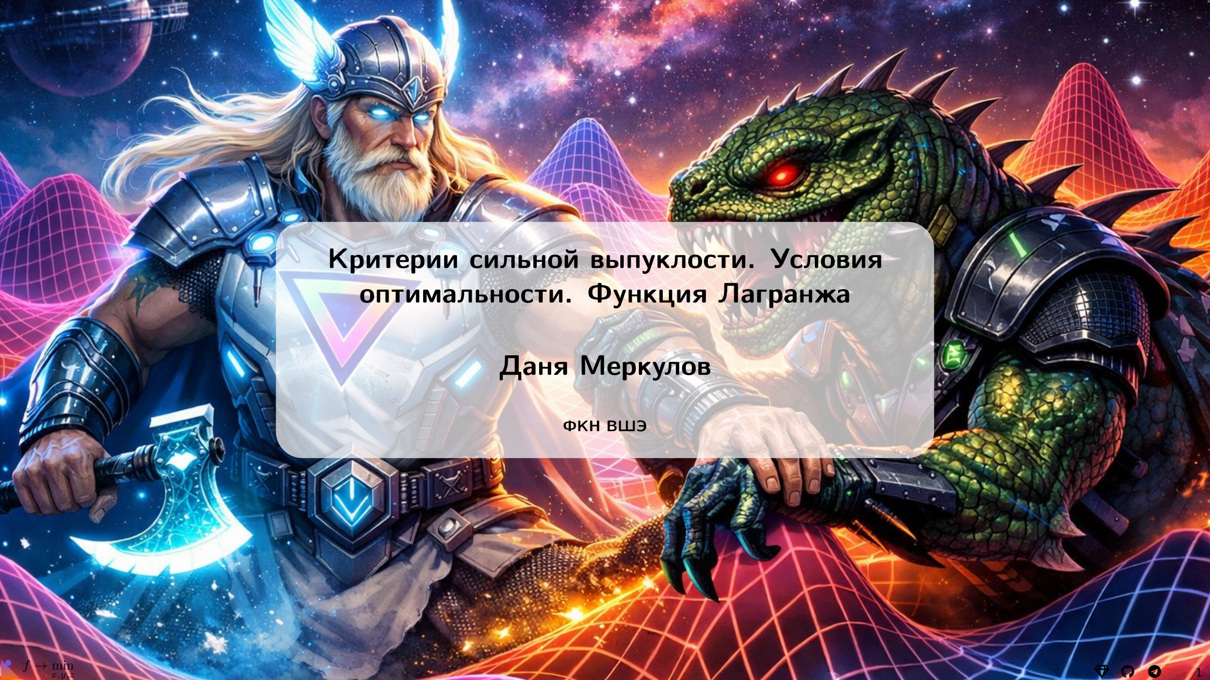


Критерии сильной выпуклости. Условия оптимальности. Функция Лагранжа

Даня Меркулов

ФКН ВШЭ

Критерии сильной выпуклости

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

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$$

$$\lambda_{min}(A) = 1$$

$$\lambda_{max}(A) = 5$$

$$\mu = \lambda_{min}(A) = 1$$

$$L = \lambda_{max}(A) = 5$$

константа сильной вып. $f(x)$

константа гладкости $f(x)$

$\parallel$

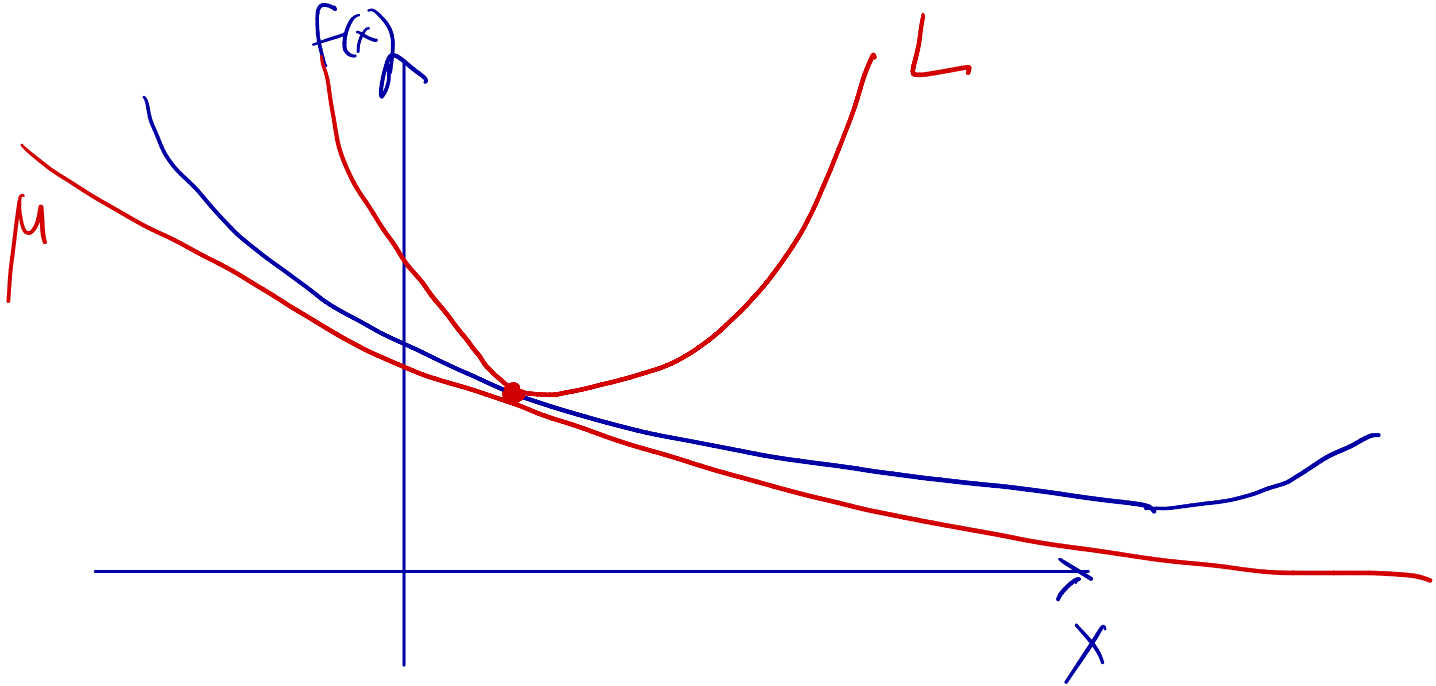
константа Липшица градиента

$$\forall x, y \in \mathbb{R}^n \quad \|\nabla f(x) - \nabla f(y)\| \le L \|x - y\|$$

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

$$\|\nabla^2 f(x)\| \le L$$

$$\sigma_{max}(\nabla^2 f(x)) \le L$$



f(x)
L
μ
X

Дифференциальный критерий выпуклости первого порядка

Дифференцируемая функция $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ выпукла тогда и только тогда, когда $\forall x, y \in S$:

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

Пусть $y = x + \Delta x$, тогда критерий запишется в виде:

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

$$f(x+\Delta x) \geq f^{I}_{x}(x+\Delta x)$$

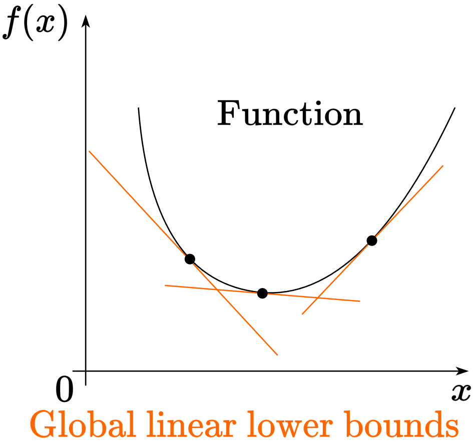


Figure 1: Выпуклая функция больше или равна линейной аппроксимации Тейлора в любой точке

Дифференциальный критерий выпуклости второго порядка

Дважды дифференцируемая функция $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ выпукла тогда и только тогда, когда $\forall x \in \mathbf{int}(S) \neq \emptyset$:

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

Другими словами, $\forall y \in \mathbb{R}^n$:

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

Сильная выпуклость

$f(x)$, **определенная на выпуклом множестве** $S \subseteq \mathbb{R}^n$, называется μ-сильно выпуклой (сильно выпуклой) на S, если:

опр.

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

для любых $x_1, x_2 \in S$ и $0 \leq \lambda \leq 1$ для некоторого $\mu > 0$.

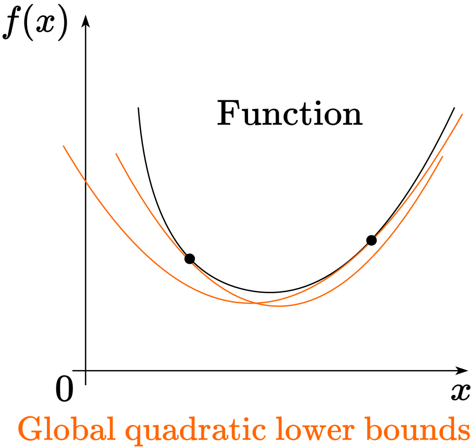


Figure 2: Сильно выпуклая функция не меньше некоторой параболы в любой точке

Дифференциальный критерий сильной выпуклости первого порядка

Дифференцируемая $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ является μ-сильно выпуклой тогда и только тогда, когда $\forall x, y \in S$:

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

Дифференциальный критерий сильной выпуклости первого порядка

Дифференцируемая $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ является μ-сильно выпуклой тогда и только тогда, когда $\forall x, y \in S$:

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

Пусть $y = x + \Delta x$, тогда критерий запишется в виде:

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

Дифференциальный критерий сильной выпуклости первого порядка

Дифференцируемая $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ является μ-сильно выпуклой тогда и только тогда, когда $\forall x, y \in S$:

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

Пусть $y = x + \Delta x$, тогда критерий запишется в виде:

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

Theorem

Пусть $f(x)$ дифференцируемая функция на выпуклом множестве $X \subseteq \mathbb{R}^n$. Тогда $f(x)$ сильно выпукла на X с константой $\mu > 0$ тогда и только тогда, когда

$$\boxed{f(x) - f(x_0) \geq \langle \nabla f(x_0), x - x_0 \rangle + \frac{\mu}{2}\|x - x_0\|^2}$$

для всех $x, x_0 \in X$.

Доказательство. Необходимость

$f(x)$ - μ-сильн. вып $\Rightarrow$ крит. 1 пор.

Пусть $0 < \lambda \leq 1$. Согласно определению сильно выпуклой функции,

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

определение

Доказательство. Необходимость

Пусть $0 < \lambda \leq 1$. Согласно определению сильно выпуклой функции,

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

$1 : \lambda \neq 0$

или эквивалентно,

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

Доказательство. Необходимость

Пусть $0 < \lambda \leq 1$. Согласно определению сильно выпуклой функции,

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

или эквивалентно,

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

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

дифф. f (форм. Тейлор с ост. чл. в форме Пеано)

Доказательство. Необходимость

Пусть $0 < \lambda \leq 1$. Согласно определению сильно выпуклой функции,

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

или эквивалентно,

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

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

$$= \langle\nabla f(x_0), x - x_0\rangle + \cancel{\frac{o(\lambda)}{\lambda}} \to 0.$$

Таким образом, переходя к пределу при $\lambda \downarrow 0$, мы приходим к исходному утверждению.

Доказательство. Достаточность

Предположим, что неравенство в теореме выполняется для всех $x, x_0 \in X$. Возьмем $x_0 = \lambda x_1 + (1-\lambda)x_2$, где $x_1, x_2 \in X$, $0 \leq \lambda \leq 1$. Согласно неравенству, следующие неравенства выполняются:

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

Доказательство. Достаточность

Предположим, что неравенство в теореме выполняется для всех $x, x_0 \in X$. Возьмем $\underline{x_0 = \lambda x_1 + (1-\lambda)x_2}$, где $x_1, x_2 \in X$, $0 \leq \lambda \leq 1$. Согласно неравенству, следующие неравенства выполняются:

$$
\begin{aligned}
x_1, x_0 \qquad & f(x_1) - f(x_0) \geq \langle \nabla f(x_0), x_1 - x_0 \rangle + \frac{\mu}{2}\|x_1 - x_0\|^2, \\
x_2, x_0 \qquad & f(x_2) - f(x_0) \geq \langle \nabla f(x_0), x_2 - x_0 \rangle + \frac{\mu}{2}\|x_2 - x_0\|^2.
\end{aligned}
$$

Доказательство. Достаточность

Предположим, что неравенство в теореме выполняется для всех $x, x_0 \in X$. Возьмем $\boxed{x_0 = \lambda x_1 + (1-\lambda)x_2}$, где $x_1, x_2 \in X$, $\boxed{0 \leq \lambda \leq 1}$. Согласно неравенству, следующие неравенства выполняются:

$$f(x_1) - f(x_0) \geq \langle \nabla f(x_0), x_1 - x_0 \rangle + \frac{\mu}{2}\|x_1 - x_0\|^2, \qquad \lambda$$

$$f(x_2) - f(x_0) \geq \langle \nabla f(x_0), x_2 - x_0 \rangle + \frac{\mu}{2}\|x_2 - x_0\|^2. \qquad (1-\lambda)$$

Умножая первое неравенство на λ и второе на $1-\lambda$ и складывая их, учитывая, что

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

Доказательство. Достаточность

Предположим, что неравенство в теореме выполняется для всех $x, x_0 \in X$. Возьмем $x_0 = \lambda x_1 + (1-\lambda)x_2$, где $x_1, x_2 \in X$, $0 \leq \lambda \leq 1$. Согласно неравенству, следующие неравенства выполняются:

$$
f(x_1) - f(x_0) \geq \langle \nabla f(x_0), x_1 - x_0 \rangle + \frac{\mu}{2}\|x_1 - x_0\|^2,
$$

$$
f(x_2) - f(x_0) \geq \langle \nabla f(x_0), x_2 - x_0 \rangle + \frac{\mu}{2}\|x_2 - x_0\|^2.
$$

Теорема Васила

$\lambda + 1 - \lambda = 1$

Умножая первое неравенство на λ и второе на $1-\lambda$ и складывая их, учитывая, что

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

и $\lambda(1-\lambda)^2 + \lambda^2(1-\lambda) = \lambda(1-\lambda)$, мы получаем

$$
\begin{aligned}
\lambda f(x_1) + (1-\lambda)f(x_2) - f(x_0) - \frac{\mu}{2}\lambda(1-\lambda)\|x_1 - x_2\|^2 \geq \\
\langle \nabla f(x_0), \underbrace{\lambda x_1 + (1-\lambda)x_2}_{x_0} - x_0 \rangle = 0.
\end{aligned}
$$

Таким образом, неравенство из определения сильно выпуклой функции выполняется. Важно отметить, что $\mu = 0$ соответствует случаю выпуклой функции и соответствующему дифференциальному критерию.

Дифференциальный критерий сильной выпуклости второго порядка

Дважды дифференцируемая функция $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ называется μ-сильно выпуклой тогда и только тогда, когда $\forall x \in \mathbf{int}(S) \neq \emptyset$:

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

Другими словами:

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

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

Дифференциальный критерий сильной выпуклости второго порядка

Дважды дифференцируемая функция $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ называется μ-сильно выпуклой тогда и только тогда, когда $\forall x \in \mathbf{int}(S) \neq \emptyset$:

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

Другими словами:

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

Theorem

Пусть $X \subseteq \mathbb{R}^n$ выпуклое множество, с $\text{int}X \neq \emptyset$. Кроме того, пусть $f(x)$ дважды непрерывно дифференцируемая функция на X. Тогда $f(x)$ сильно выпукла на X с константой $\mu > 0$ тогда и только тогда, когда

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

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

для всех $x \in X$ и $y \in \mathbb{R}^n$.

Доказательство. Необходимость

Целевое неравенство тривиально, когда $y = \mathbf{0}_n$, поэтому мы предполагаем $y \neq \mathbf{0}_n$.

Предположим, что x является внутренней точкой множества X. Тогда $\boxed{x + \alpha y \in X}$ для всех $y \in \mathbb{R}^n$ и достаточно малых α. Поскольку $f(x)$ дважды дифференцируема,

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

Доказательство. Необходимость

Целевое неравенство тривиально, когда $y = \mathbf{0}_n$, поэтому мы предполагаем $y \neq \mathbf{0}_n$.

Предположим, что x является внутренней точкой множества X. Тогда $x + \alpha y \in X$ для всех $y \in \mathbb{R}^n$ и достаточно малых α. Поскольку $f(x)$ дважды дифференцируема,

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

На основании первого дифференциального критерия сильной выпуклости:

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

$|:\alpha^2 \neq 0$

$\alpha \to 0$

Это неравенство сводится к целевому неравенству после деления обеих сторон на α^2 и перехода к пределу при $\alpha \downarrow 0$.

Если $x \in X$ но $x \notin \text{int}X$, рассмотрим последовательность $\{x_k\}$ такую, что $x_k \in \text{int}X$ и $x_k \to x$ при $k \to \infty$. Тогда, мы приходим к целевому неравенству после перехода к пределу.

Доказательство. Достаточность

$O(?)$

Используя формулу Тейлора с остаточным членом Лагранжа и целевое неравенство, мы получаем для $x + y \in X$:

$$f(x+y) - f(x) - \langle \nabla f(x), y \rangle = \frac{1}{2} \langle y, \nabla^2 f(x + \alpha y) y \rangle \geq \frac{\mu}{2} \|y\|^2,$$

где $0 \leq \alpha \leq 1$. Следовательно,

Доказательство. Достаточность

$\exists \alpha$

Используя формулу Тейлора с остаточным членом Лагранжа и целевое неравенство, мы получаем для $x + y \in X$:

$$
f(x+y) - f(x) - \langle \nabla f(x), y \rangle = \frac{1}{2} \langle y, \nabla^2 f(x + \alpha y) y \rangle \geq \frac{\mu}{2} \|y\|^2,
$$

где $0 \leq \alpha \leq 1$. Следовательно,

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

с.в.I

Следовательно, согласно первому дифференциальному критерию сильной выпуклости, функция $f(x)$ сильно выпукла с константой μ. Важно отметить, что $\mu = 0$ соответствует случаю выпуклой функции и соответствующему дифференциальному критерию.

Выпуклая и вогнутая функция

Example

Покажите, что $f(x) = c^\top x + b$ выпукла и вогнута.

Простейшая сильно выпуклая функция

$A \succ 0 \Rightarrow f(x) - \mu$ - сильно выпукл.

Example

Покажите, что $\boxed{f(x) = x^\top Ax, \text{ где } A \succeq 0}$ - выпукла на $\mathbb{R}^n$. Является ли она сильно выпуклой?

Выпуклость и непрерывность

Пусть $f(x)$ - выпуклая функция на выпуклом множестве $S \subseteq \mathbb{R}^n$. Тогда $f(x)$ непрерывна $\forall x \in \mathbf{ri}(S)$.

Собственная выпуклая функция

Функция $f : \mathbb{R}^n \to \mathbb{R}$ называется **собственной выпуклой функцией**, если она никогда не принимает значения $-\infty$ и не равна ∞ тождественно.

Индикаторная функция

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

является собственной выпуклой функцией.

Выпуклость и непрерывность

Пусть $f(x)$ - выпуклая функция на выпуклом множестве $S \subseteq \mathbb{R}^n$. Тогда $f(x)$ непрерывна $\forall x \in \textbf{ri}(S)$.

Собственная выпуклая функция

Функция $f : \mathbb{R}^n \to \mathbb{R}$ называется **собственной выпуклой функцией**, если она никогда не принимает значения $-\infty$ и не равна ∞ тождественно.

Индикаторная функция

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

является собственной выпуклой функцией.

Замкнутая функция

Функция $f : \mathbb{R}^n \to \mathbb{R}$ называется **замкнутой**, если для каждого $\alpha \in \mathbb{R}$, множество подуровня замкнуто. Эквивалентно, если надграфик замкнут, то функция f замкнута.

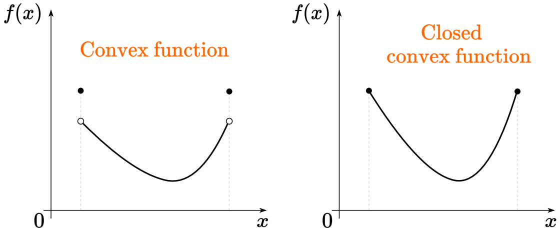


Figure 3: Выпуклые функции могут иметь разрывы на границе своей области определения.

Условие Поляка-Лоясиевича. Линейная сходимость градиентного спуска без выпуклости

градиентного доминирования

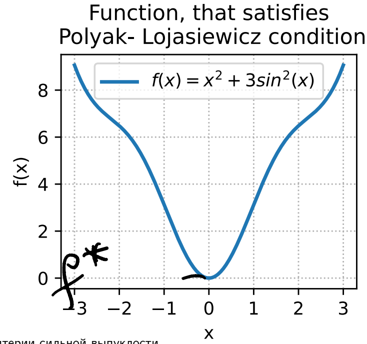
Неравенство PL выполняется, если выполняется следующее условие для некоторого $\mu > 0$,

$$
\boxed{\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*) \forall x} \quad (PL) \qquad f^* - \min_{x \in S} f(x)
$$

При выполнении условия PL алгоритм градиентного спуска имеет линейную сходимость.

Следующие функции удовлетворяют условию PL, но не являются выпуклыми. Ссылка на код

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



Условие Поляка-Лоясиевича. Линейная сходимость градиентного спуска без выпуклости

Неравенство PL выполняется, если выполняется следующее условие для некоторого $\mu > 0$,

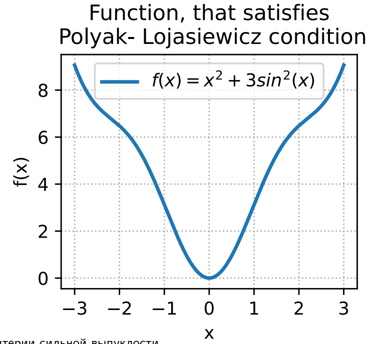
$$\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*) \forall x$$

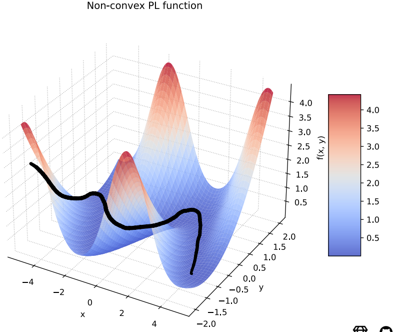
При выполнении условия PL алгоритм градиентного спуска имеет линейную сходимость.

Следующие функции удовлетворяют условию PL, но не являются выпуклыми. Ссылка на код

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

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





Условия оптимальности

Теория

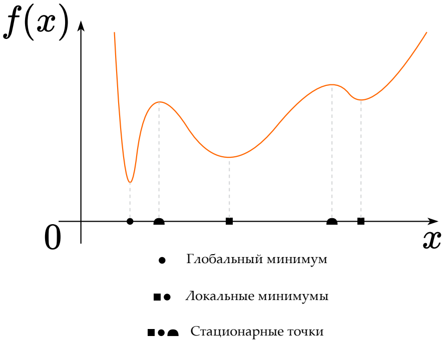


Figure 7: Иллюстрация различных стационарных (критических) точек

мин.

функция

$$\boxed{f(x) \to \min_{x \in S}}$$

бюджетное множество

feasible set

Теория

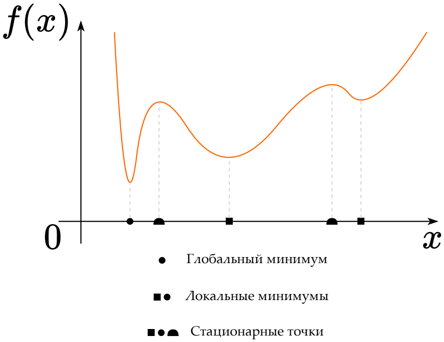


Figure 7: Иллюстрация различных стационарных (критических) точек

$$f(x) \to \min_{x \in S}$$

Множество S обычно называется **допустимым множеством** (или **бюджетным множеством**).

Теория

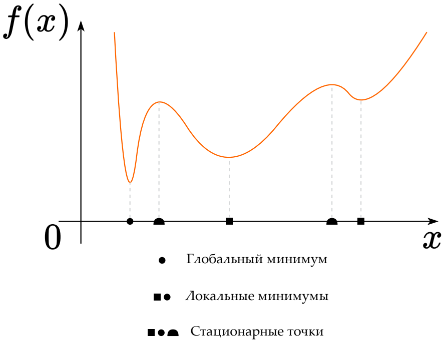


Figure 7: Иллюстрация различных стационарных (критических) точек

$$f(x) \to \min_{x \in S}$$

Множество S обычно называется **допустимым множеством** (или **бюджетным множеством**).

Мы говорим, что задача имеет решение, если допустимое множество **не пусто** и существует точка $x^* \in S$, в которой достигается минимум или инфимум данной функции.

Теория

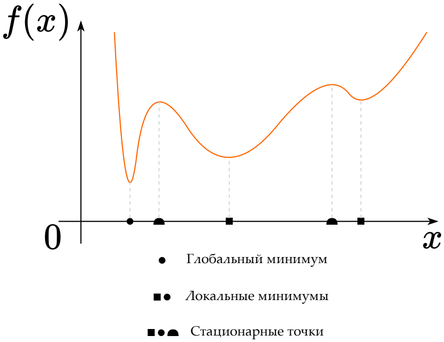


Figure 7: Иллюстрация различных стационарных (критических) точек

$$f(x) \to \min_{x \in S}$$

Множество S обычно называется **допустимым множеством** (или **бюджетным множеством**).

Мы говорим, что задача имеет решение, если допустимое множество **не пусто** и существует точка $x^* \in S$, в которой достигается минимум или инфимум данной функции.

- Точка x^* является **глобальным минимумом**, если $f(x^*) \leq f(x)$ для всех $x \in S$.

Теория

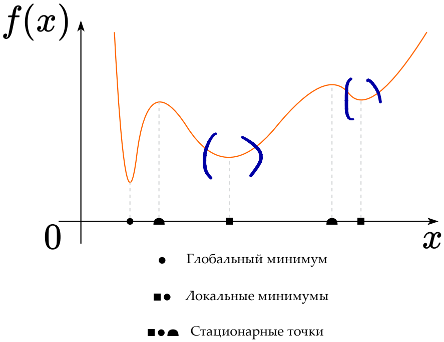


Figure 7: Иллюстрация различных стационарных (критических) точек

$$f(x) \to \min_{x \in S}$$

Множество S обычно называется **допустимым множеством** (или **бюджетным множеством**).

Мы говорим, что задача имеет решение, если допустимое множество **не пусто** и существует точка $x^* \in S$, в которой достигается минимум или инфимум данной функции.

- Точка x^* является **глобальным минимумом**, если $f(x^*) \leq f(x)$ для всех $x \in S$.
- Точка x^* является **локальным минимумом**, если существует окрестность N точки x^* такая, что $f(x^*) \leq f(x)$ для всех $x \in N \cap S$.

Теория

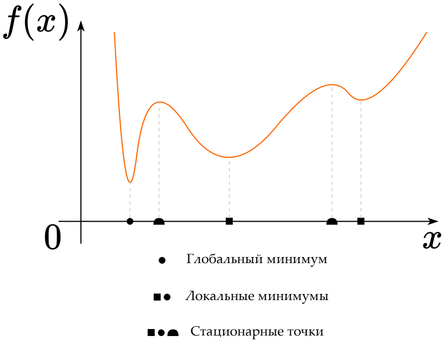


Figure 7: Иллюстрация различных стационарных (критических) точек

$$f(x) \to \min_{x \in S}$$

Множество S обычно называется **допустимым множеством** (или **бюджетным множеством**).

Мы говорим, что задача имеет решение, если допустимое множество **не пусто** и существует точка $x^* \in S$, в которой достигается минимум или инфимум данной функции.

- Точка x^* является **глобальным минимумом**, если $f(x^*) \leq f(x)$ для всех $x \in S$.
- Точка x^* является **локальным минимумом**, если существует окрестность N точки x^* такая, что $f(x^*) \leq f(x)$ для всех $x \in N \cap S$.
- Точка x^* является **строгим локальным минимумом**, если существует окрестность N точки x^* такая, что $f(x^*) < f(x)$ для всех $x \in N \cap S$ с $x \neq x^*$.

Теория

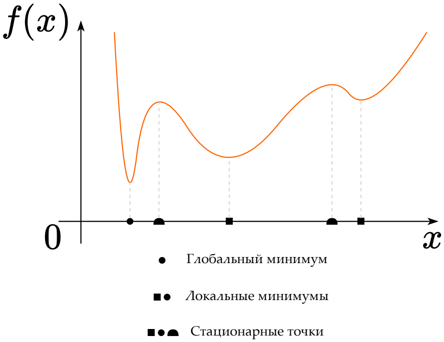


Figure 7: Иллюстрация различных стационарных (критических) точек

$$f(x) \to \min_{x \in S}$$

Множество S обычно называется **допустимым множеством** (или **бюджетным множеством**).

Мы говорим, что задача имеет решение, если допустимое множество **не пусто** и существует точка $x^* \in S$, в которой достигается минимум или инфимум данной функции.

- Точка x^* является **глобальным минимумом**, если $f(x^*) \leq f(x)$ для всех $x \in S$.
- Точка x^* является **локальным минимумом**, если существует окрестность N точки x^* такая, что $f(x^*) \leq f(x)$ для всех $x \in N \cap S$.
- Точка x^* является **строгим локальным минимумом**, если существует окрестность N точки x^* такая, что $f(x^*) < f(x)$ для всех $x \in N \cap S$ с $x \neq x^*$.
- Мы называем точку x^* **стационарной точкой** (или критической точкой), если $\nabla f(x^*) = 0$. Любой локальный минимум дифференцируемой функции должен быть стационарной точкой.

Теорема Вейерштрасса об экстремальных значениях

Theorem

Пусть $S \subset \mathbb{R}^n$ - компактное множество и $f(x)$ - непрерывная функция на S. Тогда точка глобального минимума функции $f(x)$ на S существует.

Теорема Вейерштрасса об экстремальных значениях

Theorem

Пусть $S \subset \mathbb{R}^n$ - компактное множество и $f(x)$ - непрерывная функция на S. Тогда точка глобального минимума функции $f(x)$ на S существует.

Figure 8: Многие практические задачи теоретически разрешимы

Теорема Вейерштрасса об экстремальных значениях

Theorem

Пусть $S \subset \mathbb{R}^n$ - компактное множество и $f(x)$ - непрерывная функция на S. Тогда точка глобального минимума функции $f(x)$ на S существует.

Figure 8: Многие практические задачи теоретически разрешимы

Теорема Тейлора

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ - непрерывно дифференцируемая функция и $p \in \mathbb{R}^n$. Тогда мы имеем:

$$f(x+p) = f(x) + \nabla f(x+tp)^T p \quad \text{для некоторого } t \in (0,1)$$

Теорема Вейерштрасса об экстремальных значениях

Theorem

Пусть $S \subset \mathbb{R}^n$ - компактное множество и $f(x)$ - непрерывная функция на S. Тогда точка глобального минимума функции $f(x)$ на S существует.

Figure 8: Многие практические задачи теоретически разрешимы

Теорема Тейлора

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ - непрерывно дифференцируемая функция и $p \in \mathbb{R}^n$. Тогда мы имеем:

$$f(x+p) = f(x) + \nabla f(x+tp)^T p \quad \text{для некоторого } t \in (0,1)$$

Кроме того, если f дважды непрерывно дифференцируема, то мы имеем:

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)p\, dt$$

$$\boxed{f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp) p}$$

для некоторого $t \in (0,1)$.

Безусловная оптимизация

$$S = \mathbb{R}^n$$

Необходимые условия

Необходимое условие оптимальности первого порядка

Если x^* - локальный минимум и f непрерывно дифференцируема в открытой окрестности, то

$$\nabla f(x^*) = 0$$

Необходимые условия

Необходимое условие оптимальности первого порядка

Если x^* - локальный минимум и f непрерывно дифференцируема в открытой окрестности, то

$$\nabla f(x^*) = 0$$

Доказательство
Предположим от противного, что $\boxed{\nabla f(x^*) \neq 0}$. Определим вектор $\boxed{p = -\nabla f(x^*)}$ и заметим, что

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$$

Необходимые условия

Необходимое условие оптимальности первого порядка

Если x^* - локальный минимум и f непрерывно дифференцируема в открытой окрестности, то

$$\nabla f(x^*) = 0$$

Доказательство
Предположим от противного, что $\nabla f(x^*) \neq 0$. Определим вектор $p = -\nabla f(x^*)$ и заметим, что

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$$

Поскольку $\boxed{\nabla f \text{ непрерывна в окрестности } x^*}$, существует скаляр $T > 0$ такой, что

$$\boxed{p^T \nabla f(x^* + tp) < 0,} \text{ для всех } \ t \in [0, T]$$

Необходимые условия

Необходимое условие оптимальности первого порядка

Если x^* - локальный минимум и f непрерывно дифференцируема в открытой окрестности, то

$$\nabla f(x^*) = 0$$

Доказательство
Предположим от противного, что $\nabla f(x^*) \neq 0$. Определим вектор $p = -\nabla f(x^*)$ и заметим, что

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$$

Поскольку ∇f непрерывна в окрестности x^*, существует скаляр $T > 0$ такой, что

$$p^T \nabla f(x^* + tp) < 0, \text{ для всех } t \in [0, T]$$

Для любого $\bar{t} \in (0, T]$, мы имеем по теореме Тейлора, что

$$\boxed{f(x^* + \bar{t}p) = f(x^*) + \bar{t}\, p^T \nabla f(x^* + tp),} \text{ для некоторого } t \in (0, \bar{t})$$

Необходимые условия

Необходимое условие оптимальности первого порядка

Если x^* - локальный минимум и f непрерывно дифференцируема в открытой окрестности, то

$$\nabla f(x^*) = 0$$

Доказательство
Предположим от противного, что $\nabla f(x^*) \neq 0$. Определим вектор $p = -\nabla f(x^*)$ и заметим, что

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$$

Поскольку ∇f непрерывна в окрестности x^*, существует скаляр $T > 0$ такой, что

$$p^T \nabla f(x^* + tp) < 0, \text{ для всех } t \in [0, T]$$

$f(x^*)$ — неминимум

Для любого $\bar{t} \in (0, T]$, мы имеем по теореме Тейлора, что

< 0

$$f(x^* + \bar{t}p) = f(x^*) + \bar{t}\, p^T \nabla f(x^* + tp), \text{ для некоторого } t \in (0, \bar{t})$$

Следовательно, $f(x^* + \bar{t}p) < f(x^*)$ для всех $\bar{t} \in (0, T]$. Мы нашли направление из x^*, вдоль которого f убывает, поэтому x^* не является локальным минимумом, что приводит к противоречию.

Достаточные условия

Достаточные условия оптимальности второго порядка

Пусть $\nabla^2 f$ непрерывна в открытой окрестности x^*, и выполнено

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0.$$

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

Тогда x^* является строгим локальным минимумом функции f.

Достаточные условия

Достаточные условия оптимальности второго порядка

Пусть $\nabla^2 f$ непрерывна в открытой окрестности x^*, и выполнено

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0.$$

Тогда x^* является строгим локальным минимумом функции f.

Доказательство

Поскольку гессиан непрерывен и положительно определён в x^*, мы можем выбрать радиус $r > 0$ такой, что $\nabla^2 f(x)$ остаётся положительно определённым для всех x в открытом шаре $B = \{z \mid \|z - x^*\| < r\}$. Возьмём любой ненулевой вектор p с $\|p\| < r$, тогда $x^* + p \in B$ и для некоторого $t \in (0, 1)$ выполняется

Достаточные условия

Достаточные условия оптимальности второго порядка

Пусть $\nabla^2 f$ непрерывна в открытой окрестности x^*, и выполнено

$$
\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0.
$$

Тогда x^* является строгим локальным минимумом функции f.

Доказательство

Поскольку гессиан непрерывен и положительно определён в x^*, мы можем выбрать радиус $r > 0$ такой, что $\nabla^2 f(x)$ остаётся положительно определённым для всех x в открытом шаре $B = \{z \mid \|z - x^*\| < r\}$. Возьмём любой ненулевой вектор p с $\|p\| < r$, тогда $x^* + p \in B$ и для некоторого $t \in (0, 1)$ выполняется

$$
\boxed{f(x^* + p) = f(x^*) + \underbrace{p^T \nabla f(x^*)}_{=0} + \underbrace{\frac{1}{2} p^T \nabla^2 f(x^* + tp) p}_{>0},}
$$

$$
f(x^* + p) > f(x^*)
$$

Достаточные условия

ℹ Достаточные условия оптимальности второго порядка

Пусть $\nabla^2 f$ непрерывна в открытой окрестности x^*, и выполнено

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0.$$

Тогда x^* является строгим локальным минимумом функции f.

Доказательство

Поскольку гессиан непрерывен и положительно определён в x^*, мы можем выбрать радиус $r > 0$ такой, что $\nabla^2 f(x)$ остаётся положительно определённым для всех x в открытом шаре $B = \{z \mid \|z - x^*\| < r\}$. Возьмём любой ненулевой вектор p с $\|p\| < r$, тогда $x^* + p \in B$ и для некоторого $t \in (0, 1)$ выполняется

$$f(x^* + p) = f(x^*) + p^T \nabla f(x^*) + \frac{1}{2} p^T \nabla^2 f(x^* + tp) p,$$

$$= f(x^*) + \frac{1}{2} p^T \nabla^2 f(x^* + tp) p.$$

Достаточные условия

Достаточные условия оптимальности второго порядка

Пусть $\nabla^2 f$ непрерывна в открытой окрестности x^*, и выполнено

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0.$$

Тогда x^* является строгим локальным минимумом функции f.

Доказательство

Поскольку гессиан непрерывен и положительно определён в x^*, мы можем выбрать радиус $r > 0$ такой, что $\nabla^2 f(x)$ остаётся положительно определённым для всех x в открытом шаре $B = \{z \mid \|z - x^*\| < r\}$. Возьмём любой ненулевой вектор p с $\|p\| < r$, тогда $x^* + p \in B$ и для некоторого $t \in (0, 1)$ выполняется

$$f(x^* + p) = f(x^*) + p^T \nabla f(x^*) + \frac{1}{2} p^T \nabla^2 f(x^* + tp) p,$$

$$= f(x^*) + \frac{1}{2} p^T \nabla^2 f(x^* + tp) p.$$

Поскольку $x^* + tp \in B$, то $p^T \nabla^2 f(x^* + tp) p > 0$, и поэтому $f(x^* + p) > f(x^*)$, что доказывает утверждение.

Контрпример Пеано

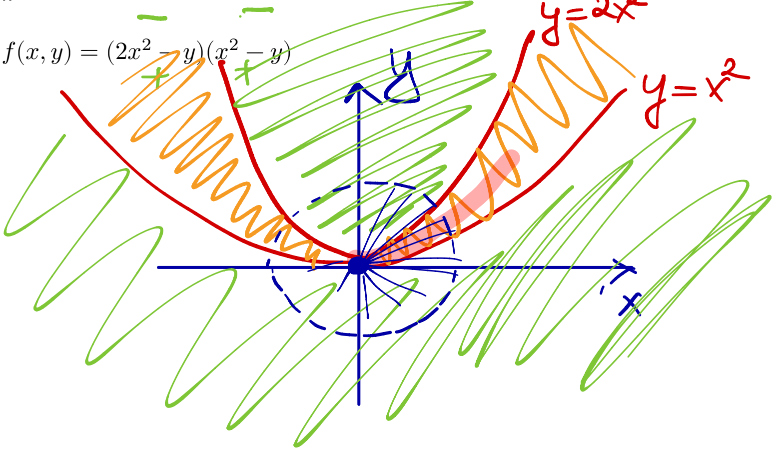
Заметим, что если $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succeq 0$ (гессиан положительно полуопределён), то мы не можем быть уверены, что x^* является локальным минимумом.

Контрпример Пеано

Заметим, что если $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succeq 0$ (гессиан положительно полуопределён), то мы не можем быть уверены, что x^* является локальным минимумом.

почему $\cancel{\nabla^2 f \succeq 0}$

$$
f(x,y) = (2x^2 - y)(x^2 - y)
$$



Контрпример Пеано

Заметим, что если $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succeq 0$ (гессиан положительно полуопределён), то мы не можем быть уверены, что x^* является локальным минимумом.

$$f(x, y) = (2x^2 - y)(x^2 - y)$$

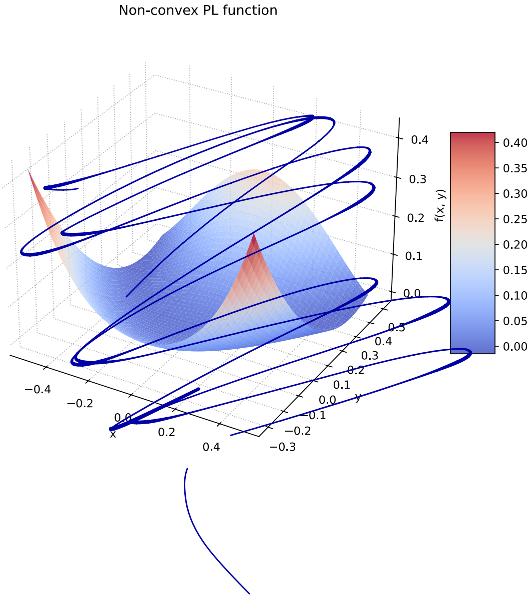
Хотя поверхность не имеет локального минимума в начале координат, её пересечение с любой вертикальной плоскостью, проходящей через начало координат (плоскость с уравнением $y = mx$ или $x = 0$) является кривой, которая имеет локальный минимум в начале координат. Другими словами, если точка начинает движение из начала координат $(0, 0)$ вдоль любой прямой линии, то значение $(2x^2 - y)(x^2 - y)$ будет увеличиваться в начале движения. Тем не менее, $(0, 0)$ не является локальным минимумом функции, потому что движение вдоль параболы, такой как $y = \sqrt{2}x^2$, приведёт к уменьшению значения функции.

Контрпример Пеано

Заметим, что если $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succeq 0$ (гессиан положительно полуопределён), то мы не можем быть уверены, что x^* является локальным минимумом.

$$f(x,y) = (2x^2 - y)(x^2 - y)$$

Хотя поверхность не имеет локального минимума в начале координат, её пересечение с любой вертикальной плоскостью, проходящей через начало координат (плоскость с уравнением $y = mx$ или $x = 0$) является кривой, которая имеет локальный минимум в начале координат. Другими словами, если точка начинает движение из начала координат $(0,0)$ вдоль любой прямой линии, то значение $(2x^2 - y)(x^2 - y)$ будет увеличиваться в начале движения. Тем не менее, $(0,0)$ не является локальным минимумом функции, потому что движение вдоль параболы, такой как $y = \sqrt{2}x^2$, приведёт к уменьшению значения функции.



Условная оптимизация

$$S \neq \mathbb{R}^n$$

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.

Рассмотрим множество $S \subseteq \mathbb{R}^n$ и функцию $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума f на S, и что f непрерывно дифференцируема в окрестности x^*.

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.

Рассмотрим множество $S \subseteq \mathbb{R}^n$ и функцию $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума f на S, и что f непрерывно дифференцируема в окрестности x^*.

1. Тогда для любого допустимого направления $d \in \mathbb{R}^n$ в x^* выполняется $\nabla f(x^*)^\top d \geq 0$.

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.

Рассмотрим множество $S \subseteq \mathbb{R}^n$ и функцию $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума f на S, и что f непрерывно дифференцируема в окрестности x^*.

1. Тогда для любого допустимого направления $d \in \mathbb{R}^n$ в x^* выполняется $\nabla f(x^*)^\top d \geq 0$.
2. Если, кроме того, S выпукло, то

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

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.

Рассмотрим множество $S \subseteq \mathbb{R}^n$ и функцию $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума f на S, и что f непрерывно дифференцируема в окрестности x^*.

1. Тогда для любого допустимого направления $d \in \mathbb{R}^n$ в x^* выполняется $\nabla f(x^*)^\top d \geq 0$.
2. Если, кроме того, S выпукло, то

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

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.

Рассмотрим множество $S \subseteq \mathbb{R}^n$ и функцию $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума f на S, и что f непрерывно дифференцируема в окрестности x^*.

1. Тогда для любого допустимого направления $d \in \mathbb{R}^n$ в x^* выполняется $\nabla f(x^*)^\top d \geq 0$.
2. Если, кроме того, S выпукло, то

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

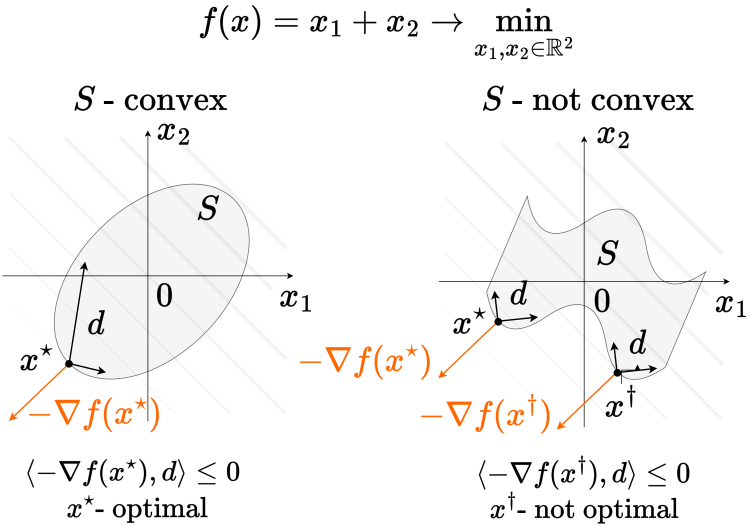


Figure 9: Общее условие локальной оптимальности первого порядка

Выпуклый случай

Следует отметить, что в **выпуклом** случае (то есть при выпуклых f и S) необходимое условие становится достаточным.

Выпуклый случай

Следует отметить, что в **выпуклом** случае (то есть при выпуклых f и S) необходимое условие становится достаточным.

Ещё один важный результат для выпуклого случая звучит следующим образом: если $f(x) : S \to \mathbb{R}$ - выпуклая функция, определённая на выпуклом множестве S, то:

Выпуклый случай

Следует отметить, что в **выпуклом** случае (то есть при выпуклых f и S) необходимое условие становится достаточным.

Ещё один важный результат для выпуклого случая звучит следующим образом: если $f(x) : S \to \mathbb{R}$ - выпуклая функция, определённая на выпуклом множестве S, то:

- Любой локальный минимум является глобальным.

Выпуклый случай

Следует отметить, что в **выпуклом** случае (то есть при выпуклых f и S) необходимое условие становится достаточным.

Ещё один важный результат для выпуклого случая звучит следующим образом: если $f(x) : S \to \mathbb{R}$ - выпуклая функция, определённая на выпуклом множестве S, то:

- Любой локальный минимум является глобальным.
- Множество локальных (= глобальных) минимумов S^* выпукло.

Выпуклый случай

Следует отметить, что в **выпуклом** случае (то есть при выпуклых f и S) необходимое условие становится достаточным.

Ещё один важный результат для выпуклого случая звучит следующим образом: если $f(x) : S \to \mathbb{R}$ - выпуклая функция, определённая на выпуклом множестве S, то:

- Любой локальный минимум является глобальным.
- Множество локальных (= глобальных) минимумов S^* выпукло.
- Если $f(x)$ - строго или сильно выпуклая функция, то S^* содержит только одну точку: $S^* = \{x^*\}$.

Задачи с ограничениями-равенствами

В задачах без ограничений всё довольно интуитивно. В этом разделе мы добавим одно ограничение-равенство, то есть:

Задачи с ограничениями-равенствами

В задачах без ограничений всё довольно интуитивно. В этом разделе мы добавим одно ограничение-равенство, то есть:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & h(x) = 0
\end{split}
$$

Задачи с ограничениями-равенствами

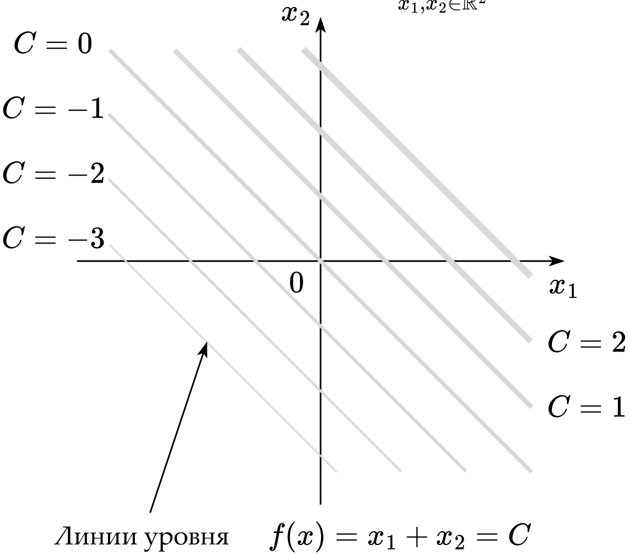
В задачах без ограничений всё довольно интуитивно. В этом разделе мы добавим одно ограничение-равенство, то есть:

$$
\begin{aligned}
f(x) &\to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } h(x) &= 0
\end{aligned}
$$

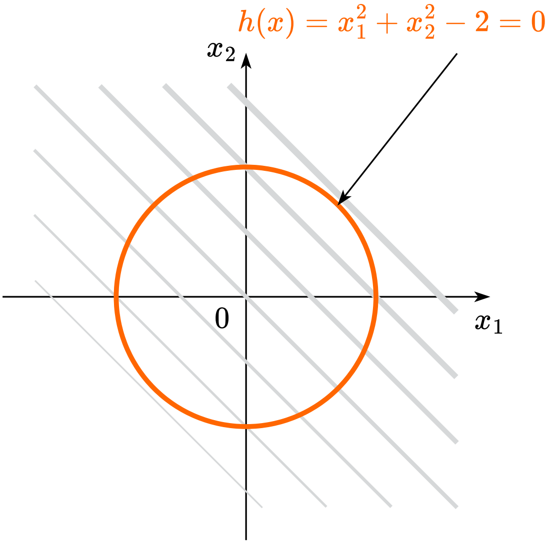
Мы попробуем проиллюстрировать подход к решению этой задачи через простой пример с $f(x) = x_1 + x_2$ и $h(x) = x_1^2 + x_2^2 - 2$.

Задачи с ограничениями-равенствами

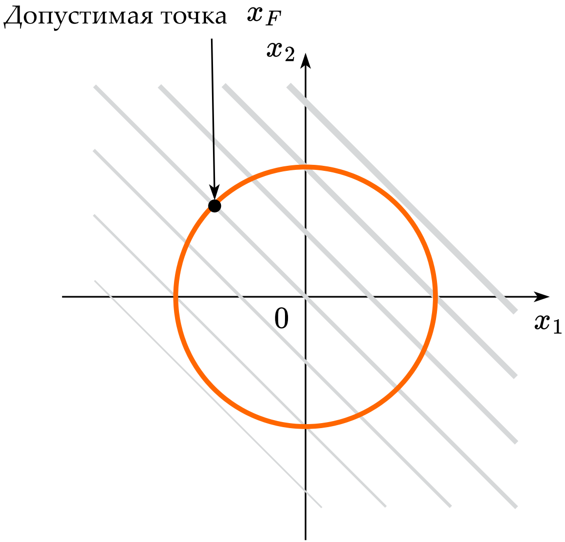
$$f(x) = x_1 + x_2 \to \min_{x_1, x_2 \in \mathbb{R}^2}$$



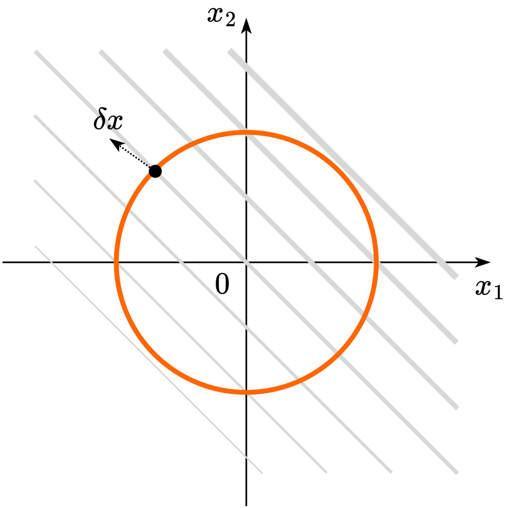
Задачи с ограничениями-равенствами



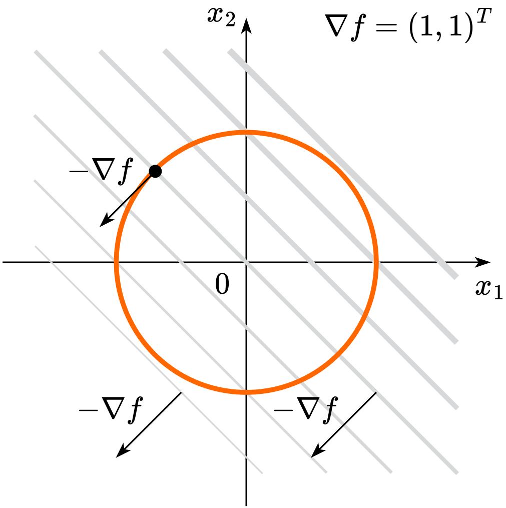
Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами

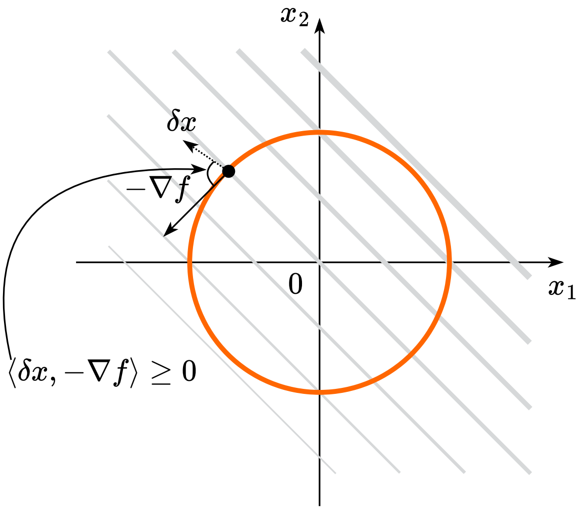


Задачи с ограничениями-равенствами



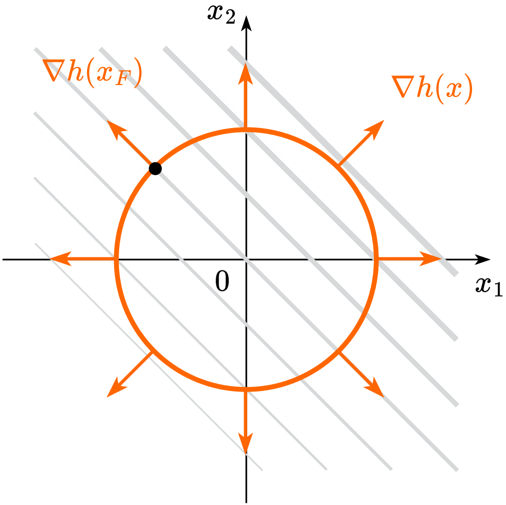
Задачи с ограничениями-равенствами

Мы хотим: $f(x_F + \delta x) \leq f(x_F)$

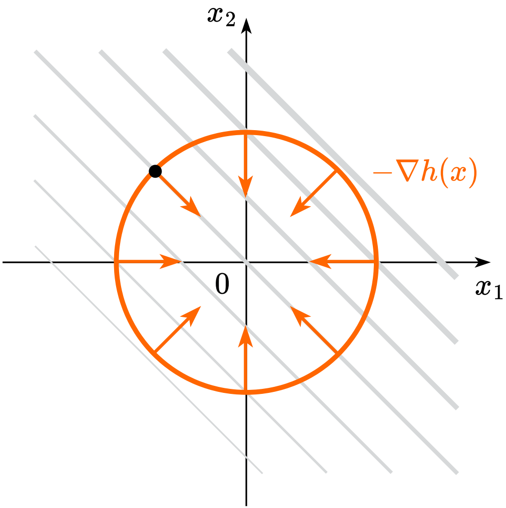


Задачи с ограничениями-равенствами

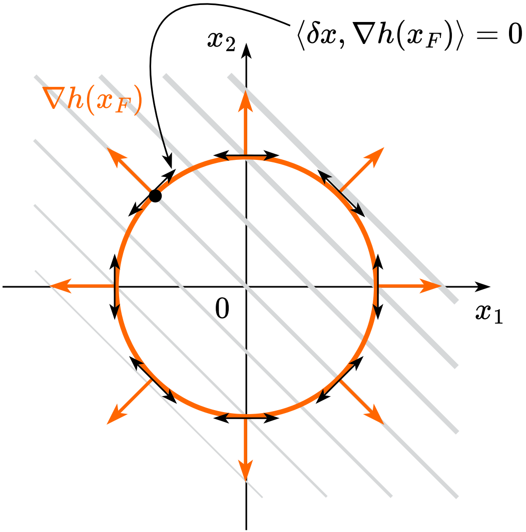
$$\nabla h = (2x_1, 2x_2)^T$$



Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Предположим, что в процессе такого движения мы пришли в точку, где

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Предположим, что в процессе такого движения мы пришли в точку, где

$$-\nabla f(x) = \nu \nabla h(x)$$

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Предположим, что в процессе такого движения мы пришли в точку, где

$$-\nabla f(x) = \nu \nabla h(x)$$

$$\langle \delta x, -\nabla f(x) \rangle = \langle \delta x, \nu \nabla h(x) \rangle = 0$$

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

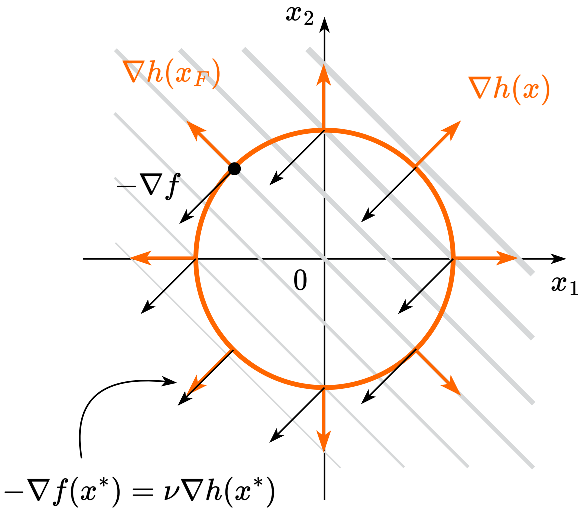
Предположим, что в процессе такого движения мы пришли в точку, где

$$-\nabla f(x) = \nu \nabla h(x)$$

$$\langle \delta x, -\nabla f(x) \rangle = \langle \delta x, \nu \nabla h(x) \rangle = 0$$

Тогда мы достигли такой точки допустимого множества, из которой нельзя уменьшить значение функции при допустимых малых сдвигах. Это и есть условие локального минимума в задаче с ограничением.

Задачи с ограничениями-равенствами



Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$
L(x, \nu) = f(x) + \nu h(x)
$$

Если задача регулярная (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача регулярная (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

Важно отметить, что $L(x^*, \nu^*) = f(x^*)$.

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача регулярная (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

$\nabla_x L(x^*, \nu^*) = 0$ это мы уже написали выше

Важно отметить, что $L(x^*, \nu^*) = f(x^*)$.

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача регулярная (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

$\nabla_x L(x^*, \nu^*) = 0$ это мы уже написали выше

$\nabla_\nu L(x^*, \nu^*) = 0$ бюджетное ограничение

Важно отметить, что $L(x^*, \nu^*) = f(x^*)$.

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача регулярная (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

$\nabla_x L(x^*, \nu^*) = 0$ это мы уже написали выше

$\nabla_\nu L(x^*, \nu^*) = 0$ бюджетное ограничение

Достаточные условия

Важно отметить, что $L(x^*, \nu^*) = f(x^*)$.

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача регулярная (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

$\nabla_x L(x^*, \nu^*) = 0$ это мы уже написали выше

$\nabla_\nu L(x^*, \nu^*) = 0$ бюджетное ограничение

Достаточные условия

$\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$

Важно отметить, что $L(x^*, \nu^*) = f(x^*)$.

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача регулярная (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

$\nabla_x L(x^*, \nu^*) = 0$ это мы уже написали выше

$\nabla_\nu L(x^*, \nu^*) = 0$ бюджетное ограничение

Достаточные условия

$\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$

$\forall y \neq 0 \in \mathbb{R}^n : \nabla h(x^*)^\top y = 0$

Важно отметить, что $L(x^*, \nu^*) = f(x^*)$.

Задачи с ограничениями-равенствами

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

$$
L(x, \nu) = f(x) + \sum_{i=1}^p \nu_i h_i(x) = f(x) + \nu^\top h(x)
$$

Пусть $f(x)$ и $h_i(x)$ дважды дифференцируемы в точке x^* и непрерывно дифференцируемы в некоторой окрестности x^*. Условия локального минимума для $x \in \mathbb{R}^n, \nu \in \mathbb{R}^p$ записываются как

Необходимые условия

$$\nabla_x L(x^*, \nu^*) = 0$$

$$\nabla_\nu L(x^*, \nu^*) = 0$$

ECP: Достаточные условия

$$\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$$

$$\forall y \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^\top y = 0$$

Задача наименьших квадратов

Example

Поставим задачу оптимизации и решим её для линейной системы $Ax = b, A \in \mathbb{R}^{m \times n}$ для трёх случаев (предполагая, что матрица имеет полный ранг):

- $m < n$

Задача наименьших квадратов

Example

Поставим задачу оптимизации и решим её для линейной системы $Ax = b, A \in \mathbb{R}^{m \times n}$ для трёх случаев (предполагая, что матрица имеет полный ранг):

- $m < n$
- $m = n$

Задача наименьших квадратов

Example

Поставим задачу оптимизации и решим её для линейной системы $Ax = b, A \in \mathbb{R}^{m \times n}$ для трёх случаев (предполагая, что матрица имеет полный ранг):

- $m < n$
- $m = n$
- $m > n$

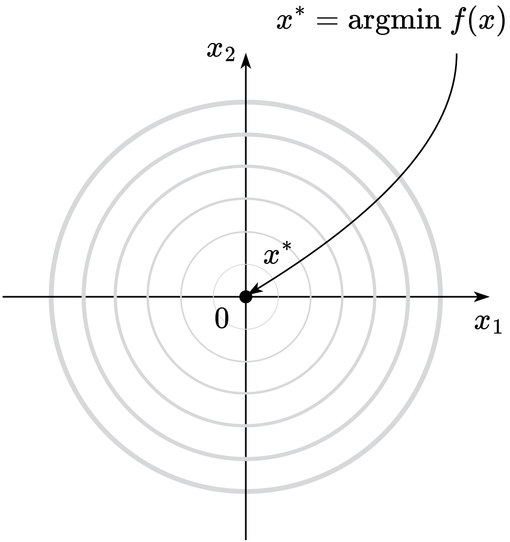
Задачи с ограничениями-неравенствами

Пример задачи с ограничениями-неравенствами

$$f(x) = x_1^2 + x_2^2 \quad g(x) = x_1^2 + x_2^2 - 1$$

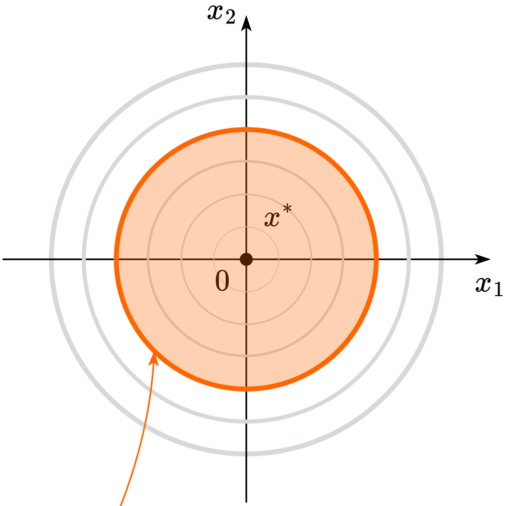
$$\begin{split} & f(x) \to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } & g(x) \leq 0 \end{split}$$

Задачи с ограничениями-неравенствами



Линии уровня $f(x) = x_1^2 + x_2^2 = C$

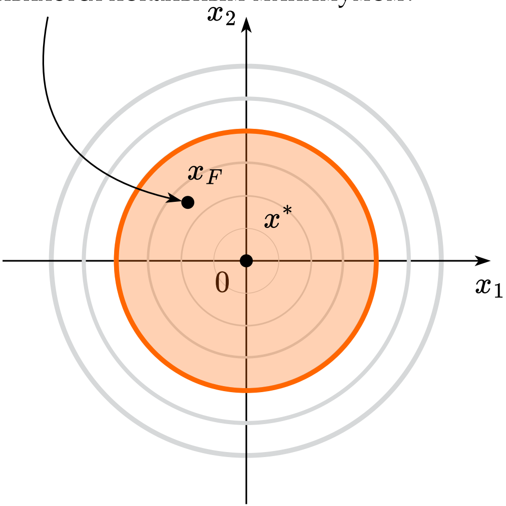
Задачи с ограничениями-неравенствами



Бюджетное множество $g(x) = x_1^2 + x_2^2 - 1 \leq 0$

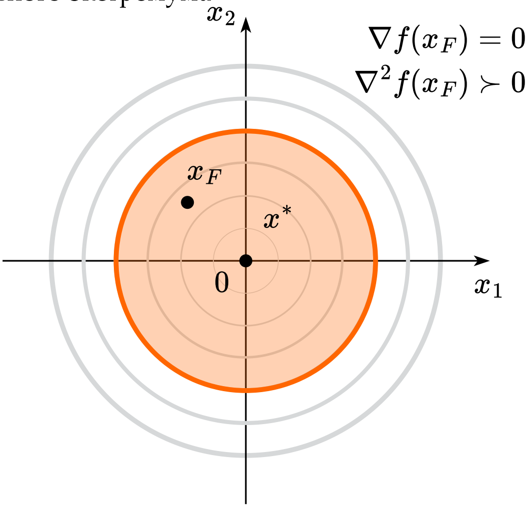
Задачи с ограничениями-неравенствами

Как понять, что некоторая допустимая
точка является локальным минимумом?



Задачи с ограничениями-неравенствами

Просто! Проверим достаточные условия локального экстремума



Задачи с ограничениями-неравенствами

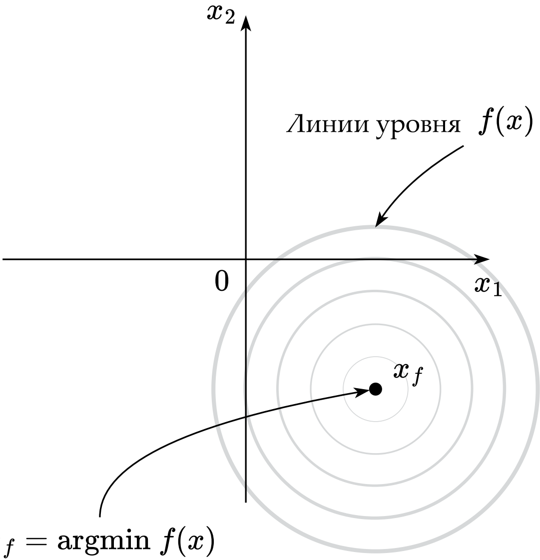
Таким образом, если ограничения типа неравенства неактивны в условной задаче, то мы можем решать задачу без ограничений. Однако так бывает не всегда. Рассмотрим второй простой пример.

$$f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 \quad g(x) = x_1^2 + x_2^2 - 1$$

$$
\begin{aligned}
&f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } &g(x) \leq 0
\end{aligned}
$$

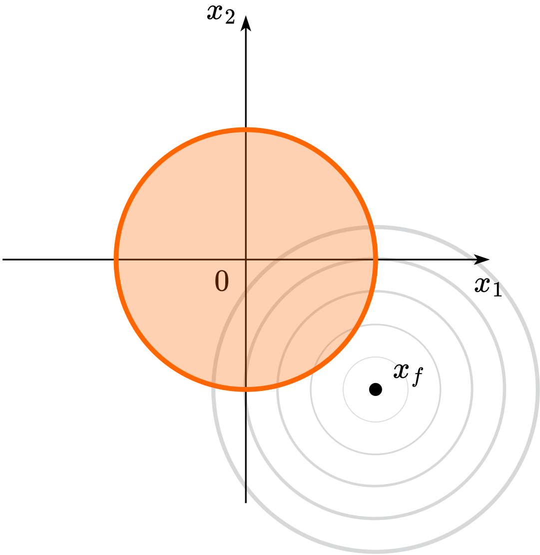
Задачи с ограничениями-неравенствами

$$f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 = C$$



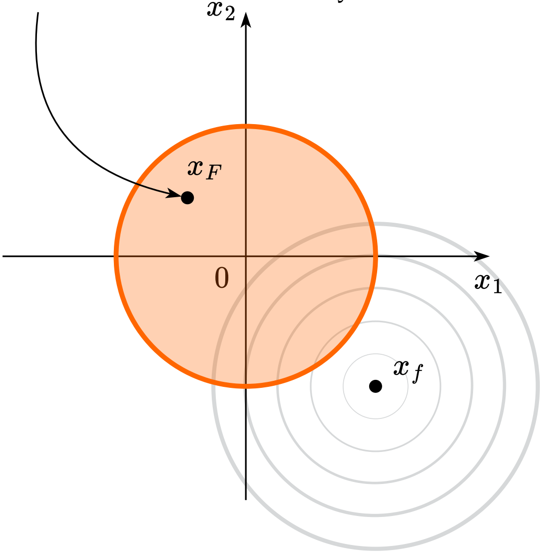
Задачи с ограничениями-неравенствами

Бюджетное множество $g(x) = x_1^2 + x_2^2 - 1 \leq 0$



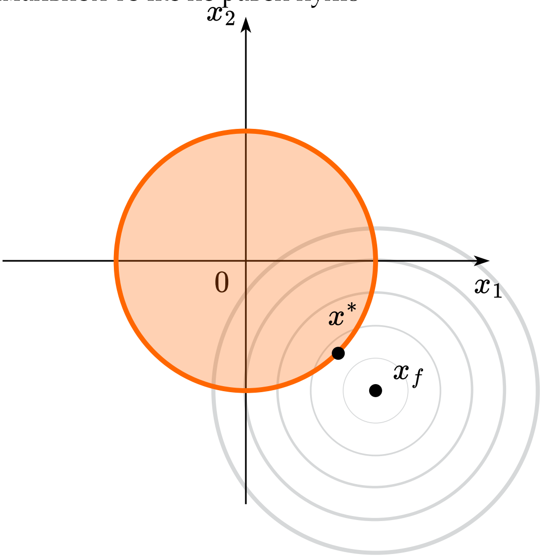
Задачи с ограничениями-неравенствами

Как понять, что некоторая допустимая точка является локальным минимумом?



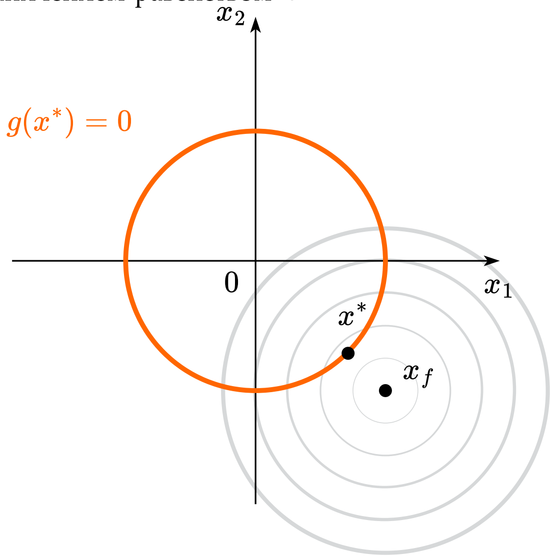
Задачи с ограничениями-неравенствами

Не так просто! Даже градиент
в оптимальной точке не равен нулю 😭

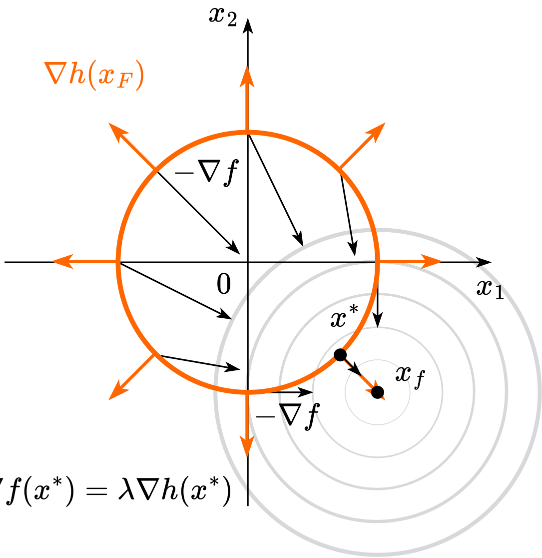


Задачи с ограничениями-неравенствами

Фактически имеем задачу
с ограничением-равенством 💡



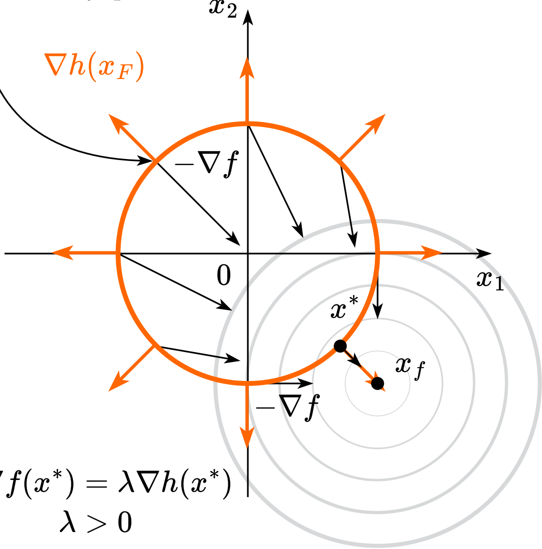
Задачи с ограничениями-неравенствами



$$-\nabla f(x^*) = \lambda \nabla h(x^*)$$

Задачи с ограничениями-неравенствами

Не является локальным минимумом, т.к. $-\nabla f(x)$ направлен внутрь бюджетного множества



$$-\nabla f(x^*) = \lambda \nabla h(x^*)$$
$$\lambda > 0$$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

$$
\text{s.t. } g(x) \leq 0
$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) \succ 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

$$
\text{s.t. } g(x) \leq 0
$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) \succ 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

$$\text{s.t. } g(x) \leq 0$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) \succ 0$

$g(x) \leq 0$ активно: $g(x^*) = 0$

- $g(x^*) = 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

$$
f(x) \to \min_{x \in \mathbb{R}^n}
$$
$$
\text{s.t. } g(x) \leq 0
$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) \succ 0$

$g(x) \leq 0$ активно: $g(x^*) = 0$

- $g(x^*) = 0$
- Необходимые условия: $-\nabla f(x^*) = \lambda \nabla g(x^*)$, $\lambda > 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

$$
\text{s.t. } g(x) \leq 0
$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) \succ 0$

$g(x) \leq 0$ активно: $g(x^*) = 0$

- $g(x^*) = 0$
- Необходимые условия: $-\nabla f(x^*) = \lambda \nabla g(x^*)$, $\lambda > 0$
- Достаточные условия: $\langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0, \forall y \neq 0 \in \mathbb{R}^n : \nabla g(x^*)^\top y = 0$

Лагранжиан для задач с ограничениями-неравенствами

Объединяя два возможных случая, мы можем записать общие условия для задачи:

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

$$\text{s.t. } g(x) \leq 0$$

Определим функцию Лагранжа:

$$L(x, \lambda) = f(x) + \lambda g(x)$$

Классические условия оптимальности первого и второго порядка Каруша-Куна-Таккера для локального минимума x^*, сформулированные при некоторых условиях регулярности, можно записать следующим образом.

Лагранжиан для задач с ограничениями-неравенствами

Объединяя два возможных случая, мы можем записать общие условия для задачи:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

Определим функцию Лагранжа:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

Классические условия оптимальности первого и второго порядка Каруша-Куна-Таккера для локального минимума x^*, сформулированные при некоторых условиях регулярности, можно записать следующим образом.

Если x^* является локальным минимумом для описанной выше задачи, то существует единственный множитель Лагранжа λ^* такой, что:

$$
(1)\ \nabla_x L(x^*, \lambda^*) = 0
$$

Лагранжиан для задач с ограничениями-неравенствами

Объединяя два возможных случая, мы можем записать общие условия для задачи:

$$
f(x) \to \min_{x \in \mathbb{R}^n}
$$
$$
\text{s.t. } g(x) \leq 0
$$

Определим функцию Лагранжа:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

Классические условия оптимальности первого и второго порядка Каруша-Куна-Таккера для локального минимума x^*, сформулированные при некоторых условиях регулярности, можно записать следующим образом.

Если x^* является локальным минимумом для описанной выше задачи, то существует единственный множитель Лагранжа λ^* такой, что:

$$
(1)\ \nabla_x L(x^*, \lambda^*) = 0
$$
$$
(2)\ \lambda^* \geq 0
$$

Лагранжиан для задач с ограничениями-неравенствами

Объединяя два возможных случая, мы можем записать общие условия для задачи:

$$
\begin{aligned}
&f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } &g(x) \leq 0
\end{aligned}
$$

Определим функцию Лагранжа:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

Классические условия оптимальности первого и второго порядка Каруша-Куна-Таккера для локального минимума x^*, сформулированные при некоторых условиях регулярности, можно записать следующим образом.

Если x^* является локальным минимумом для описанной выше задачи, то существует единственный множитель Лагранжа λ^* такой, что:

$$
\begin{aligned}
&(1)\ \nabla_x L(x^*, \lambda^*) = 0 \\
&(2)\ \lambda^* \geq 0 \\
&(3)\ \lambda^* g(x^*) = 0
\end{aligned}
$$

Лагранжиан для задач с ограничениями-неравенствами

Объединяя два возможных случая, мы можем записать общие условия для задачи:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

Определим функцию Лагранжа:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

Классические условия оптимальности первого и второго порядка Каруша-Куна-Таккера для локального минимума x^*, сформулированные при некоторых условиях регулярности, можно записать следующим образом.

Если x^* является локальным минимумом для описанной выше задачи, то существует единственный множитель Лагранжа λ^* такой, что:

$$
\begin{split}
& (1) \; \nabla_x L(x^*, \lambda^*) = 0 \\
& (2) \; \lambda^* \geq 0 \\
& (3) \; \lambda^* g(x^*) = 0 \\
& (4) \; g(x^*) \leq 0
\end{split}
$$

Общая формулировка

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

Данная формулировка представляет собой общую задачу математического программирования.

Решение включает в себя построение лагранжиана:

$$
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)
$$

Необходимые условия

Пусть x^*, (λ^*, ν^*) является решением задачи математического программирования с нулевым двойственным зазором (оптимальное значение прямой задачи p^* равно оптимальному значению двойственной задачи d^*). Пусть также функции f_0, f_i, h_i дифференцируемы.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$

Необходимые условия

Пусть x^*, (λ^*, ν^*) является решением задачи математического программирования с нулевым двойственным зазором (оптимальное значение прямой задачи p^* равно оптимальному значению двойственной задачи d^*). Пусть также функции f_0, f_i, h_i дифференцируемы.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$

Необходимые условия

Пусть x^*, (λ^*, ν^*) является решением задачи математического программирования с нулевым двойственным зазором (оптимальное значение прямой задачи p^* равно оптимальному значению двойственной задачи d^*). Пусть также функции f_0, f_i, h_i дифференцируемы.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda^*_i \geq 0, i = 1, \ldots, m$

Необходимые условия

Пусть x^*, (λ^*, ν^*) является решением задачи математического программирования с нулевым двойственным зазором (оптимальное значение прямой задачи p^* равно оптимальному значению двойственной задачи d^*). Пусть также функции f_0, f_i, h_i дифференцируемы.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda^*_i \geq 0, i = 1, \ldots, m$
- $\lambda^*_i f_i(x^*) = 0, i = 1, \ldots, m$

Необходимые условия

Пусть x^*, (λ^*, ν^*) является решением задачи математического программирования с нулевым двойственным зазором (оптимальное значение прямой задачи p^* равно оптимальному значению двойственной задачи d^*). Пусть также функции f_0, f_i, h_i дифференцируемы.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda_i^* \geq 0, i = 1, \ldots, m$
- $\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m$
- $f_i(x^*) \leq 0, i = 1, \ldots, m$

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, при наличии регулярности можно записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с полуопределённым гессианом лагранжиана.

- **Условие Слейтера.** Если для выпуклой задачи (при минимизации, с выпуклыми f_0, f_i и аффинными h_i) существует точка x такая, что $h(x) = 0$ и $f_i(x) < 0$ (существует строго допустимая точка), то двойственный зазор равен нулю и условия Каруша-Куна-Таккера становятся необходимыми и достаточными.

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, при наличии регулярности можно записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с полуопределённым гессианом лагранжиана.

- **Условие Слейтера.** Если для выпуклой задачи (при минимизации, с выпуклыми f_0, f_i и аффинными h_i) существует точка x такая, что $h(x) = 0$ и $f_i(x) < 0$ (существует строго допустимая точка), то двойственный зазор равен нулю и условия Каруша-Куна-Таккера становятся необходимыми и достаточными.
- **Условие линейной квалификации ограничений.** Если f_i и h_i являются аффинными функциями, то никаких других условий не требуется.

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, при наличии регулярности можно записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с полуопределённым гессианом лагранжиана.

- **Условие Слейтера.** Если для выпуклой задачи (при минимизации, с выпуклыми f_0, f_i и аффинными h_i) существует точка x такая, что $h(x) = 0$ и $f_i(x) < 0$ (существует строго допустимая точка), то двойственный зазор равен нулю и условия Каруша-Куна-Таккера становятся необходимыми и достаточными.
- **Условие линейной квалификации ограничений.** Если f_i и h_i являются аффинными функциями, то никаких других условий не требуется.
- **Условие линейной независимости ограничений.** Градиенты активных ограничений неравенства и градиенты ограничений равенства линейно независимы в точке x^*.

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, при наличии регулярности можно записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с полуопределённым гессианом лагранжиана.

- **Условие Слейтера.** Если для выпуклой задачи (при минимизации, с выпуклыми f_0, f_i и аффинными h_i) существует точка x такая, что $h(x) = 0$ и $f_i(x) < 0$ (существует строго допустимая точка), то двойственный зазор равен нулю и условия Каруша-Куна-Таккера становятся необходимыми и достаточными.
- **Условие линейной квалификации ограничений.** Если f_i и h_i являются аффинными функциями, то никаких других условий не требуется.
- **Условие линейной независимости ограничений.** Градиенты активных ограничений неравенства и градиенты ограничений равенства линейно независимы в точке x^*.
- Для других примеров см. wiki.

Пример. Проекция на гиперплоскость

$$
\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.
$$

Пример. Проекция на гиперплоскость

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Решение

Лагранжиан:

Пример. Проекция на гиперплоскость

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Решение

Лагранжиан:

$$L(\mathbf{x}, \nu) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 + \nu(\mathbf{a}^T\mathbf{x} - b)$$

Пример. Проекция на гиперплоскость

$$
\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.
$$

Решение

Лагранжиан:

$$
L(\mathbf{x}, \nu) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2 + \nu(\mathbf{a}^T\mathbf{x} - b)
$$

Производная L по $\mathbf{x}$:

$$
\frac{\partial L}{\partial \mathbf{x}} = \mathbf{x} - \mathbf{y} + \nu\mathbf{a} = 0, \qquad \mathbf{x} = \mathbf{y} - \nu\mathbf{a}
$$

Пример. Проекция на гиперплоскость

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Решение

Лагранжиан:

$$L(\mathbf{x},\nu) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 + \nu(\mathbf{a}^T\mathbf{x} - b)$$

Производная L по $\mathbf{x}$:

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{x} - \mathbf{y} + \nu\mathbf{a} = 0, \qquad \mathbf{x} = \mathbf{y} - \nu\mathbf{a}$$

$$\mathbf{a}^T\mathbf{x} = \mathbf{a}^T\mathbf{y} - \nu\mathbf{a}^T\mathbf{a} \qquad \nu = \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}$$

Пример. Проекция на гиперплоскость

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Решение

Лагранжиан:

$$L(\mathbf{x}, \nu) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 + \nu(\mathbf{a}^T\mathbf{x} - b)$$

Производная L по $\mathbf{x}$:

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{x} - \mathbf{y} + \nu\mathbf{a} = 0, \qquad \mathbf{x} = \mathbf{y} - \nu\mathbf{a}$$

$$\mathbf{a}^T\mathbf{x} = \mathbf{a}^T\mathbf{y} - \nu\mathbf{a}^T\mathbf{a} \qquad \nu = \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}$$

$$\mathbf{x} = \mathbf{y} - \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}\mathbf{a}$$

Пример. Проекция на единичный симплекс

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0.$$

Пример. Проекция на единичный симплекс

$$
\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0.
$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$
L = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^\top \mathbf{1} - 1)
$$

Пример. Проекция на единичный симплекс

$$
\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0.
$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$
L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^\top \mathbf{1} - 1)
$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$

Пример. Проекция на единичный симплекс

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0.$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^\top \mathbf{1} - 1)$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$

Пример. Проекция на единичный симплекс

$$
\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^\top\mathbf{1} = 1, \quad \mathbf{x} \geq 0.
$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$
L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^\top\mathbf{1} - 1)
$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$

Пример. Проекция на единичный симплекс

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0.$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^\top \mathbf{1} - 1)$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $\mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0$

Пример. Проекция на единичный симплекс

$$
\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0.
$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$
L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^\top \mathbf{1} - 1)
$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $\mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0$

Пример. Проекция на единичный симплекс

$$
\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0.
$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$
L = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^\top \mathbf{1} - 1)
$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $\mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0$

Question

Решите систему выше за $O(n \log n)$.

Пример. Проекция на единичный симплекс

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0.$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^\top \mathbf{1} - 1)$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $\mathbf{x}^\top \mathbf{1} = 1, \quad \mathbf{x} \geq 0$

> **Question**
>
> Решите систему выше за $O(n \log n)$.

> **Question**
>
> Решите систему выше за $O(n)$.

Ссылки

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе "Элементы статистического обучения" @ KTH.

Ссылки

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе "Элементы статистического обучения" @ KTH.
- Однострочное доказательство ККТ

Ссылки

* Лекция по условиям ККТ (очень интуитивное объяснение) в курсе "Элементы статистического обучения" @ KTH.
* Однострочное доказательство ККТ
* О втором порядке оптимальности для задач оптимизации с ограничениями неравенства

Ссылки

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе "Элементы статистического обучения" @ KTH.
- Однострочное доказательство ККТ
- О втором порядке оптимальности для задач оптимизации с ограничениями неравенства
- О втором порядке оптимальности в нелинейной оптимизации

Ссылки

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе "Элементы статистического обучения" @ KTH.
- Однострочное доказательство ККТ
- О втором порядке оптимальности для задач оптимизации с ограничениями неравенства
- О втором порядке оптимальности в нелинейной оптимизации
- Численная оптимизация by Jorge Nocedal and Stephen J. Wright.