg\min\limits_{x \in S} f(x)$ will be the global one. Strong convexity guarantees the uniqueness of the solution.

Operations that preserve convexity

- Non-negative sum of the convex functions: $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0)$
- Composition with affine function $f(Ax+b)$ is convex, if $f(x)$ is convex
- Pointwise maximum (supremum): 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 $\mathrm{range}(f_1) \subseteq S_2$. 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-convex: $\log f$ is convex; Log convexity implies convexity.
- Log-concavity: $\log f$ concave; **not** closed under addition!
- Exponentially convex: $[f(x_i + x_j)] \succeq 0$, for $x_1, \ldots, x_n$
- Operator convex: $f(\lambda X + (1-\lambda)Y) \preceq \lambda f(X) + (1-\lambda) f(Y)$
- Quasiconvex: $f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$
- Pseudoconvex: $\langle \nabla f(y), x - y\rangle \geq 0 \longrightarrow f(x) \geq f(y)$

- Discrete convexity: $f : \mathbb{Z}^n \to \mathbb{Z}$; “convexity + matroid theory.”

References

- Steven Boyd lectures
- Suvrit Sra lectures
- Martin Jaggi lectures

$f(x) = kx+b$
$f'(x) = k$
$f''(x) = 0$

Example 4

вып. вогнутая

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

Решение:

1) Рассмотрим $\nabla^2 f$. $df = \langle c, dx\rangle \Rightarrow \nabla f = c$

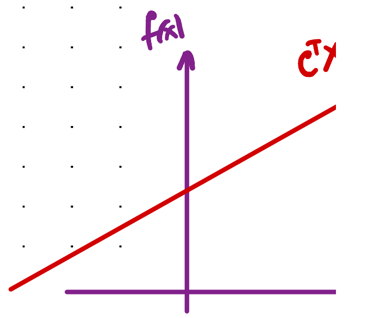
$d^2 f = \langle dc, dx_1\rangle = 0 \Rightarrow \boxed{\nabla^2 f = \mathbb{0}^{n\times n}}$

$\nabla^2 f \succeq 0 \quad \forall x \in \mathbb{R}^n \to x^T \cdot \nabla^2 f \cdot x = 0$

$\Rightarrow$ f – выпукла.

2) Вогнутость $f(x)$ = выпуклость $-f(x)$

$\nabla^2(-f(x)) = \mathbb{0}^{n\times n} \succeq 0$



Example 5

$A = A^T$

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

Решение: 1) $df = d(\langle x, Ax\rangle) = \langle (A+A^T)x, dx\rangle = \langle 2Ax, dx\rangle$

$\nabla f = 2Ax$

$d^2 f = \langle d(2Ax), dx_1\rangle = \langle 2A\,dx, dx_1\rangle = \langle 2A\,dx_1, dx\rangle$

$\Rightarrow \boxed{\nabla^2 f = 2A} \succeq 0$

если $A \succeq 0 \to \forall x \in \mathbb{R}^n \quad x^T A x \geq 0 \ \big|\cdot 2$

$x^T \cdot (2A) \cdot x \geq 0 \cdot 2$

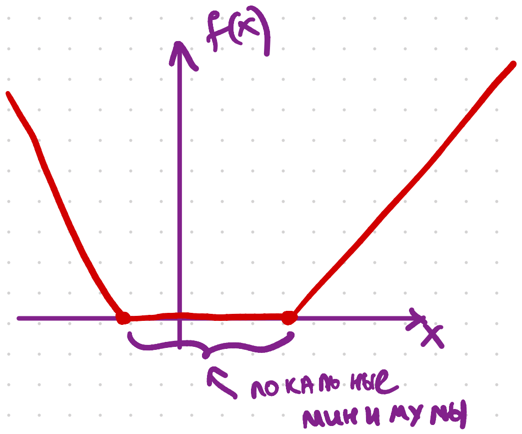
Example 6

Show, that $f(x)$ is convex, using first and second order criteria, if $f(x) = \sum\limits_{i=1}^{n} x_i^4$.

Example 7

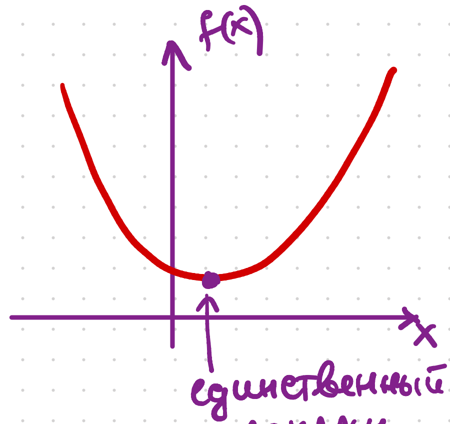
Find the set of $x \in \mathbb{R}^n$, where the function $f(x) = \dfrac{-1}{2(1 + x^\top x)}$ is convex, strictly convex, strongly convex?

f(x) выпукла



любой $\boxed{\text{локальный}}$ минимум является $\boxed{\text{глобальным}}$

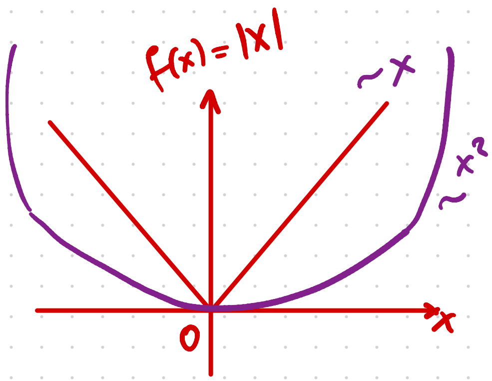
f(x) сильно выпукла



локальный $\boxed{\text{минимум}}$ $\boxed{\text{единственный}}$

+

такая функция растёт $\boxed{\text{быстрее}}$ $\boxed{\text{некоторой параболы}}$



выпуклая, но не сильно выпуклая

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 \qquad f: \mathbb{R}^n \to \mathbb{R}$$

$A \in \mathbb{R}^{m\times n}$

$b \in \mathbb{R}^m$

$$df = \frac{1}{2} d(\langle Ax-b, Ax-b\rangle) =$$

$$= \frac{1}{2} \cdot \langle 2(Ax-b), A dx\rangle$$

$$\Rightarrow df = \langle A^T(Ax-b), dx\rangle \to \nabla f = A^T(Ax-b)$$

$$d^2 f = \langle d(A^T(Ax-b)), dx_1\rangle = \langle A^T A \cdot dx, dx_1\rangle =$$

$$= \langle A^T A\, dx_1, dx\rangle$$

$$\Rightarrow \boxed{\nabla^2 f = \underbrace{A^T}_{n\times m} \underbrace{A}_{m\times n}}$$

$n\times n$

критерий Сильвестра ~~(странно)~~

$\forall x \in \mathbb{R}^n$:

$$(x^T A^T) A x \geq 0$$

$$(Ax)^T Ax$$

$$\langle Ax, Ax\rangle = \|Ax\|_2^2 \geq 0$$

Критерий сильной выпуклости:

$\forall x_0 \in \mathbb{R}^n / \{0\}$ $\qquad \forall x \in S \quad f: S \to \mathbb{R}$

$$\boxed{x_0^T \cdot \nabla^2 f(x) \cdot x_0 > 0}$$

$$x_0^T \nabla^2 f(x)\, x_0 \geq \mu \cdot x_0^T \cdot I\, x_0$$

$$\mu \cdot x_0^T x_0$$

$$\mu \cdot \|x_0\|_2^2$$

$$\boxed{\mu > 0}$$

чем больше μ, тем круче функция

μ - константа сильной выпуклости.

$$f(x) = x^T A x \qquad A \in S^n_+$$

$$\Rightarrow \nabla^2 f = 2A$$

если $A \in S^n_+$

$A \succeq 0$

f - выпуклая

f - не сильно выпуклая

если $A \in S^n_{++}$

$A \succ 0$

$\nabla^2 f \succ 0$

f - сильновыпуклая

$\nabla^2 f = 2A$

$\forall\, x \in \mathbb{R}^n$:

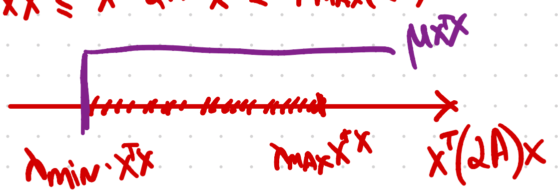
$$x^T \cdot 2A \cdot x \geq \mu \cdot x^T x$$

$$\underline{\mu = \lambda_{min}(2A)}$$

$$x^T(2A)\cdot x \geq \lambda_{min}(2A)\, x^T x$$

$$\Rightarrow \boxed{\mu = 2\cdot\lambda_{min}(A)} > 0$$

$$\lambda_{min}\, x^T x \leq x^T \cdot 2A \cdot x \leq \lambda_{max}(2A)\cdot x^T x$$



$$f(x) = \frac{1}{2} x^T A x \quad - \text{ квадр.}$$

$\lambda_{min}(A) < 0$

$f(x)$ - не выпуклая

$\lambda_{min}(A) = 0$

выпуклая

$\lambda_{min}(A) > 0$

сильно выпуклая

Optimality conditions. KKT

Background

Extreme value (Weierstrass) theorem

Let $S \subset \mathbb{R}^n$ be compact set and $f(x)$ continuous function on S. So that, the point of the global minimum of the function $f(x)$ on S exists.

Lagrange multipliers

Consider simple yet practical case of equality constraints:

$$
\begin{split}
& f(x) \to \min\limits_{x \in \mathbb{R}^n} \\
\text{s.t. } & h_i(x) = 0, i = 1, \ldots, p
\end{split}
$$

The basic idea of Lagrange method implies switch from conditional to unconditional optimization through increasing the dimensionality of the problem:

$$
L(x, \nu) = f(x) + \sum\limits_{i=1}^m \nu_i h_i(x) \to \min\limits_{x \in \mathbb{R}^n, \nu \in \mathbb{R}^p}
$$

General formulations and conditions

$$
f(x) \to \min\limits_{x \in S}
$$

We say that the problem has a solution if the budget set **is not empty**: $x^* \in S$, in which the minimum or the infimum of the given function is achieved.

Optimization on the general set S.

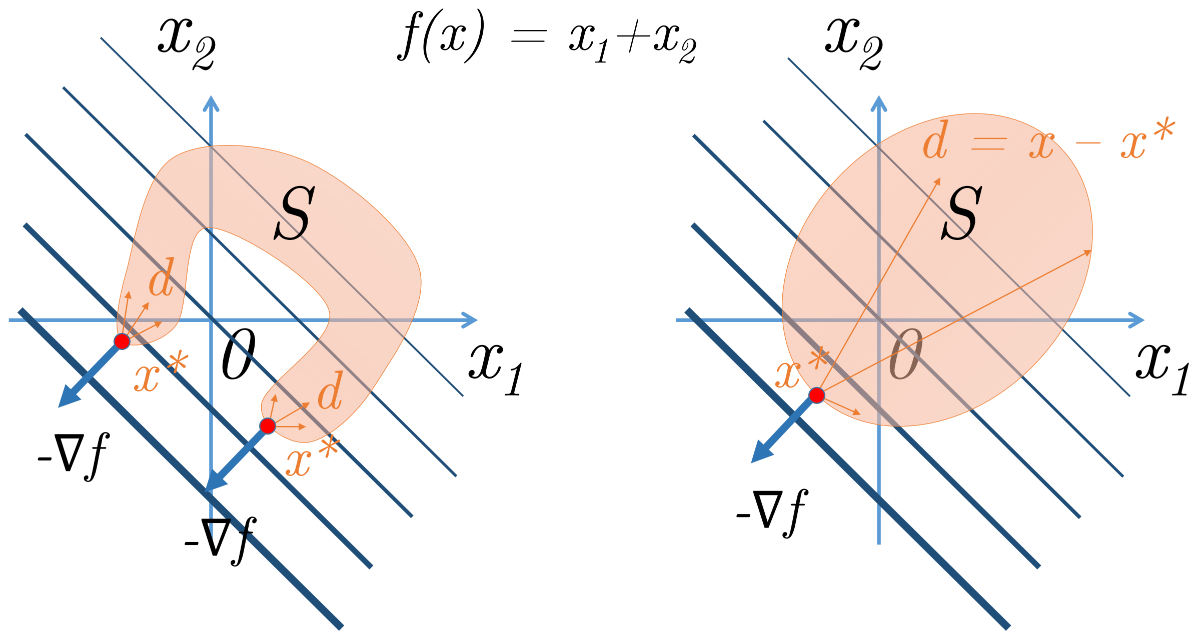
Direction $d \in \mathbb{R}^n$ is a feasible direction at $x^* \in S \subseteq \mathbb{R}^n$ if small steps along d do not take us outside of S.

Consider a set $S \subseteq \mathbb{R}^n$ and a function $f : \mathbb{R}^n \to \mathbb{R}$. Suppose that $x^* \in S$ is a point of local minimum for f over S, and further assume that f is continuously differentiable around x^*.

1. Then for every feasible direction $d \in \mathbb{R}^n$ at x^* it holds that $\nabla f^{(}x^*)^\top d \geq 0$

2. If, additionally, S is convex then

$$\nabla f(x^*)^\top(x - x^*) \geq 0, \forall x \in S.$$



Unconstrained optimization

Безусловная задача оптимизации

General case

Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function.

$$\boxed{f(x) \to \min_{x \in \mathbb{R}^n}} \tag{UP}$$

If x^* - is a local minimum of $f(x)$, then:

Необходимое условие лок. экстр.

$$\boxed{\nabla f(x^*) = 0} \tag{UP:Nec.}$$

1. $\nabla f(x^*) = 0$
2. $\nabla^2 f(x^*) \succ 0$ – ми

$\nabla^2 f(x^*) \prec 0$ – м

If $f(x)$ at some point x^* satisfies the following conditions:

достаточное условие

$$H_f(x^*) = \boxed{\nabla^2 f(x^*) \succ (\prec) 0,} \tag{UP:Suff.}$$

then (if necessary condition is also satisfied) x^* is a local minimum(maximum) of $f(x)$.

Note, that if $\nabla f(x^*) = 0, \nabla^2 f(x^*) = 0$, i.e. the hessian is positive *semidefinite*, we cannot be sure if x^* is a local minimum (see [Peano surface] $f(x, y) = (2x^2 - y)(y - x^2)$).

Convex case

It should be mentioned, that in **convex** case (i.e., $f(x)$ is convex) necessary condition becomes sufficient. Moreover, we can generalize this result on the class of non-differentiable convex functions.

Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ - convex function, then the point x^* is the solution of (UP) if and only if:

$$0_n \in \partial f(x^*)$$

One more important result for convex constrained case sounds as follows. If $f(x) : S \to \mathbb{R}$ - convex function defined on the convex set S, then:

- Any local minima is the global one.
- The set of the local minimizers S^* is convex.

пример: лин. регрессия:

$$f(x) = \frac{1}{2}\|Ax-b\|_2^2 \to \min_{x\in\mathbb{R}^n}$$

$A \in \mathbb{R}^{m\times n}$

1. $\nabla f = A^T(Ax-b) = 0$

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

$$A^TAx - A^Tb = 0$$

$$A^TAx = A^Tb \,\Big|\, (A^TA)^{-1}\cdot$$

$$(A^TA)^{-1}\cdot A^TA\,x = (A^TA)^{-1}A^Tb$$

$$\boxed{x = (A^TA)^{-1}\cdot A^Tb}$$

$(A^TA)^{-1}$ существует

$\det(A^TA) > 0$

$A^T \quad A$

$n\times m \quad m\times n$

m - число ур-ий

n - число неизвестных

$m > n$

$mn \to n^2$

$\det A \neq 0$

$$\boxed{x^* = (A^TA)^{-1}A^Tb}$$

$$\boxed{x^* = A^\dagger b}$$

\cross

\dagger

$A^\dagger = (A^TA)^{-1}A^T$

псевдообратная матрица

$m = n$

если $\operatorname{rg} A = n$

$\det A \neq 0$

$\det A^TA \neq 0$

$\det A^TA = \det A^T \cdot \det A =$

$= (\det A)^2$

$A^TAx = A^Tb \;\Big|\; \cdot (A^T)^{-1}$

$(A^T)^{-1}A^TA\,x = (A^T)^{-1}A^Tb$

$Ax = b \quad \Big|\cdot A^{-1}$

$$\boxed{x = A^{-1}b}$$

$m < n$

всех матрице mn

в A^TA $\quad n^2 \quad mn < n^2$

$\det(A^TA) = 0$