

# Субградиент и субдифференциал

Даниил Меркулов

Методы оптимизации. МФТИ

## Негладкие задачи

# Задача наименьших квадратов с $\ell_1$ -регуляризацией

$\ell_1$  induces sparsity

$\ell_2$  regularization.  $\|Xw - y\|_2^2 \rightarrow \min_{\|w\|_2 \leq 1}$



$\ell_1$  regularization.  $\|Xw - y\|_2^2 \rightarrow \min_{\|w\|_1 \leq 1}$



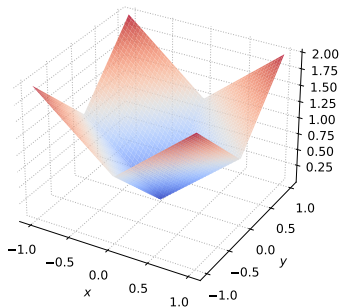
@fminxyz

# Нормы не являются гладкими

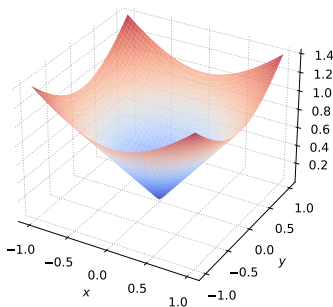
$$\min_{x \in \mathbb{R}^n} f(x),$$

Рассмотрим классическую выпуклую задачу оптимизации. Мы предполагаем, что  $f(x)$  является выпуклой функцией, но теперь мы не требуем гладкости.

$p = 1$  Norm Cone



$p = 2$  Norm Cone



$p = \infty$  Norm Cone

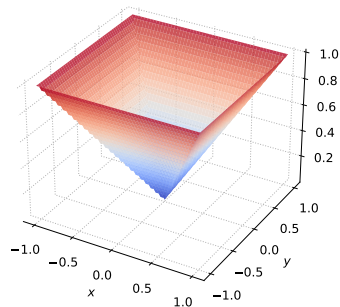


Рис. 1: Нормы конусов для разных  $p$  — нормы не являются гладкими

# Пример Вульфа

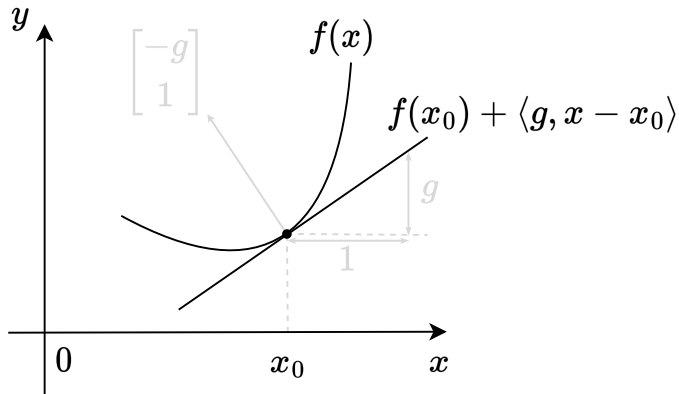
Wolfe's example



Рис. 2: Пример Вульфа. [Открыть в Colab](#)

## Вычисление субградиента

## Линейная нижняя оценка выпуклых функций



Важное свойство непрерывной выпуклой функции  $f(x)$  заключается в том, что для любой выбранной точки  $x_0$  для всех  $x \in \text{dom } f$  выполняется неравенство:

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

Рис. 3: Линейная аппроксимация Тейлора служит глобальной нижней оценкой для выпуклой функции

## Линейная нижняя оценка выпуклых функций



Важное свойство непрерывной выпуклой функции  $f(x)$  заключается в том, что для любой выбранной точки  $x_0$  для всех  $x \in \text{dom } f$  выполняется неравенство:

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

для некоторого вектора  $g$ , т.е. касательная к графику функции является *глобальной* нижней оценкой для функции.

- Если  $f(x)$  дифференцируема, то  $g = \nabla f(x_0)$ .

Рис. 3: Линейная аппроксимация Тейлора служит глобальной нижней оценкой для выпуклой функции



## Линейная нижняя оценка выпуклых функций



Важное свойство непрерывной выпуклой функции  $f(x)$  заключается в том, что для любой выбранной точки  $x_0$  для всех  $x \in \text{dom } f$  выполняется неравенство:

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

для некоторого вектора  $g$ , т.е. касательная к графику функции является *глобальной* нижней оценкой для функции.

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

Рис. 3: Линейная аппроксимация Тейлора служит глобальной нижней оценкой для выпуклой функции

## Линейная нижняя оценка выпуклых функций



Важное свойство непрерывной выпуклой функции  $f(x)$  заключается в том, что для любой выбранной точки  $x_0$  для всех  $x \in \text{dom } f$  выполняется неравенство:

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

для некоторого вектора  $g$ , т.е. касательная к графику функции является *глобальной* нижней оценкой для функции.

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

Рис. 3: Линейная аппроксимация Тейлора служит глобальной нижней оценкой для выпуклой функции

## Линейная нижняя оценка выпуклых функций



Важное свойство непрерывной выпуклой функции  $f(x)$  заключается в том, что для любой выбранной точки  $x_0$  для всех  $x \in \text{dom } f$  выполняется неравенство:

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

для некоторого вектора  $g$ , т.е. касательная к графику функции является *глобальной* нижней оценкой для функции.

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

Мы не хотим потерять такое удобное свойство.

Рис. 3: Линейная аппроксимация Тейлора служит глобальной нижней оценкой для выпуклой функции

## Субградиент и субдифференциал

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

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

## Субградиент и субдифференциал

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

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

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

## Субградиент и субдифференциал

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

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

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



Рис. 4: Субдифференциал — это множество всех возможных субградиентов

# Субградиент и субдифференциал

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

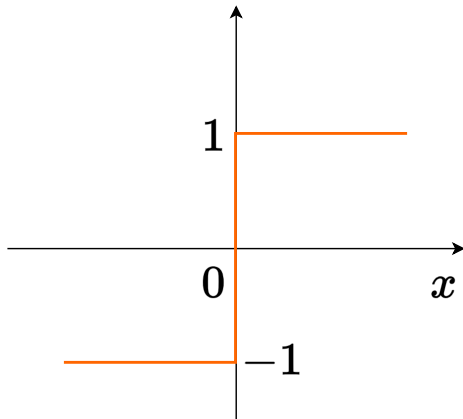
# Субградиент и субдифференциал

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

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



$$\partial f(x)$$





## Свойства субдифференциала

- Если  $x_0 \in \text{ri}(S)$ , то  $\partial f(x_0)$  является выпуклым компактным множеством.

## Свойства субдифференциала

- Если  $x_0 \in \text{ri}(S)$ , то  $\partial f(x_0)$  является выпуклым компактным множеством.
- Выпуклая функция  $f(x)$  дифференцируема в точке  $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$ .

## Свойства субдифференциала

- Если  $x_0 \in \text{ri}(S)$ , то  $\partial f(x_0)$  является выпуклым компактным множеством.
- Выпуклая функция  $f(x)$  дифференцируема в точке  $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$ .
- Если  $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$ , то  $f(x)$  выпукла на  $S$ .

## Свойства субдифференциала

- Если  $x_0 \in \text{ri}(S)$ , то  $\partial f(x_0)$  является выпуклым компактным множеством.
- Выпуклая функция  $f(x)$  дифференцируема в точке  $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$ .
- Если  $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$ , то  $f(x)$  выпукла на  $S$ .

# Свойства субдифференциала

- Если  $x_0 \in \text{ri}(S)$ , то  $\partial f(x_0)$  является выпуклым компактным множеством.
- Выпуклая функция  $f(x)$  дифференцируема в точке  $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$ .
- Если  $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$ , то  $f(x)$  выпукла на  $S$ .

**i** Субдифференциал дифференцируемой функции

Пусть  $f : S \rightarrow \mathbb{R}$  — функция, определенная на множестве  $S$  в евклидовом пространстве  $\mathbb{R}^n$ . Если  $x_0 \in \text{ri}(S)$  и  $f$  дифференцируема в точке  $x_0$ , то либо  $\partial f(x_0) = \emptyset$  либо  $\partial f(x_0) = \{\nabla f(x_0)\}$ . Более того, если функция  $f$  выпукла, то первая ситуация невозможна.

# Свойства субдифференциала

- Если  $x_0 \in \text{ri}(S)$ , то  $\partial f(x_0)$  является выпуклым компактным множеством.
- Выпуклая функция  $f(x)$  дифференцируема в точке  $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$ .
- Если  $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$ , то  $f(x)$  выпукла на  $S$ .

**i** Субдифференциал дифференцируемой функции

Пусть  $f : S \rightarrow \mathbb{R}$  — функция, определенная на множестве  $S$  в евклидовом пространстве  $\mathbb{R}^n$ . Если  $x_0 \in \text{ri}(S)$  и  $f$  дифференцируема в точке  $x_0$ , то либо  $\partial f(x_0) = \emptyset$  либо  $\partial f(x_0) = \{\nabla f(x_0)\}$ . Более того, если функция  $f$  выпукла, то первая ситуация невозможна.

## Доказательство

1. Пусть  $s \in \partial f(x_0)$  для некоторого  $s \in \mathbb{R}^n$  отличного от  $\nabla f(x_0)$ . Пусть  $v \in \mathbb{R}^n$  — единичный вектор. Поскольку  $x_0$  является внутренней точкой множества  $S$ , существует  $\delta > 0$  такое, что  $x_0 + tv \in S$  для всех  $0 < t < \delta$ . По определению субградиента:

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v \rangle$$

# Свойства субдифференциала

- Если  $x_0 \in \text{ri}(S)$ , то  $\partial f(x_0)$  является выпуклым компактным множеством.
- Выпуклая функция  $f(x)$  дифференцируема в точке  $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$ .
- Если  $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$ , то  $f(x)$  выпукла на  $S$ .

**i** Субдифференциал дифференцируемой функции

Пусть  $f : S \rightarrow \mathbb{R}$  — функция, определенная на множестве  $S$  в евклидовом пространстве  $\mathbb{R}^n$ . Если  $x_0 \in \text{ri}(S)$  и  $f$  дифференцируема в точке  $x_0$ , то либо  $\partial f(x_0) = \emptyset$  либо  $\partial f(x_0) = \{\nabla f(x_0)\}$ . Более того, если функция  $f$  выпукла, то первая ситуация невозможна.

## Доказательство

1. Пусть  $s \in \partial f(x_0)$  для некоторого  $s \in \mathbb{R}^n$  отличного от  $\nabla f(x_0)$ . Пусть  $v \in \mathbb{R}^n$  — единичный вектор. Поскольку  $x_0$  является внутренней точкой множества  $S$ , существует  $\delta > 0$  такое, что  $x_0 + tv \in S$  для всех  $0 < t < \delta$ . По определению субградиента:

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v \rangle$$

# Свойства субдифференциала

- Если  $x_0 \in \text{ri}(S)$ , то  $\partial f(x_0)$  является выпуклым компактным множеством.
- Выпуклая функция  $f(x)$  дифференцируема в точке  $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$ .
- Если  $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$ , то  $f(x)$  выпукла на  $S$ .

**i** Субдифференциал дифференцируемой функции

Пусть  $f : S \rightarrow \mathbb{R}$  — функция, определенная на множестве  $S$  в евклидовом пространстве  $\mathbb{R}^n$ . Если  $x_0 \in \text{ri}(S)$  и  $f$  дифференцируема в точке  $x_0$ , то либо  $\partial f(x_0) = \emptyset$  либо  $\partial f(x_0) = \{\nabla f(x_0)\}$ . Более того, если функция  $f$  выпукла, то первая ситуация невозможна.

## Доказательство

1. Пусть  $s \in \partial f(x_0)$  для некоторого  $s \in \mathbb{R}^n$  отличного от  $\nabla f(x_0)$ . Пусть  $v \in \mathbb{R}^n$  — единичный вектор. Поскольку  $x_0$  является внутренней точкой множества  $S$ , существует  $\delta > 0$  такое, что  $x_0 + tv \in S$  для всех  $0 < t < \delta$ . По определению субградиента:

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v \rangle$$

что влечёт:

$$\frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v \rangle$$

для всех  $0 < t < \delta$ . Переходя к пределу при  $t \rightarrow 0$  и используя определение градиента, получаем:

$$\langle \nabla f(x_0), v \rangle = \lim_{t \rightarrow 0; 0 < t < \delta} \frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v \rangle$$



# Свойства субдифференциала

2. Отсюда  $\langle s - \nabla f(x_0), v \rangle \geq 0$ . В силу произвольности  $v$  можно выбрать

$$v = -\frac{s - \nabla f(x_0)}{\|s - \nabla f(x_0)\|},$$

которое приводит к  $s = \nabla f(x_0)$ .

# Свойства субдифференциала

2. Отсюда  $\langle s - \nabla f(x_0), v \rangle \geq 0$ . В силу произвольности  $v$  можно выбрать

$$v = -\frac{s - \nabla f(x_0)}{\|s - \nabla f(x_0)\|},$$

которое приводит к  $s = \nabla f(x_0)$ .

3. Более того, если функция  $f$  выпукла, то согласно дифференциальному условию выпуклости  $f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle$  для всех  $x \in S$ . Но по определению это означает, что  $\nabla f(x_0) \in \partial f(x_0)$ .

# Субдифференцируемость и выпуклость

## i Question

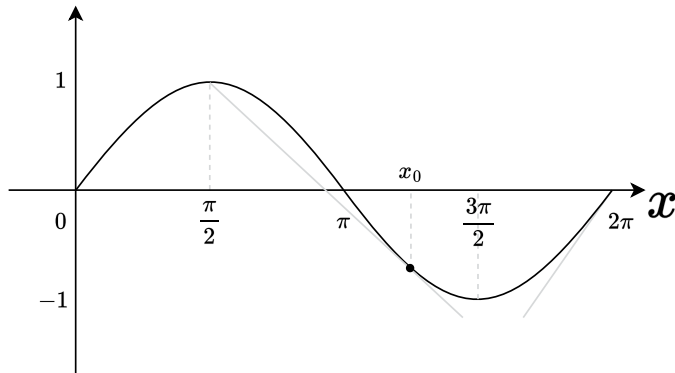
Верно ли, что если функция имеет субдифференциал в некоторой точке, то функция выпукла?

# Субдифференцируемость и выпуклость

## i Question

Верно ли, что если функция имеет субдифференциал в некоторой точке, то функция выпукла?

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



# Субдифференцируемость и выпуклость

## Question

Верно ли, что если функция выпукла, то она имеет субградиент в любой точке?

# Субдифференцируемость и выпуклость

## i Question

Верно ли, что если функция выпукла, то она имеет субградиент в любой точке?

Выпуклость следует из субдифференцируемости в любой точке. Естественный вопрос заключается в том, верно ли обратное: является ли всякая выпуклая функция субдифференцируемой? Оказывается, в общем случае ответ на этот вопрос отрицателен.

Пусть  $f : [0, \infty) \rightarrow \mathbb{R}$  определена как  $f(x) := -\sqrt{x}$ . Тогда,  $\partial f(0) = \emptyset$ .

Предположим, что  $s \in \partial f(0)$  для некоторого  $s \in \mathbb{R}$ . Тогда, по определению, мы должны иметь  $sx \leq -\sqrt{x}$  для всех  $x \geq 0$ . Из этого мы можем вывести  $s \leq -\sqrt{1/x}$  для всех  $x > 0$ . Переходя к пределу при  $x$  стремящемся к 0 справа, мы получаем  $s \leq -\infty$ , что невозможно.

# Вычисление субдифференциалов

**i** Теорема Моро — Роккафеллара  
(субдифференциал линейной комбинации)

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

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

# Вычисление субдифференциалов

**i** Теорема Морро — Роккафеллара  
(субдифференциал линейной комбинации)

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

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

**i** Теорема Дубовицкого — Милютина  
(субдифференциал поточечного максимума)

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

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

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



# Вычисление субдифференциала

- $\partial(\alpha f)(x) = \alpha \partial f(x)$ , для  $\alpha \geq 0$

# Вычисление субдифференциала

- $\partial(\alpha f)(x) = \alpha \partial f(x)$ , для  $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$ ,  $f_i$  — выпуклые функции

# Вычисление субдифференциала

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

# Вычисление субдифференциала

- $\partial(\alpha f)(x) = \alpha \partial f(x)$ , для  $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$ ,  $f_i$  — выпуклые функции
- $\partial(f(Ax + b))(x) = A^T \partial f(Ax + b)$ ,  $f$  — выпуклая функция
- $z \in \partial f(x)$  тогда и только тогда, когда  $x \in \partial f^*(z)$ .

## Связь с выпуклой геометрией

Для выпуклого множества  $S \subseteq \mathbb{R}^n$ , рассмотрим индикаторную функцию  $I_S : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$I_S(x) = I\{x \in S\} = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases}$$

## Связь с выпуклой геометрией

Для выпуклого множества  $S \subseteq \mathbb{R}^n$ , рассмотрим индикаторную функцию  $I_S : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$I_S(x) = I\{x \in S\} = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases}$$

Для  $x \in S$ ,  $\partial I_S(x) = \mathcal{N}_S(x)$ , **нормальный конус** для  $S$  в  $x$ :

$$\mathcal{N}_S(x) = \{g \in \mathbb{R}^n : g^T x \geq g^T y \text{ for any } y \in S\}$$

## Связь с выпуклой геометрией

Для выпуклого множества  $S \subseteq \mathbb{R}^n$ , рассмотрим индикаторную функцию  $I_S : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$I_S(x) = I\{x \in S\} = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases}$$

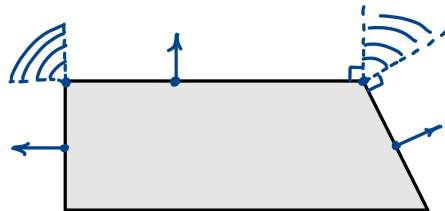
Для  $x \in S$ ,  $\partial I_S(x) = \mathcal{N}_S(x)$ , **нормальный конус** для  $S$  в  $x$ :

$$\mathcal{N}_S(x) = \{g \in \mathbb{R}^n : g^T x \geq g^T y \text{ for any } y \in S\}$$

**Почему?** По определению субградиента  $g$ ,

$$I_S(y) \geq I_S(x) + g^T(y - x) \quad \text{for all } y$$

- При  $y \notin S$ ,  $I_S(y) = \infty$



## Связь с выпуклой геометрией

Для выпуклого множества  $S \subseteq \mathbb{R}^n$ , рассмотрим индикаторную функцию  $I_S : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$I_S(x) = I\{x \in S\} = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases}$$

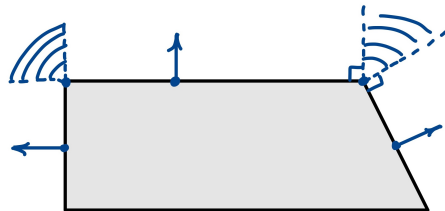
Для  $x \in S$ ,  $\partial I_S(x) = \mathcal{N}_S(x)$ , **нормальный конус** для  $S$  в  $x$ :

$$\mathcal{N}_S(x) = \{g \in \mathbb{R}^n : g^T x \geq g^T y \text{ for any } y \in S\}$$

**Почему?** По определению субградиента  $g$ ,

$$I_S(y) \geq I_S(x) + g^T(y - x) \quad \text{for all } y$$

- При  $y \notin S$ ,  $I_S(y) = \infty$
- При  $y \in S$ , this means  $0 \geq g^T(y - x)$





## Условия Оптимальности

Для любой  $f$  (выпуклой или нет),

$$f(x^*) = \min_x f(x) \iff 0 \in \partial f(x^*)$$

То есть,  $x^*$  является точкой минимума тогда и только тогда, когда  $0$  является субградиентом функции  $f$  в точке  $x^*$ . Это утверждение называется **субградиентное условие оптимальности**

Почему? Легко: если  $g = 0$  является субградиентом, это значит что для всех  $y$

$$f(y) \geq f(x^*) + 0^T(y - x^*) = f(x^*)$$

Отметим, что для выпуклой и дифференцируемой функций  $f$  верно

$$\partial f(x) = \{\nabla f(x)\}$$

## Получение условия оптимальности первого порядка

Попробуем записать общее **условие оптимальности первого порядка**. Вспомним, что решением задачи

$$\min_x f(x) \text{ subject to } x \in S$$

является точка  $x$ , для выпуклой и дифференцируемой  $f$ , в том и только в том случае, если

$$\nabla f(x)^T (y - x) \geq 0 \quad \text{for all } y \in S$$

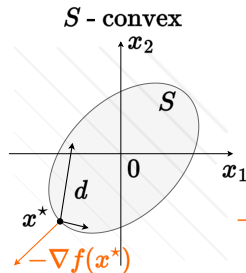
Интуитивно: написанное выше означает, что функция увеличивается по мере движения от точки  $x$ . Как это доказать? Во-первых, перепишем задачу в следующем виде:

$$\min_x f(x) + I_S(x)$$

Теперь воспользуемся условием оптимальности в субградиентной форме:

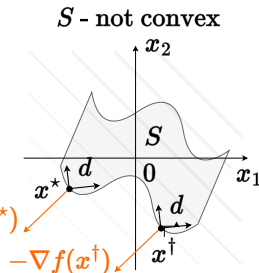
$$0 \in \partial(f(x) + I_S(x))$$

$$f(x) = x_1 + x_2 \rightarrow \min_{x_1, x_2 \in \mathbb{R}^2}$$



$$\langle -\nabla f(x^*), d \rangle \leq 0$$

$x^*$  - optimal



$$\langle -\nabla f(x^\dagger), d \rangle \leq 0$$

$x^\dagger$  - not optimal

## Получение условия оптимальности первого порядка

Заметим, что

$$0 \in \partial(f(x) + I_S(x))$$

## Получение условия оптимальности первого порядка

Заметим, что

$$0 \in \partial(f(x) + I_S(x))$$

$$\Leftrightarrow 0 \in \{\nabla f(x)\} + \mathcal{N}_S(x)$$

## Получение условия оптимальности первого порядка

Заметим, что

$$0 \in \partial(f(x) + I_S(x))$$

$$\Leftrightarrow 0 \in \{\nabla f(x)\} + \mathcal{N}_S(x)$$

$$\Leftrightarrow -\nabla f(x) \in \mathcal{N}_S(x)$$

## Получение условия оптимальности первого порядка

Заметим, что

$$0 \in \partial(f(x) + I_S(x))$$

$$\Leftrightarrow 0 \in \{\nabla f(x)\} + \mathcal{N}_S(x)$$

$$\Leftrightarrow -\nabla f(x) \in \mathcal{N}_S(x)$$

$$\Leftrightarrow -\nabla f(x)^T x \geq -\nabla f(x)^T y \text{ for all } y \in S$$

# Получение условия оптимальности первого порядка

Заметим, что

$$0 \in \partial(f(x) + I_S(x))$$

$$\Leftrightarrow 0 \in \{\nabla f(x)\} + \mathcal{N}_S(x)$$

$$\Leftrightarrow -\nabla f(x) \in \mathcal{N}_S(x)$$

$$\Leftrightarrow -\nabla f(x)^T x \geq -\nabla f(x)^T y \text{ for all } y \in S$$

$$\Leftrightarrow \nabla f(x)^T (y - x) \geq 0 \text{ for all } y \in S$$

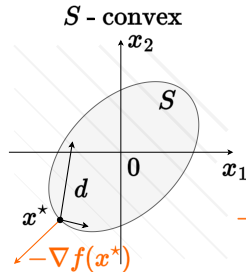
что и требовалось.

Замечание: условие  $0 \in \partial f(x) + \mathcal{N}_S(x)$  является

**общим условием** оптимальности для выпуклых задач.

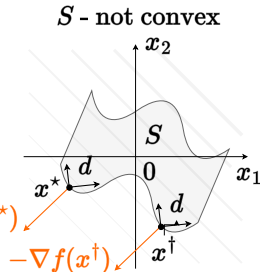
Однако с ним не всегда удобно работать (ККТ удобнее, про них позже).

$$f(x) = x_1 + x_2 \rightarrow \min_{x_1, x_2 \in \mathbb{R}^2}$$



$$\langle -\nabla f(x^*), d \rangle \leq 0$$

$x^*$  - optimal



$$\langle -\nabla f(x^\dagger), d \rangle \leq 0$$

$x^\dagger$  - not optimal

## Пример 1

### Example

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



## Пример 1

### Example

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

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

## Пример 1

### Example

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

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

Итак,

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

## Пример 2

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

## Пример 2

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

Согласно теореме о производной композиции функций (функция  $\varphi(x) = x^q$  дифференцируема) и обозначая  $g(x) = \max(0, f_0(x))$ , имеем:

$$\partial f(x) = q(g(x))^{q-1} \partial g(x)$$

## Пример 2

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

Согласно теореме о производной композиции функций (функция  $\varphi(x) = x^q$  дифференцируема) и обозначая  $g(x) = \max(0, f_0(x))$ , имеем:

$$\partial f(x) = q(g(x))^{q-1} \partial g(x)$$

По теореме о субдифференциале поточечного максимума

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

### Пример 3. Субдифференциал нормы

Пусть  $V$  - конечномерное евклидово пространство, и  $x_0 \in V$ . Пусть  $\|\cdot\|$  - произвольная норма в пространстве  $V$ , и пусть  $\|\cdot\|_*$  - соответствующая сопряженная норма. Тогда,

$$\partial\|\cdot\|(x_0) = \begin{cases} B_{\|\cdot\|_*}(0, 1), & \text{если } x_0 = 0, \\ \{s \in V : \|s\|_* \leq 1; \langle s, x_0 \rangle = \|x_0\|\} = \{s \in V : \|s\|_* = 1; \langle s, x_0 \rangle = \|x_0\|\}, & \text{иначе.} \end{cases}$$

Где  $B_{\|\cdot\|_*}(0, 1)$  есть замкнутый единичный относительно сопряженной нормы шар с центром в нуле. Другими словами, вектор  $s \in V$  с  $\|s\|_* = 1$  является субградиентом нормы  $\|\cdot\|$  в точке  $x_0 \neq 0$  тогда и только тогда, когда неравенство Гёльдера  $\langle s, x_0 \rangle \leq \|x_0\|$  переходит в равенство.

### Пример 3. Субдифференциал нормы

Пусть  $V$  - конечномерное евклидово пространство, и  $x_0 \in V$ . Пусть  $\|\cdot\|$  - произвольная норма в пространстве  $V$ , и пусть  $\|\cdot\|_*$  - соответствующая сопряженная норма. Тогда,

$$\partial\|\cdot\|(x_0) = \begin{cases} B_{\|\cdot\|_*}(0, 1), & \text{если } x_0 = 0, \\ \{s \in V : \|s\|_* \leq 1; \langle s, x_0 \rangle = \|x_0\|\} = \{s \in V : \|s\|_* = 1; \langle s, x_0 \rangle = \|x_0\|\}, & \text{иначе.} \end{cases}$$

Где  $B_{\|\cdot\|_*}(0, 1)$  есть замкнутый единичный относительно сопряженной нормы шар с центром в нуле. Другими словами, вектор  $s \in V$  с  $\|s\|_* = 1$  является субградиентом нормы  $\|\cdot\|$  в точке  $x_0 \neq 0$  тогда и только тогда, когда неравенство Гёльдера  $\langle s, x_0 \rangle \leq \|x_0\|$  переходит в равенство.

Пусть  $s \in V$ . По определению  $s \in \partial\|\cdot\|(x_0)$  если и только если

$$\langle s, x \rangle - \|x\| \leq \langle s, x_0 \rangle - \|x_0\|, \text{ for all } x \in V,$$

что равносильно

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} \leq \langle s, x_0 \rangle - \|x_0\|.$$

По определению супремума, последнее равносильно

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} = \langle s, x_0 \rangle - \|x_0\|.$$

### Пример 3. Субдифференциал нормы

Пусть  $V$  - конечномерное евклидово пространство, и  $x_0 \in V$ . Пусть  $\|\cdot\|$  - произвольная норма в пространстве  $V$ , и пусть  $\|\cdot\|_*$  - соответствующая сопряженная норма. Тогда,

$$\partial\|\cdot\|(x_0) = \begin{cases} B_{\|\cdot\|_*}(0, 1), & \text{если } x_0 = 0, \\ \{s \in V : \|s\|_* \leq 1; \langle s, x_0 \rangle = \|x_0\|\} = \{s \in V : \|s\|_* = 1; \langle s, x_0 \rangle = \|x_0\|\}, & \text{иначе.} \end{cases}$$

Где  $B_{\|\cdot\|_*}(0, 1)$  есть замкнутый единичный относительно сопряженной нормы шар с центром в нуле. Другими словами, вектор  $s \in V$  с  $\|s\|_* = 1$  является субградиентом нормы  $\|\cdot\|$  в точке  $x_0 \neq 0$  тогда и только тогда, когда неравенство Гёльдера  $\langle s, x_0 \rangle \leq \|x_0\|$  переходит в равенство.

Пусть  $s \in V$ . По определению  $s \in \partial\|\cdot\|(x_0)$  если и только если

$$\langle s, x \rangle - \|x\| \leq \langle s, x_0 \rangle - \|x_0\|, \text{ for all } x \in V,$$

что равносильно

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} \leq \langle s, x_0 \rangle - \|x_0\|.$$

По определению супремума, последнее равносильно

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} = \langle s, x_0 \rangle - \|x_0\|.$$

Важно отметить, что выражение слева есть супремум из определения сопряженной функции по Фенхелю для нормы, которая, как известно, записывается так:

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} = \begin{cases} 0, & \text{if } \|s\|_* \leq 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

Таким образом, выражение равносильно  $\|s\|_* \leq 1$  и  $\langle s, x_0 \rangle = \|x_0\|$ .



### Пример 3. Субдифференциал нормы

Следовательно, остаётся заметить, что для  $x_0 \neq 0$  неравенство  $\|s\|_* \leq 1$  должно переходить в равенство, поскольку при  $\|s\|_* < 1$  неравенство Гёльдера влечёт  $\langle s, x_0 \rangle \leq \|s\|_* \|x_0\| < \|x_0\|$ .

Сопряженная норма в примере выше появилась не случайно. Оказывается, что совершенно аналогичным образом для произвольной функции  $f$  (не только для нормы) её субдифференциал может быть описан в терминах двойственного объекта - сопряженной по Фенхелю функции.