Линеаризация

- $f(x)$ - дифференцируемая функция
- $\nabla f(x)$ - градиент
- $l(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle$
- $l(x)$ - касательная гиперплоскость к графику функции f в точке $(x_0, f(x_0))$

Определение 4

$f(x)$ - функция, определенная на множестве $X \subseteq \mathbb{R}^n$, $a \in \mathbb{R}^n$ - **субградиент** функции $f(x)$ в точке $x_0 \in X$, если

$$f(x) - f(x_0) \geq \langle a, x - x_0 \rangle \qquad \forall x \in X$$

Определение 5

$\partial_X f(x_0)$ - **субдифференциал** в точке x_0 функции $f(x)$ - это множество всех субградиентов $f(x)$ в точке x_0

Пример

Пусть $f : \mathbb{R} \to \mathbb{R}$ - функция $f(x) = |x|$, тогда $\partial f(0) = [-1, 1]$.

Доказательство

- 1. $[-1, 1] \subseteq \partial f(0)$ - ?
  - a - субградиент в т. x_0, если $f(x) \geq f(x_0) + \langle a, x - x_0 \rangle \quad \forall x \in \mathrm{X}$
  - $x_0 = 0, \ \mathrm{X} = \mathbb{R}$
  - пусть $a \in [-1, 1]$
  - тогда $ax \leq |a||x| \leq |x| \quad \forall x \in \mathbb{R}$
  - т.е. по определению $a \in \partial f(0)$
  - $[-1, 1] \subseteq \partial f(0)$

Пример

Пусть $f : \mathbb{R} \to \mathbb{R}$ - функция $f(x) = |x|$, тогда $\partial f(0) = [-1, 1]$.

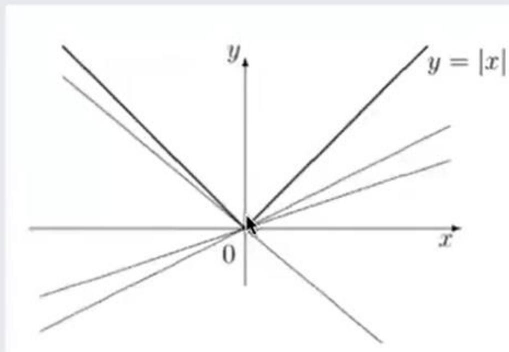
Доказательство

- 2. $[-1, 1] \supseteq \partial f(0)$ - ?
  - от противного
  - a - субградиент в т. x_0, если $f(x) \geq f(x_0) + \langle a, x - x_0 \rangle \quad \forall x \in \mathrm{X}$
  - по опр-ию $|x| \geq ax \quad \forall x \in \mathbb{R}$
  - пусть $\exists a \in \partial f(0)$ и $|a| > 1$
  - пусть $x = a$, тогда $|a| \geq |a|^2 \to$ противоречие
  - $[-1, 1] \supseteq \partial f(0)$

Пример

Пусть $f(x) = |x|$, тогда

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



Утверждение 1

Пусть $f : \mathrm{X} \to \mathbb{R}$, $x_0 \in \mathrm{int}\mathrm{X}$, $f(x)$ дифференцируема в точке x_0. Тогда либо $\partial f(x_0) = \varnothing$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$.
Если $f(x)$ - выпуклая, то первый вариант невозможен.

intX – внутренние точки X

Доказательство

- пусть $a \in \partial f(x_0)$, $a \neq \nabla f(x_0)$
- h - единичный вектор, $\|h\| = 1$
- $x_0 \in \mathrm{int} \mathrm{X} \ \rightarrow \ \exists \varepsilon > 0$:
$$x_0 + th \in \mathrm{X} \quad \forall\, 0 < t < \varepsilon$$
- субградиент:
$$f(x_0 + th) \geq f(x_0) + t\langle a, h\rangle$$
-
$$\langle \nabla f(x_0), h\rangle = \lim_{t \to 0,\, 0 < t < \varepsilon} \frac{f(x_0 + th) - f(x_0)}{t} \geq \langle a, h\rangle$$

Доказательство

- $\langle a - \nabla f(x_0),\, h\rangle \leq 0$
- положим $h = \frac{a - \nabla f(x_0)}{\|a - \nabla f(x_0)\|}$
- $\langle a - \nabla f(x_0),\, h\rangle = \|a - \nabla f(x_0)\| \leq 0 \quad \rightarrow \quad a = \nabla f(x_0)$
- если $f(x)$ - выпуклая функция, то
$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0\rangle \ \ \forall x \in \mathrm{X}$$
- по определению субградиента
$$\nabla f(x_0) \in \partial f(x_0)$$

Утверждение 2

Пусть $f : \mathrm{X} \to \mathbb{R}$, X - выпуклое множество. Если $f(x)$ субдифференцируема, т.е. субдифференциал функции f не является пустым множеством для всех точек X, то $f(x)$ – выпуклая.

Доказательство

Пусть $x, y \in \mathrm{X}$, $\alpha \in [0,1]$, $y_\alpha = x + \alpha(y - x) \in \mathrm{X}$, $g \in \partial_{\mathrm{X}} f(y_\alpha)$. Тогда

$$\begin{aligned} f(y) &\geq f(y_\alpha) + \langle g, y - y_\alpha \rangle = f(y_\alpha) + (1-\alpha)\langle g, y - x \rangle, \\ f(x) &\geq f(y_\alpha) + \langle g, x - y_\alpha \rangle = f(y_\alpha) - \alpha \langle g, y - x \rangle. \end{aligned}$$

Складывая полученные неравенства с коэффициентами α и $1-\alpha$ соответственно, получим $\alpha f(y) + (1-\alpha) f(x) \geq f(y_\alpha)$, то есть функция f является выпуклой.

Верно ли обратное?

Нет

Пример

Пусть $f(x) = -\sqrt{x}, \quad x \in \mathbb{R}_+$. Тогда $\partial f(0) = \varnothing$.

Доказательство

- Пусть $a \in \mathbb{R}$, $a \in \partial f(0)$,
- тогда $ax \leq -\sqrt{x} \quad \forall x \geq 0$,
- $a \leq -\frac{1}{\sqrt{x}} \quad \forall x > 0$,
- переходим к пределу $x \to 0$, $x > 0$,
- $a \leq -\infty$ - невозможно.

Теорема 8 (см. Лемму 3.1.7 из книги Нестерова 2010 года)

Пусть $f : \mathrm{X} \to \mathbb{R}$ - выпуклая функция, $x_0 \in \mathrm{int}\mathrm{X}$. Тогда $f(x)$ дифференцируема в точке x_0, тогда и только тогда, когда субдифференциал $\partial f(x_0)$ содержит ровно один элемент. В этом случае $\partial f(x_0) = \{\nabla f(x_0)\}$.

Геометрический смысл субградиента

- Критерий выпуклости 1го порядка

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

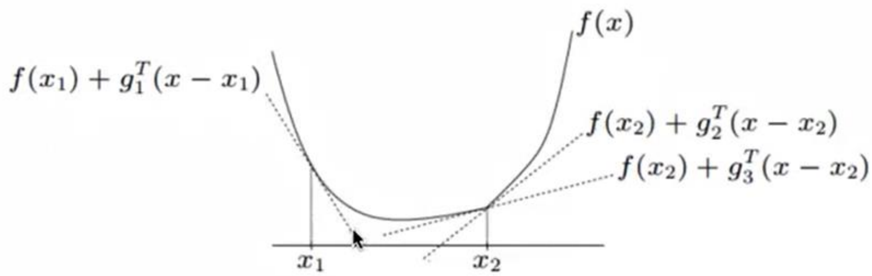
$$\forall x \in \mathrm{X}$$

- Субградиент в т. x_0 - вектор a

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

$$\forall x \in \mathrm{X}$$

Геометрический смысл субдифференциала



- Опорная гиперплоскость к графику выпуклой функции $f(x)$ в точке $(x_2, f(x_2))$

$$\{(x,t) \mid t = f(x_2) + g^\top (x - x_2)\}$$

- g - субградиент

Следствие 1

Для функции $f(x)$ выпуклой на выпуклом числовом множестве $\mathrm{X} \subseteq \mathbb{R}$:

$$\partial_{\mathrm{X}} f(x^*) = [f'_-(x^*),\ f'_+(x^*)],$$

$f'_-(x^*)$, $f'_+(x^*)$ - лево- и правосторонние производные f в точке x^*.

Следствие 2

В любой точке $x_0 \in \mathrm{X}$ субдифференциал $\partial f(x_0)$ является выпуклым и замкнутым множеством.

Теорема 9 (см. Теорему 3.1.13 из книги Нестерова 2010 года)

Пусть $f : \mathrm{X} \to \mathbb{R}$ - выпуклая и замкнутая функция, $x_0 \in \mathrm{int}\mathrm{X}$.

Тогда множество $\partial f(x_0) \neq \varnothing$ и является выпуклым компактом.

Напомним, что *надграфиком* функции $f : E \to \mathbb{R}$, заданной на множестве E, называется множество $\operatorname{Epi} f := \{(x,t) \in E \times \mathbb{R} : f(x) \le t\}$.

Определение 1.1 (Замкнутые функции)**.** Пусть $f : E \to \mathbb{R}$ — функция, заданная на множестве E в нормированном пространстве V. Функция f называется *замкнутой*, если $\operatorname{Epi} f$ является замкнутым множеством в пространстве $V \oplus \mathbb{R}$.

Связь между субдифференциалом выпуклой функции и производной по направлению

Теорема 10

Пусть f – выпуклая замкнутая функция. Тогда для любых $x_0 \in \mathrm{int}(\mathrm{dom} f)$ и $p \in \mathbb{R}^d$ имеет место равенство

$$f'(x_0; p) = \max\{\langle g, p\rangle \mid g \in \partial f(x_0)\}, \tag{7}$$

где $f'(x_0; p)$ – производная функции f в точке x_0 по направлению p:

$$f'(x_0; p) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \left[f(x_0 + \alpha p) - f(x_0) \right].$$

Доказательство

Заметим, что

$$f'(x_0; p) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \left[f(x_0 + \alpha p) - f(x_0) \right] \geq \langle g, p \rangle, \tag{8}$$

где g – произвольный вектор из $\partial f(x_0)$. Поэтому субдифференциал функции $f'(x_0; p)$ в точке $p = 0$ является непустым и $\partial f(x_0) \subseteq \partial_p f'(x_0; 0)$.

Докажем, что из выпуклости функции f следует следующее утверждение: если $x, y \in \mathrm{dom} f$, $\beta \geq 0$ и $y + \beta(y - x) \in \mathrm{dom} f$, то выполнено неравенство

$$f(y + \beta(y - x)) \geq f(y) + \beta(f(y) - f(x)).$$

Доказательство

Действительно, пусть $\alpha = \frac{\beta}{1+\beta}$, $u = y + \beta(y - x)$. Тогда

$$y = \frac{1}{1+\beta}(u + \beta x) = (1 - \alpha)u + \alpha x.$$

Поэтому из выпуклости функции f получаем

$$f(y) \leq (1 - \alpha) f(u) + \alpha f(x) = \frac{1}{1+\beta} f(u) + \frac{\beta}{1+\beta} f(x),$$

а значит,

$$f(y + \beta(y - x)) \geq f(y) + \beta(f(y) - f(x)).$$

Доказательство

Теперь покажем, что

$$f(y) \geq f(x) + f'(x; y - x).$$

Действительно, пусть $y_\alpha = x + \alpha(y - x)$, $\alpha \in (0, 1]$. Тогда из предыдущего неравенства (с $\beta = \frac{1-\alpha}{\alpha}$) получаем

$$f(y) = f\left(y_\alpha + \frac{1-\alpha}{\alpha}(y_\alpha - x)\right) \geq f(y_\alpha) + \frac{(1-\alpha)f(y_\alpha) - f(x)}{\alpha}.$$

Переходя к пределу по $\alpha \downarrow 0$ получаем требуемое неравенство, из которого следует:

$$f(y) \geq f(x_0) + f'(x_0; y - x_0) \geq f(x_0) + \langle g, y - x_0 \rangle, \quad g \in \partial_p f'(x_0; 0).$$

Доказательство

Таким образом, $\partial_p f'(x_0; 0) \subseteq \partial f(x_0)$, и, значит, $\partial f(x_0) \equiv \partial_p f'(x_0; 0)$. Рассмотрим $g_p \in \partial_p f'(x_0; p)$. По определению субградиента для всех $v \in \mathbb{R}^d$ и $\tau > 0$ имеем

$$\tau f'(x_0; v) = f'(x_0; \tau v) \geq f'(x_0; p) + \langle g_p, \tau v - p \rangle.$$

Переходя к пределу при $\tau \to \infty$, заключаем, что $f'(x_0; v) \geq \langle g_p, v \rangle$, а переходя к пределу при $\tau \to 0$, получаем $f'(x_0; p) - \langle g_p, p \rangle \leq 0$. Однако неравенство $f'(x_0; v) \geq \langle g_p, v \rangle$ подразумевает, что $g_p \in \partial_p f'(x_0; 0)$. Поэтому, сравнивая неравенства (8) и $f'(x_0; p) - \langle g_p, p \rangle \leq 0$ делаем вывод, что $\langle g_p, p \rangle = f'(x_0; p)$.

Субдифференциал евклидовой нормы

Пример

Пусть $f(x) = \|x\|$, $\mathrm{X} = \mathbb{R}^d$, тогда

$$\partial f(x) = \begin{cases} \mathrm{B}_1(0) & x = 0, \\ \frac{x}{\|x\|} & x \neq 0. \end{cases}$$

Доказательство

1. - пусть $a \in \mathrm{B}_1(0) \rightarrow \|a\| \leq 1$
   - $a^\top x \leq \|x\|\|a\| = \|x\| = f(x)$
   - $a \in \partial f(0)$
2. - пусть $a \notin \mathrm{B}_1(0) \rightarrow \|a\| > 1$
   - $a^\top x = \|a\|^2 > \|a\| = f(x),\ x = a$
   - $a \notin \partial f(0)$

B1(0) – шар

1. Линейная комбинация функций (см. Лемму 3.1.9 из книги Нестерова 2010 года)

Пусть $f_1(x)$ и $f_2(x)$ – выпуклые замкнутые функции и $\alpha_1,\ \alpha_2 \geq 0$. Тогда функция $f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x)$ является выпуклой и замкнутой и

$$\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$$

для любого $x \in \mathrm{int}(\mathrm{dom}\ f) = \mathrm{int}(\mathrm{dom}\ f_1) \cap \mathrm{int}(\mathrm{dom}\ f_2)$.

Доказательство

- Первая часть утверждения следует из того, что неотрицательная выпуклая комбинация выпуклых функций - выпуклая функция.
- Докажем формулу для субдифференциала. Рассмотрим $x_0 \in \mathrm{int}(\mathrm{dom}\ f_1) \cap \mathrm{int}(\mathrm{dom}\ f_2)$. Тогда $\forall p \in \mathbb{R}^d$ имеем

$$
\begin{aligned}
f'(x_0; p) &= \alpha_1 f_1'(x_0; p) + \alpha_2 f_2'(x_0; p) \\
&= \max\{\langle g_1, \alpha_1 p\rangle \mid g_1 \in \partial f_1(x_0)\} + \max\{\langle g_2, \alpha_2 p\rangle \mid g_2 \in \partial f_2(x_0)\} \\
&= \max\{\langle \alpha_1 g_1 + \alpha_2 g_2, p\rangle \mid g_1 \in \partial f_1(x_0),\ g_2 \in \partial f_2(x_0)\} \\
&= \max\{\langle g, p\rangle \mid g \in \alpha_1 \partial f_1(x_0) + \alpha_2 \partial f_2(x_0)\}.
\end{aligned}
$$

Множества $\partial f_1(x_0)$ и $\partial f_2(x_0)$ — выпуклые компакты, отсюда, используя Теорему 11, получаем требуемое утверждение.

2. Аффинное преобразование функции

Пусть функция $f(y)$ - выпукла и замкнута функция на $\mathrm{dom}\ f \subseteq \mathbb{R}^m$.

Рассмотрим линейный оператор $\mathcal{A}(x) = Ax + b : R^n \to \mathbb{R}^m$. Функция $\varphi(x) = f(\mathcal{A}(x))$ также будет выпуклой и замкнутой с областью определения $\mathrm{dom}\varphi = \{x \mid \mathcal{A}(x) \in \mathrm{dom} f\}$. При этом для любого $x \in \mathrm{int}(\mathrm{dom}\varphi)$ выполняется неравенство

$$
\partial\varphi(x) = \mathrm{A}^\top \partial f\,(\mathcal{A}(x)).
$$

Доказательство

- Первая часть утверждения следует из утверждения, что выпуклость является аффинно-инвариантным свойством.
- Докажем формулу для субдифференциала.
  Пусть $y_0 = \mathcal{A}(x_0)$. Тогда для всех $p \in \mathbb{R}^n$ имеет место равенство

$$\varphi'(x_0, p) = f'(y_0; Ap) = \max\{\langle g, Ap\rangle \mid g \in \partial f(y_0)\}$$
$$= \max\{\langle \bar{g}, p\rangle \mid \bar{g} \in A^\top \partial f(y_0).\}$$

  Используя теорему о связи субдифференциала выпуклой функции и производной по направлению (7), а так же Теорему 11, получаем

$$\partial\varphi(x) = \mathrm{A}^\top \partial f\left(\mathcal{A}(x_0)\right).$$

3. Максимум функций

Пусть функции $f_i(x)$, $i = 1, \ldots, m$, выпуклые и замкнутые. Тогда функция $f(x) = \max\limits_{i=1,m} f_i(x)$ также является выпуклой и замкнутой. Для любого $x \in \mathrm{int}(\mathrm{dom} f) = \bigcap\limits_{i=1}^{m} \mathrm{int}(\mathrm{dom} f_i)$ имеет место равенство

$$\partial f(x) = \mathrm{conv}\{\partial f_i(x) \mid i \in I(x)\},$$

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

Доказательство

- Первая часть утверждения - это свойство выпуклых функций.
- Докажем формулу для субдифференциала. Рассмотрим $x \in \bigcap\limits_{i=1}^{m} \mathrm{int}(\mathrm{dom}\ f_i)$. Пусть $I(x) = \{1, \dots, k\}$. Тогда $\forall p \in \mathbb{R}^n$
$$f'(x; p) = \max_{i=1,k} f_i'(x; p) = \max_{i=1,k} \max\{\langle g_i, p\rangle \mid g_i \in \partial f_i(x)\}.$$
Заметим, что для любого множества значений $a_1, \dots, a_k$
$$\max_{i=1,k} a_i = \max\left\{\sum_{i=1}^{k} \lambda_i a_i \mid \{\lambda_i\} \in \Delta_k\right\},$$
где $\Delta_k = \{\lambda_i \geq 0,\ \sum\limits_{i=1}^{k} \lambda_i = 1\}$ – k-мерный симплекс.

Доказательство

Поэтому

$$f'(x; p) = \max_{\{\lambda_i\} \in \Delta_k} \left\{\sum_{i=1}^{k} \lambda_i \max\{\langle g_i, p\rangle \mid g_i \in \partial f_i(x)\}\right\} =$$
$$= \max\left\{\left\langle \sum_{i=1}^{k} \lambda_i g_i, p \right\rangle \mid g_i \in \partial f_i(x), \{\lambda_i\} \in \Delta_k\right\} =$$
$$= \max\left\{\langle g, p\rangle \mid g = \sum_{i=1}^{k} \lambda_i g_i,\ g_i \in \partial f_i(x), \{\lambda_i\} \in \Delta_k\right\} =$$
$$= \max\left\{\langle g, p\rangle \mid g \in \mathrm{conv}\left\{\partial f_i(x), i \in I(x)\right\}\right\}.$$

Применяя Теорему 11, получим требуемое.

Примеры

- Рассмотрим функцию $f(x) = \sum_{i=1}^{m} |\langle a_i, x\rangle - b_i|$. Введём обозначения

$$I_-(x) = \{i \mid \langle a_i, x\rangle - b_i < 0\},$$
$$I_+(x) = \{i \mid \langle a_i, x\rangle - b_i > 0\},$$
$$I_0(x) = \{i \mid \langle a_i, x\rangle - b_i = 0\}.$$

Тогда $\partial f(x) = \sum_{i \in I_+(x)} a_i - \sum_{i \in I_-(x)} a_i + \sum_{i \in I_0(x)} [-a_i, a_i]$.

- Рассмотрим функцию $f(x) = \max_{i=\overline{1,n}} x_i$. Пусть $I(x) = \{i \mid x_i = f(x)\}$.

Тогда $\partial f(x) = \mathrm{conv}\{e_i \mid i \in I(x)\}$. Для $x = 0$ получаем

$$\partial f(0) = \mathrm{conv}\{e_i \mid 1 \le i \le n\} \equiv \Delta_n.$$

Классы задач

1. Задача безусловной минимизации

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

2. Задача условной минимизации

$$\min_{x \in \mathrm{X}} f(x)$$

Безусловная оптимизация

Достаточное условие 1 порядка

Пусть $f(x)$ - выпуклая функция, дифференцируемая в точке x^* и $\nabla f(x^*) = 0$.
Тогда x^* – точка глобального минимума $f(x)$ на $\mathbb{R}^n$.

Необходимое условие 2 порядка

Пусть x^* точка минимума $f(x)$ на $\mathbb{R}^n$, $f(x)$ дважды дифференцируема в точке x^*. Тогда $\nabla^2 f(x^*) \geq 0$.

Достаточное условие 2 порядка

Пусть в точке x^* функция $f(x^*)$ дважды дифференцируема, выполнено необходимое условие 1 порядка (т.е. $\nabla f(x^*) = 0$) и $\nabla^2 f(x^*) > 0$. Тогда x^* – точка локального минимума.

Условная оптимизация

Теоерема

Пусть X выпукло, функция $f(x)$ дифференцируема в точке $x^* \in \mathrm{X}$. Тогда

1. если x^* - локальный минимум, то $\langle \nabla f(x^*),\ x - x^* \rangle \geq 0 \quad \forall x \in \mathrm{X}$;
2. если $f(x)$ выпукла на X и выполняется неравенство, то x^* - глобальный минимум.

Геометрический смысл

$$\langle \nabla f(x^*),\ x - x^* \rangle \geq 0 \quad \forall x \in \mathrm{X}$$

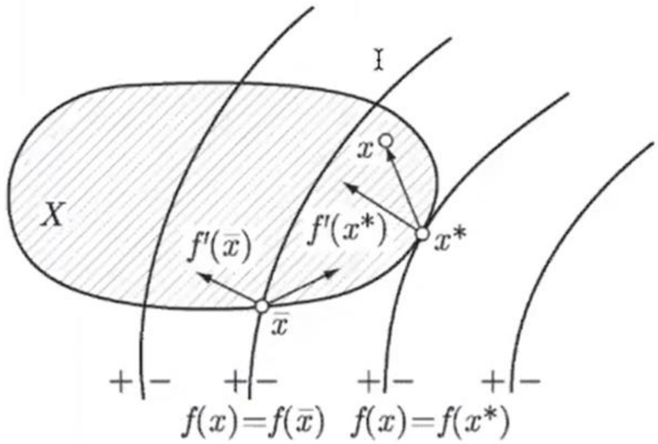


Figure: необходимое условие локальной оптимальности в условной задаче минимизации

Условная оптимизация

Теорема

Равенство $f(x^*) = \min\limits_{x \in \mathrm{dom}\ f} f(x)$ имеет место тогда и только тогда, когда

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

Доказательство

Действительно, если $0 \in \partial f(x^*)$, то $f(x) \geq f(x^*) + \langle 0, x - x^* \rangle = f(x^*)$ для всех $x \in \mathrm{dom}\ f$.
С другой стороны, если $f(x) \geq f(x^*)$ при всех $x \in \mathrm{dom}\ f$, то $0 \in \mathrm{dom}\ f(x^*)$, как следует из определения субградиента.