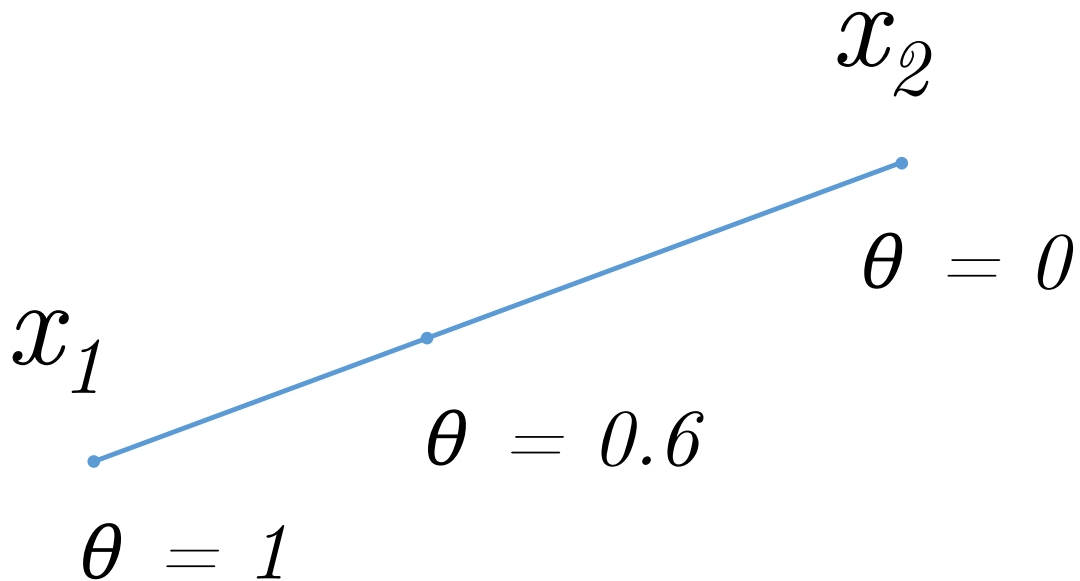


Convex set

Line segment

Suppose x_1, x_2 are two points in \mathbb{R}^m . Then the line segment between them is defined as follows:

$$x = \theta x_1 + (1 - \theta)x_2, \theta \in [0, 1]$$



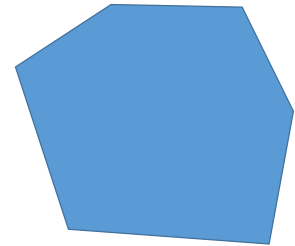
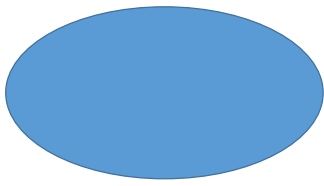
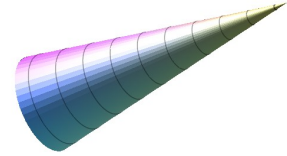
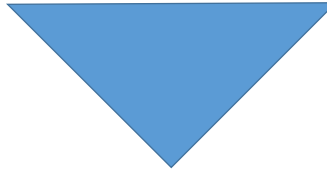
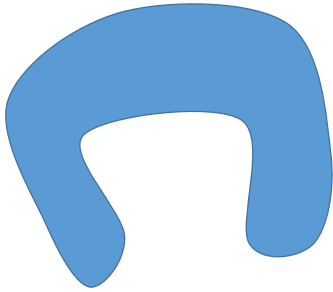
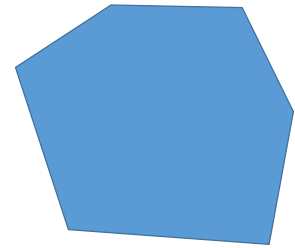
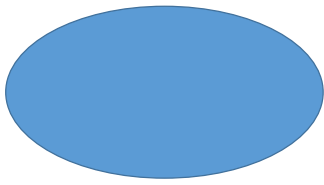
Convex set

The set S is called **convex** if for any x_1, x_2 from S the line segment between them also lies in S , i.e.

$$\forall \theta \in [0, 1], \forall x_1, x_2 \in S : \\ \theta x_1 + (1 - \theta)x_2 \in S$$

Examples:

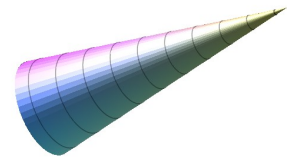
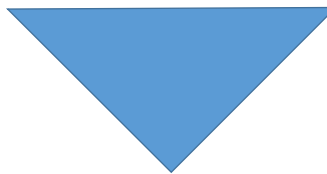
- Any affine set
- Ray
- Line segment



BRO

NOT BRO

BRO



NOT BRO

BRO

BRO

Related definitions

Convex combination

Let $x_1, x_2, \dots, x_k \in S$, then the point $\theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k$ is called the convex combination of points x_1, x_2, \dots, x_k if $\sum_{i=1}^k \theta_i = 1, \theta_i \geq 0$

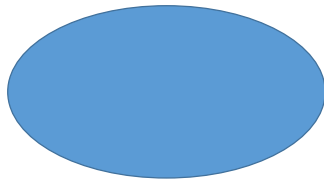
Convex hull

The set of all convex combinations of points from S is called the convex hull of the set S .

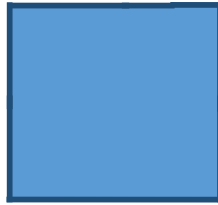
$$\mathbf{conv}(S) = \left\{ \sum_{i=1}^k \theta_i x_i \mid x_i \in S, \sum_{i=1}^k \theta_i = 1, \theta_i \geq 0 \right\}$$

- The set $\mathbf{conv}(S)$ is the smallest convex set containing S .
- The set S is convex if and only if $S = \mathbf{conv}(S)$.

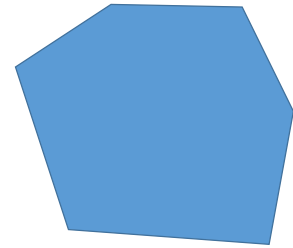
Examples:



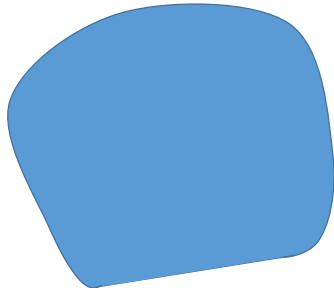
BRO



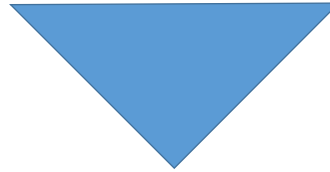
BRO



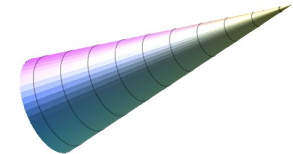
BRO



BRO



BRO



BRO

Finding convexity

In practice it is very important to understand whether a specific set is convex or not. Two approaches are used for this depending on the context.

- By definition.
- Show that S is derived from simple convex sets using operations that preserve convexity.

By definition

$$x_1, x_2 \in S, 0 \leq \theta \leq 1 \rightarrow \theta x_1 + (1 - \theta)x_2 \in S$$

Preserving convexity

The linear combination of convex sets is convex

Let there be 2 convex sets S_x, S_y , let the set

$$S = \{s \mid s = c_1x + c_2y, x \in S_x, y \in S_y, c_1, c_2 \in \mathbb{R}\}$$

Take two points from S : $s_1 = c_1x_1 + c_2y_1, s_2 = c_1x_2 + c_2y_2$ and prove that the segment between them $\theta s_1 + (1 - \theta)s_2, \theta \in [0, 1]$ also belongs to S :

$$\begin{aligned} & \theta s_1 + (1 - \theta)s_2 \\ & \theta(c_1x_1 + c_2y_1) + (1 - \theta)(c_1x_2 + c_2y_2) \\ & c_1(\theta x_1 + (1 - \theta)x_2) + c_2(\theta y_1 + (1 - \theta)y_2) \\ & c_1x + c_2y \in S \end{aligned}$$

The intersection of any (!) number of convex sets is convex

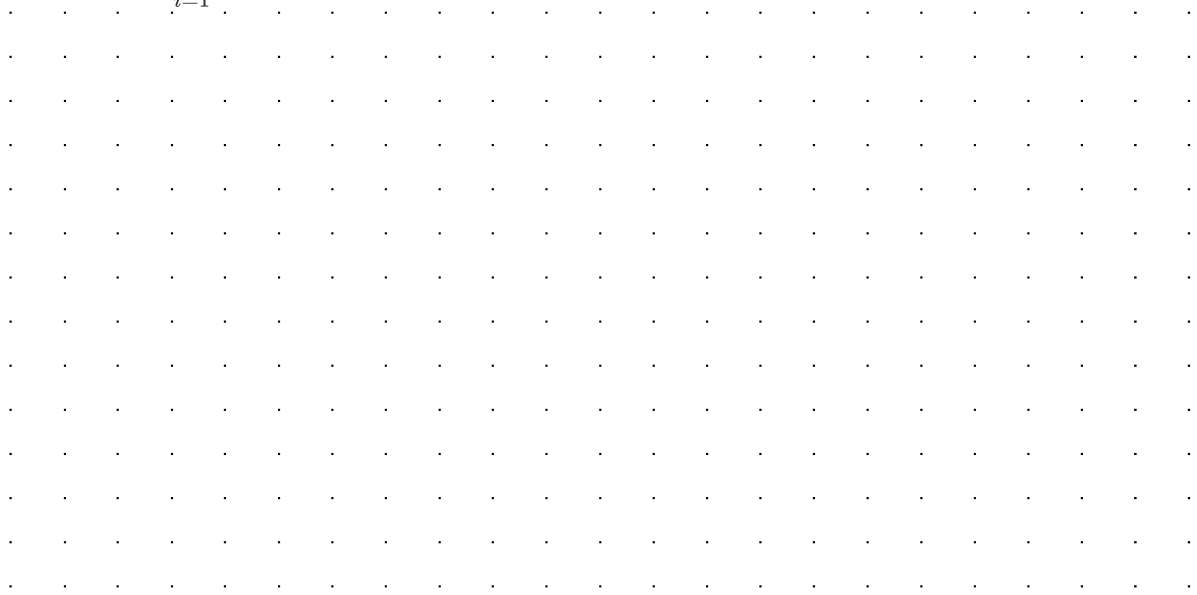
If the desired intersection is empty or contains one point, the property is proved by definition. Otherwise, take 2 points and a segment between them. These points must lie in all intersecting sets, and since they are all convex, the segment between them lies in all sets and, therefore, in their intersection.

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 $f(x) : \mathbb{R} \rightarrow \mathbb{R}$,

i.e. $\mathbb{E}f(x) = \sum_{i=1}^n p_i f(a_i)$ 1. $\mathbb{E}x^2 \leq \alpha$ 1. $\forall x \leq \alpha$



Convex function

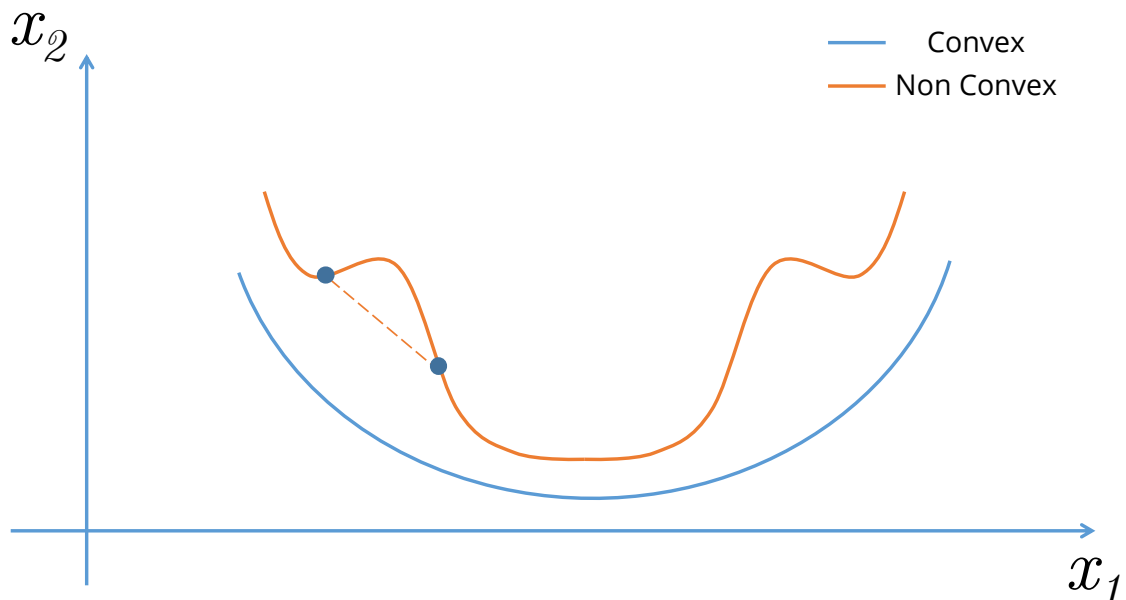
Convex function

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

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

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

If above inequality holds as strict inequality $x_1 \neq x_2$ and $0 < \lambda < 1$, then function is called strictly convex S



Examples

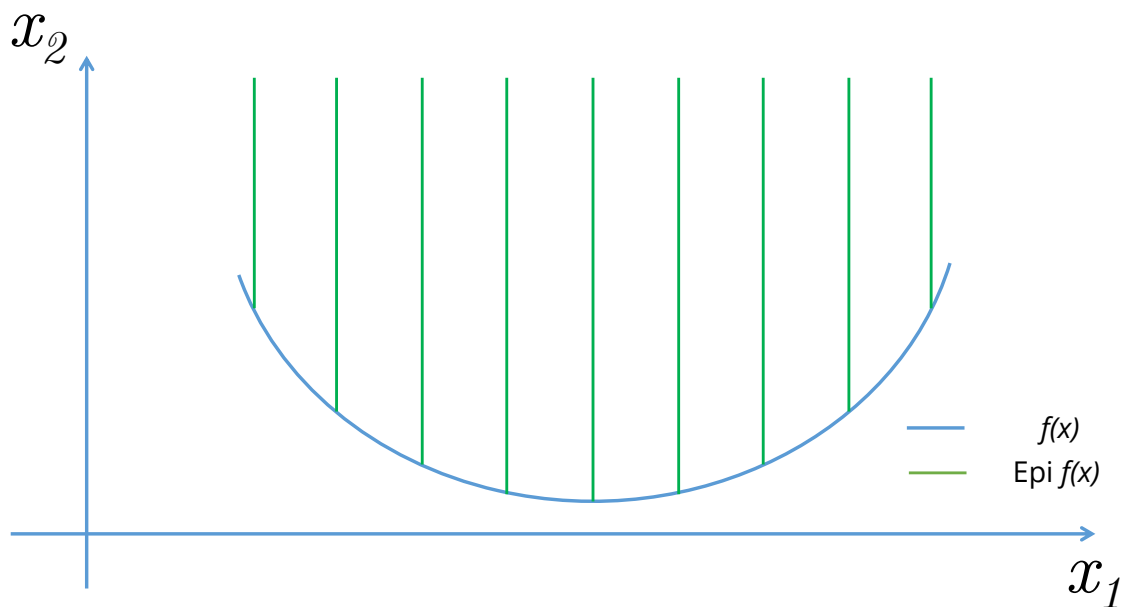
- $f(x) = x^p, p > 1, S = \mathbb{R}_+$
- $f(x) = \|x\|^p, p > 1, S = \mathbb{R}$
- $f(x) = e^{cx}, c \in \mathbb{R}, S = \mathbb{R}$
- $f(x) = -\ln x, S = \mathbb{R}_{++}$
- $f(x) = x \ln x, S = \mathbb{R}_{++}$
- The sum of the largest k coordinates $f(x) = x_{(1)} + \dots + x_{(k)}, S = \mathbb{R}^n$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

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 $f(x)$

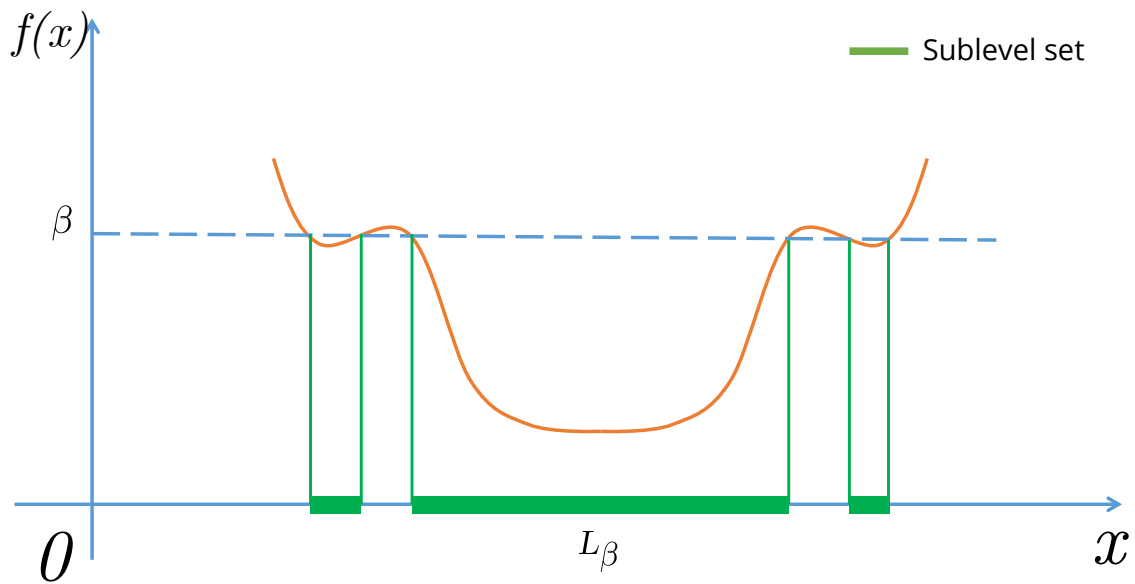


Sublevel set

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

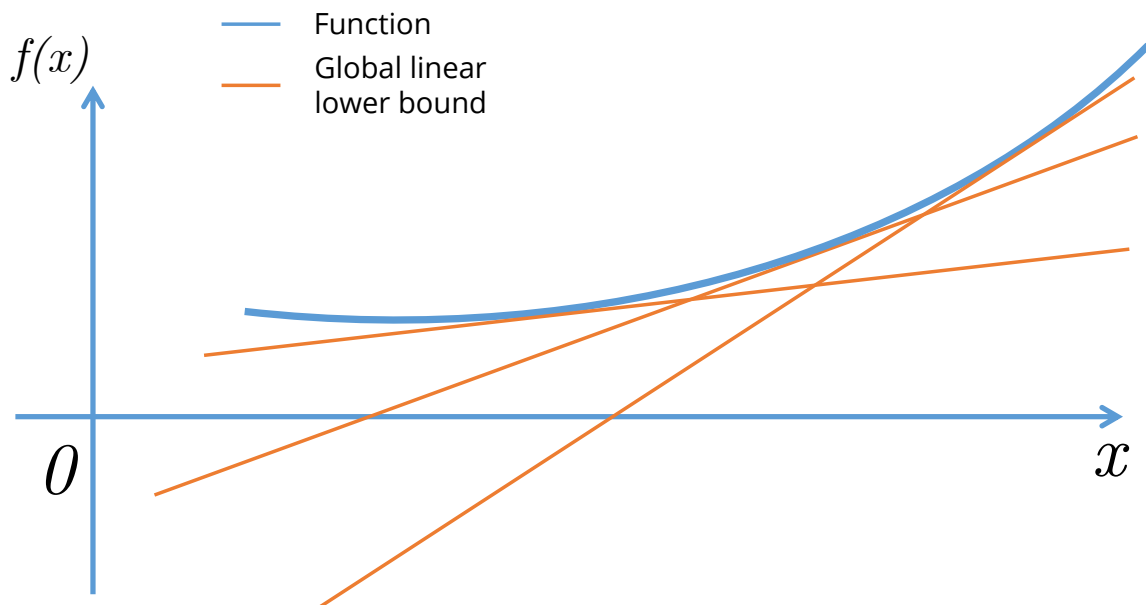
First order differential criterion of convexity

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

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

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

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



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

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

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 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}_β 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}_β is closed.

Reduction to a line

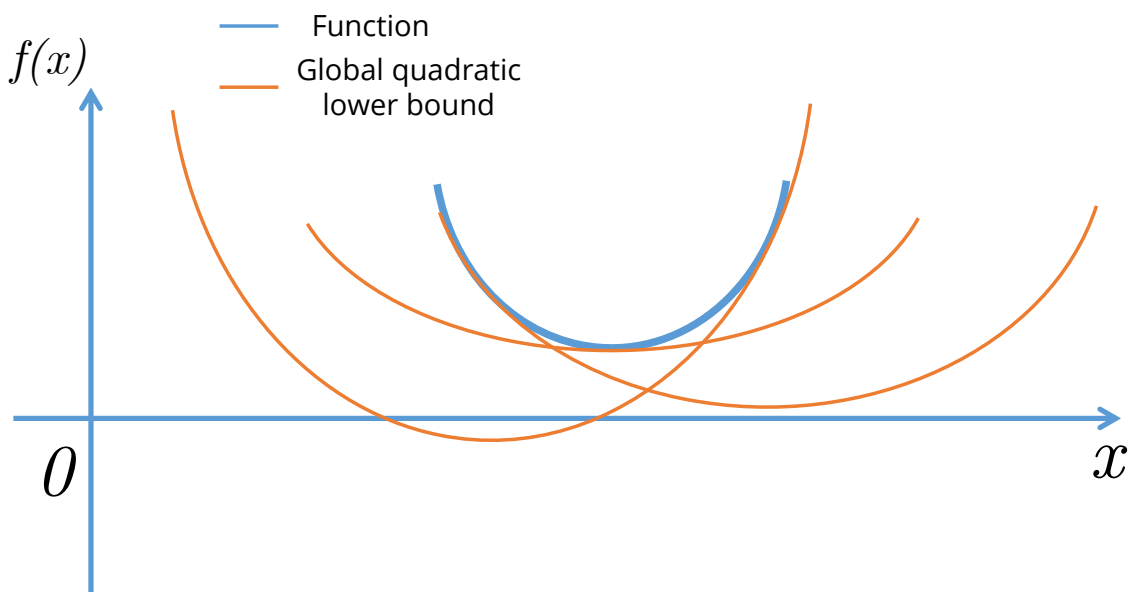
$f : S \rightarrow \mathbb{R}$ is convex if and only if S is convex set and the function $g(t) = f(x + tv)$ defined on $\{t \mid x + tv \in S\}$ is convex for any $x \in S, v \in \mathbb{R}^n$, which allows to check convexity of the scalar function in order to establish convexity of the vector function.

Strong convexity

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

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

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

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

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 $\alpha_i \geq 0$; $\sum_{i=1}^n \alpha_i = 1$ (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 $p(x) \geq 0$, $\int_S p(x) dx = 1$

- If the function $f(x)$ and the set S are convex, then any local minimum $x^* = \arg \min_{x \in S} f(x)$ 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 $f_1(x), \dots, f_m(x)$ are convex, then $f(x) = \max\{f_1(x), \dots, f_m(x)\}$ is convex
- If $f(x, y)$ is convex on x for any $y \in Y$: $g(x) = \sup_{y \in Y} f(x, y)$ 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 $f_1 : S_1 \rightarrow \mathbb{R}$ and $f_2 : S_2 \rightarrow \mathbb{R}$, 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

Other forms of convexity

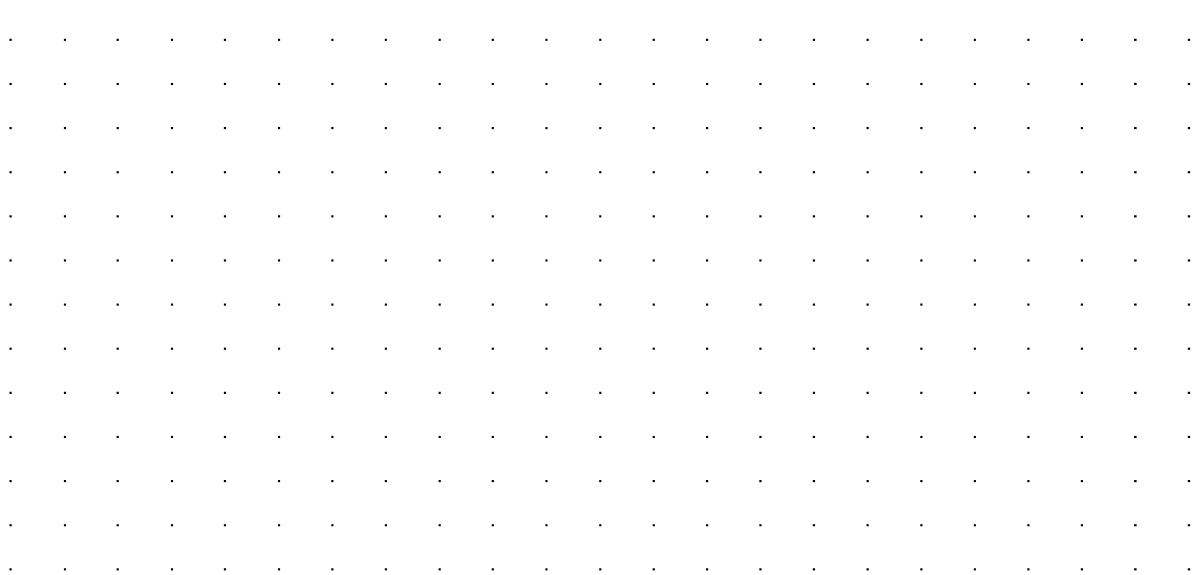
- Log-convex: $\log f$ is convex; Log convexity implies convexity.
- Log-concavity: $\log f$ concave; **not** closed under addition!
- Exponentially convex: $[f(x_i + x_j)] \geq 0$, for x_1, \dots, x_n
- Operator convex: $f(\lambda X + (1 - \lambda)Y) \preceq \lambda f(X) + (1 - \lambda)f(Y)$
- Quasiconvex: $f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}$
- Pseudoconvex: $\langle \nabla f(y), x - y \rangle \geq 0 \rightarrow f(x) \geq f(y)$
- Discrete convexity: $f : \mathbb{Z}^n \rightarrow \mathbb{Z}$; "convexity + matroid theory."

References

- [Steven Boyd lectures](#)
- [Suvrit Sra lectures](#)
- [Martin Jaggi lectures](#)

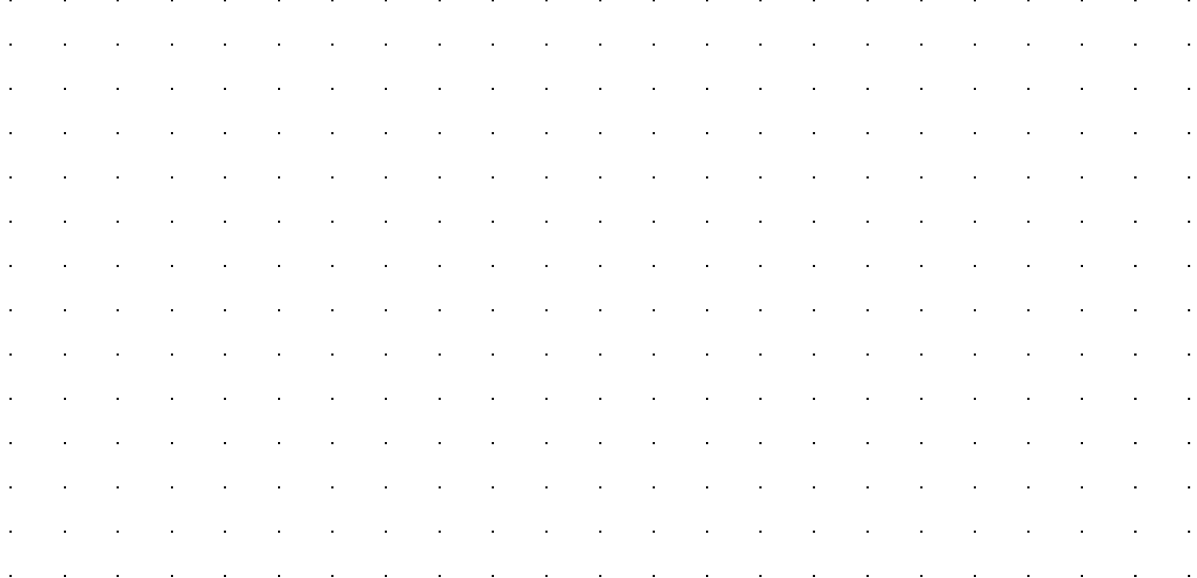
Example 4

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

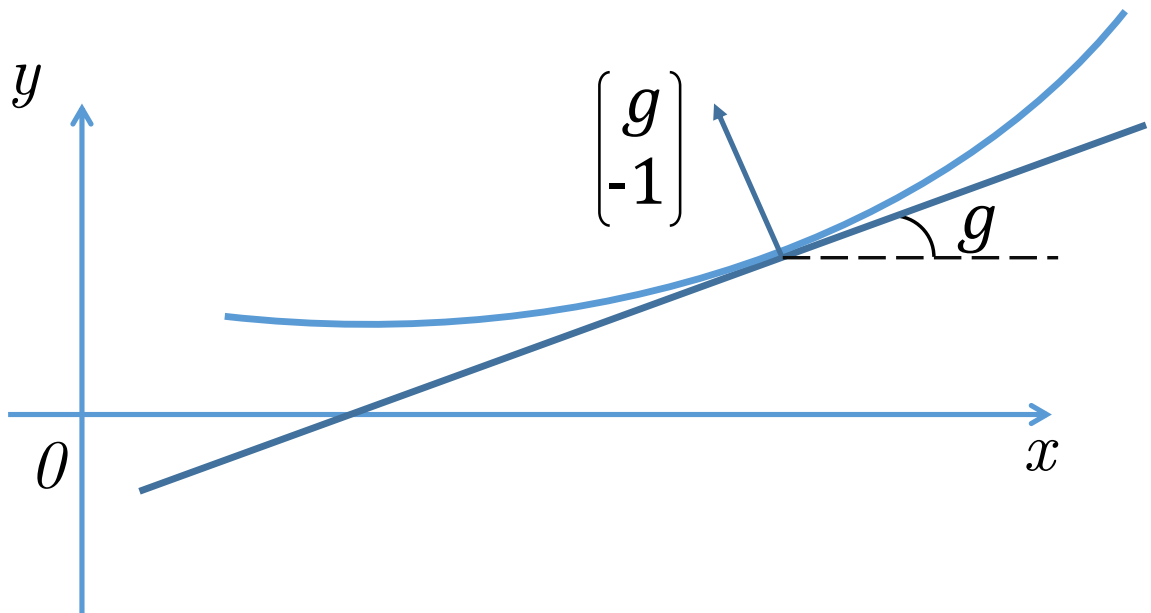


Example 5

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



Example 6



- Если $f(x)$ - дифференцируема, то $g = \nabla f(x_0)$
- Не все непрерывные выпуклые функции дифференцируемы :)

Не хочется лишаться такого вкусного свойства.

Subgradient

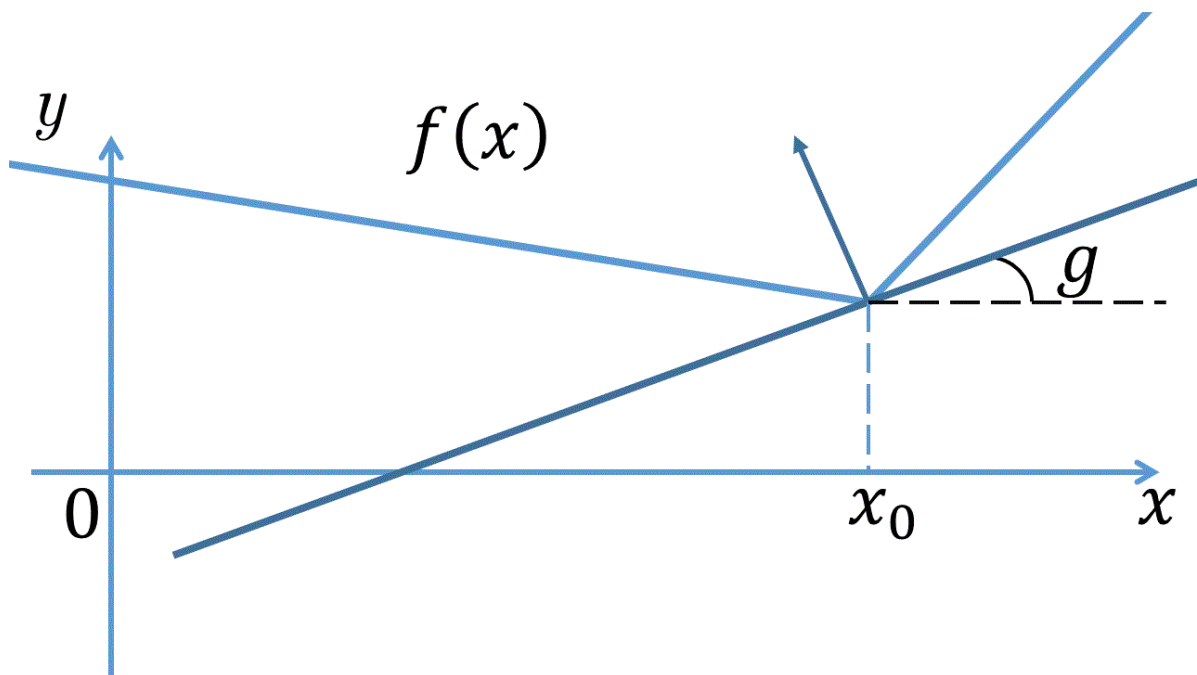
Вектор g называется **субградиентом** функции $f(x) : S \rightarrow \mathbb{R}$ в точке x_0 , если $\forall x \in S$:

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

Subdifferential

Множество всех субградиентов функции $f(x)$ в точке x_0 называется **субдифференциалом** f в x_0 и обозначается $\partial f(x_0)$.

- Если $x_0 \in \text{ri}S$, то $\partial f(x_0)$ выпуклое компактное множество.
- Выпуклая функция $f(x)$ дифференцируема в точке $x_0 \iff \partial f(x_0) = \nabla f(x_0)$
- Если $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, то $f(x)$ - выпукла на S .



Moreau - Rockafellar theorem (subdifferential of a linear combination)

Пусть $f_i(x)$ - выпуклые функции на выпуклых множествах S_i , $i = \overline{1, n}$.

Тогда, если $\bigcap_{i=1}^n \text{ri}S_i \neq \emptyset$ то функция $f(x) = \sum_{i=1}^n a_i f_i(x)$, $a_i > 0$ имеет субдифференциал

$\partial_S f(x)$ на множестве $S = \bigcap_{i=1}^n S_i$ и

$$\partial_S f(x) = \sum_{i=1}^n a_i \partial_{S_i} f_i(x)$$

Dubovitsky - Milutin theorem (subdifferential of a point-wise maximum)

Пусть $f_i(x)$ - выпуклые функции на открытом выпуклом множестве $S \subseteq \mathbb{R}^n$, $x_0 \in S$, а поточечный максимум определяется как $f(x) = \max_i f_i(x)$. Тогда:

$$\partial_S f(x_0) = \text{conv} \left\{ \bigcup_{i \in I(x_0)} \partial_S f_i(x_0) \right\},$$

где $I(x) = \{i \in [1 : m] : f_i(x) = f(x)\}$

Chain rule for subdifferentials

Пусть g_1, \dots, g_m - выпуклые функции на открытом выпуклом множестве $S \subseteq \mathbb{R}^n$, $g = (g_1, \dots, g_m)$ - образованная из них вектор - функция, φ - монотонно неубывающая выпуклая функция на открытом выпуклом множестве $U \subseteq \mathbb{R}^m$, причем $g(S) \subseteq U$. Тогда субдифференциал функции $f(x) = \varphi(g(x))$ имеет вид:

$$\partial f(x) = \bigcup_{p \in \partial \varphi(u)} \left(\sum_{i=1}^m p_i \partial g_i(x) \right),$$

где $u = g(x)$

В частности, если функция φ дифференцируема в точке $u = g(x)$, то формула запишется так:

$$\partial f(x) = \sum_{i=1}^m \frac{\partial \varphi}{\partial u_i}(u) \partial g_i(x)$$

Subdifferential calculus

- $\partial(\alpha f)(x) = \alpha \partial f(x)$, for $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$, f_i - выпуклые функции
- $\partial(f(Ax + b))(x) = A^T \partial f(Ax + b)$, f - выпуклая функция

Examples

Концептуально, различают три способа решения задач на поиск субградиента:

- Теоремы Моро - Рокафеллара, композиции, максимума
- Геометрически
- По определению

1

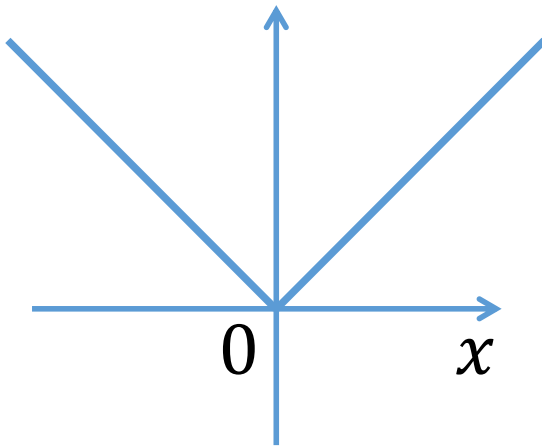
Найти $\partial f(x)$, если $f(x) = |x|$

Решение:

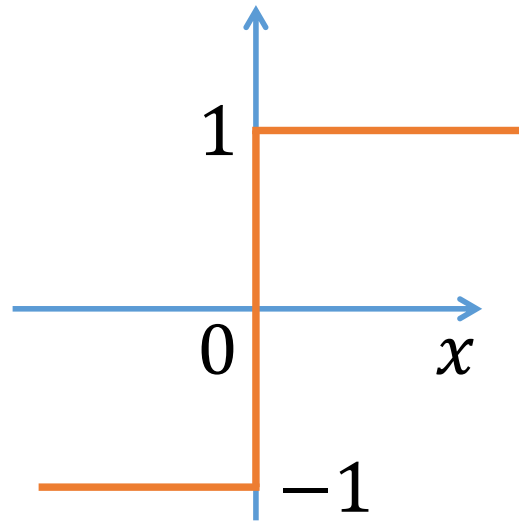
Решить задачу можно либо геометрически (в каждой точке числовой прямой указать угловые коэффициенты прямых, глобально подпирающих функцию снизу), либо по теореме Моро - Рокафеллара, рассмотрев $f(x)$ как композицию выпуклых функций:

$$f(x) = \max\{-x, x\}$$

$$f(x) = |x|$$



$$\partial f(x)$$



2

Найти $\partial f(x)$, если $f(x) = |x - 1| + |x + 1|$

Решение:

Совершенно аналогично применяем теорему Моро - Рокафеллара, учитывая следующее:

$$\partial f_1(x) = \begin{cases} -1, & x < 1 \\ [-1; 1], & x = 1 \\ 1, & x > 1 \end{cases} \quad \partial f_2(x) = \begin{cases} -1, & x < -1 \\ [-1; 1], & x = -1 \\ 1, & x > -1 \end{cases}$$

Таким образом:

$$\partial f(x) = \begin{cases} -2, & x < -1 \\ [-2; 0], & x = -1 \\ 0, & -1 < x < 1 \\ [0; 2], & x = 1 \\ 2, & x > 1 \end{cases}$$

3

Найти $\partial f(x)$, если $f(x) = [\max(0, f_0(x))]^q$. Здесь $f_0(x)$ - выпуклая функция на открытом выпуклом множестве S , $q \geq 1$.

Решение:

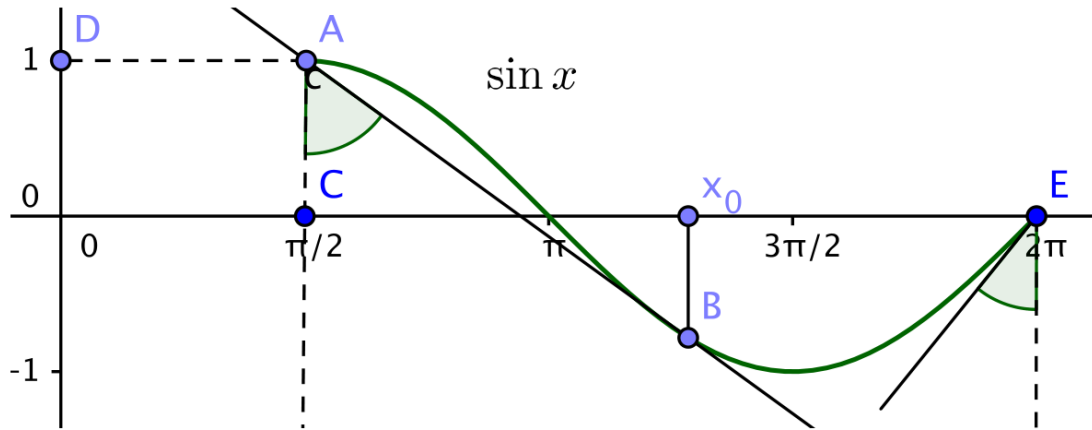
Согласно теореме о композиции (функция $\varphi(x) = x^q$ - дифференцируема), а $g(x) = \max(0, f_0(x))$ имеем: $\partial f(x) = q(g(x))^{q-1} \partial g(x)$

По теореме о поточечном максимуме:

$$\partial g(x) = \begin{cases} \partial f_0(x), & f_0(x) > 0, \\ \{0\}, & f_0(x) < 0 \\ \{a \mid a = \lambda a', 0 \leq \lambda \leq 1, a' \in \partial f_0(x)\}, & f_0(x) = 0 \end{cases}$$

4

Найти $\partial f(x)$, если $f(x) = \sin x, x \in [\pi/2; 2\pi]$



$$\partial f_G(x) = \begin{cases} (-\infty, \cos x_0], & x = \pi/2; \\ \emptyset, & x \in (\pi/2, x_0); \\ \cos x, & x \in [x_0, 2\pi); \\ [1, +\infty), & x = 2\pi. \end{cases}$$

5

Найти $\partial f(x)$, если $f(x) = |c_1^\top x| + |c_2^\top x|$

Решение: Пусть $f_1(x) = |c_1^\top x|$, а $f_2(x) = |c_2^\top x|$. Так как эти функции выпуклы, субдифференциал их суммы равен сумме субдифференциалов. Найдем каждый из них:

$$\partial f_1(x) = \partial (\max\{c_1^\top x, -c_1^\top x\}) = \begin{cases} -c_1, & c_1^\top x < 0 \\ \mathbf{conv}(-c_1; c_1), & c_1^\top x = 0 \\ c_1, & c_1^\top x > 0 \end{cases}$$

$$\partial f_2(x) = \partial (\max\{c_2^\top x, -c_2^\top x\}) = \begin{cases} -c_2, & c_2^\top x < 0 \\ \mathbf{conv}(-c_2; c_2), & c_2^\top x = 0 \\ c_2, & c_2^\top x > 0 \end{cases}$$

Далее интересными представляются лишь различные взаимные расположения векторов c_1 и c_2 , рассмотрение которых предлагается читателю.

6

Найти $\partial f(x)$, если $f(x) = \|x\|_1$

Решение: По определению

$$\|x\|_1 = |x_1| + |x_2| + \dots + |x_n| = s_1 x_1 + s_2 x_2 + \dots + s_n x_n$$

Рассмотрим эту сумму как поточечный максимум линейных функций по x : $g(x) = s^\top x$, где $s_i = \{-1, 1\}$. Каждая такая функция однозначно определяется набором коэффициентов $\{s_i\}_{i=1}^n$.

Тогда по теореме Дубовицкого - Милютина, в каждой точке $\partial f = \mathbf{conv} \left(\bigcup_{i \in I(x)} \partial g_i(x) \right)$

$$\text{Заметим, что } \partial g(x) = \partial (\max\{s^\top x, -s^\top x\}) = \begin{cases} -s, & s^\top x < 0 \\ \mathbf{conv}(-s; s), & s^\top x = 0. \\ s, & s^\top x > 0 \end{cases}$$

Причем, правило выбора "активной" функции поточечного максимума в каждой точке следующее:

- Если j -ая координата точки отрицательна, $s_i^j = -1$
- Если j -ая координата точки положительна, $s_i^j = 1$
- Если j -ая координата точки равна нулю, то подходят оба варианта коэффициентов и соответствующих им функций, а значит, необходимо включать субградиенты этих функций в объединение в теореме Дубовицкого - Милютина.

В итоге получаем ответ:

$$\partial f(x) = \{g : \|g\|_\infty \leq 1, \quad g^\top x = \|x\|_1\}$$

References

- [Lecture Notes for ORIE 6300: Mathematical Programming I by Damek Davis](#)