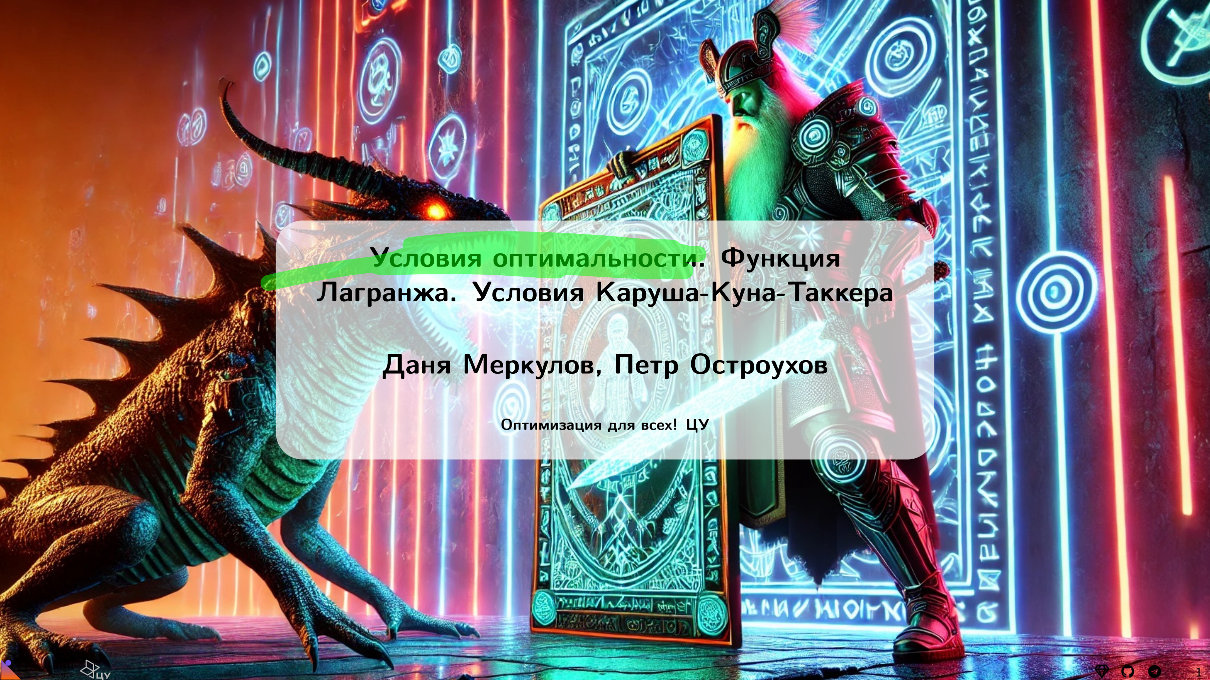


Условия оптимальности. Функция Лагранжа. Условия Каруша-Куна-Таккера

Даня Меркулов, Петр Остроухов

Оптимизация для всех! ЦУ

В этой ~~работе~~ лекции совершенно отсутствуют какие бы то ни было ~~чертежи~~ доказательства. Излагаемые мною методы не требуют ни построений, ни геометрических или механических рассуждений; они требуют только алгебраических операций, подчиненных планомерному и однообразному алгоритму.

—Предисловие к “Аналитической механике”

$$\min f(x) \leftarrow \text{целевая функция}$$

$$x \in S \leftarrow \text{бюджетное мн-во}$$

Безусловные задачи

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

Условные задачи

ограничения-равенства	ограниц-неравенства
$\min x^2$, $x=3$	$\min x^2$, $x \leq 3$
	$\min x^2$, $x \leq -3$

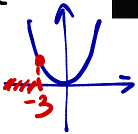
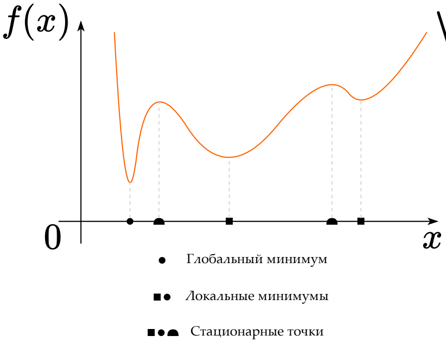


Рис. 1: Жозеф Луи Лагранж

Условия оптимальности

Теория

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



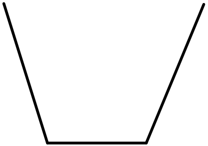


Рис. 2: Иллюстрация различных стационарных (критических) точек

Теория

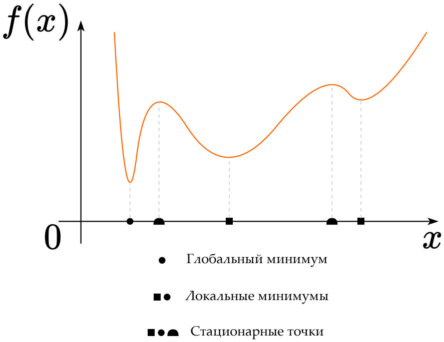


Рис. 2: Иллюстрация различных стационарных (критических) точек

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

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

Теория

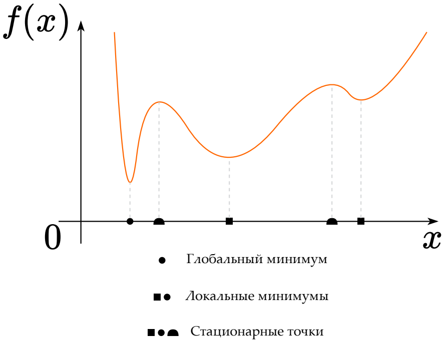


Рис. 2: Иллюстрация различных стационарных (критических) точек

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

Множество S обычно называется **допустимым множеством** (или **бюджетным множеством**).
Мы говорим, что задача имеет решение, если бюджетное множество, в котором достигается минимум или инфимум данной функции, **не пусто**: $x^* \in S$.

$$\min_{x \in \mathbb{R}} x$$

Теория

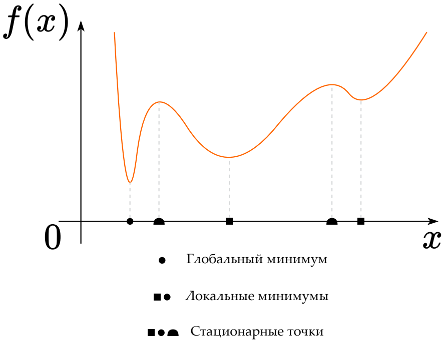


Рис. 2: Иллюстрация различных стационарных (критических) точек

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

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

Мы говорим, что задача имеет решение, если бюджетное множество, в котором достигается минимум или инфимум данной функции, **не пусто**: $x^* \in S$.

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

Теория

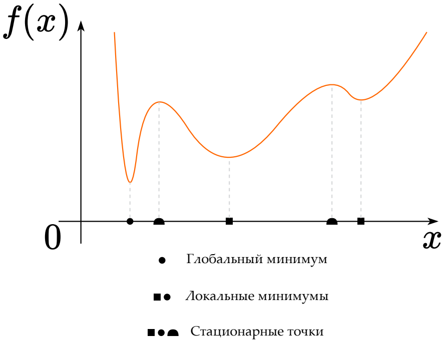


Рис. 2: Иллюстрация различных стационарных (критических) точек

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

Множество S обычно называется **допустимым множеством** (или **бюджетным множеством**).
Мы говорим, что задача имеет решение, если бюджетное множество, в котором достигается минимум или инфимум данной функции, **не пусто**: $x^* \in S$.

- 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

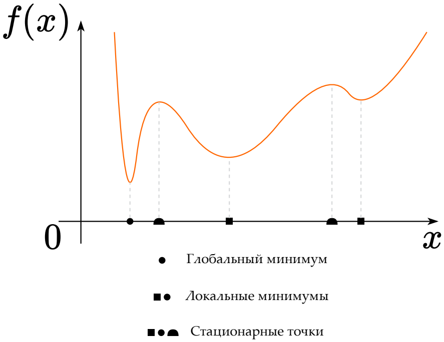


Рис. 2: Иллюстрация различных стационарных (критических) точек

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

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

Мы говорим, что задача имеет решение, если бюджетное множество, в котором достигается минимум или инфимум данной функции, **не пусто**: $x^* \in S$.

- 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

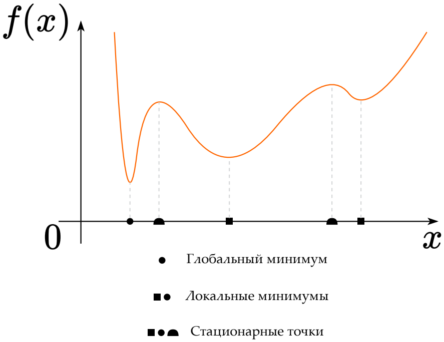


Рис. 2: Иллюстрация различных стационарных (критических) точек

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

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

Мы говорим, что задача имеет решение, если бюджетное множество, в котором достигается минимум или инфимум данной функции, **не пусто**: $x^* \in S$.

- 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.

Рис. 3: 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.

Рис. 3: 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.

Рис. 3: 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:

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

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

для некоторого $t \in (0,1)$.

EASY

Безусловная оптимизация

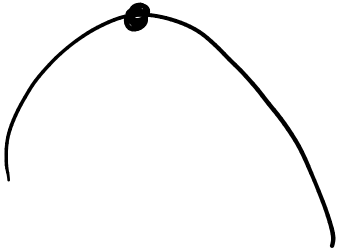
$$S = \mathbb{R}^n$$

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 $\nabla f(x^*) \neq 0$. Определим вектор $p = -\nabla f(x^*)$ 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 ∇f непрерывна в окрестности x^*, существует скаляр $T > 0$ такой, что

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

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

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

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

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

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

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

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

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

Тогда x^* является строгим локальным минимумом функции f.

пример

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

$f(x) = x^2$

$x = 1$

$f(x) = x^3$

$6x$

6

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

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

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 определен в x^*, мы можем выбрать радиус $r > 0$ такой, что $\nabla^2 f(x)$ остается положительно определенным для всех x в открытом шаре $B = \{z \mid \|z - x^*\| < r\}$. Возьмем 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 определен в x^*, мы можем выбрать радиус $r > 0$ такой, что $\nabla^2 f(x)$ остается положительно определенным для всех x в открытом шаре $B = \{z \mid \|z - x^*\| < r\}$. Возьмем 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.

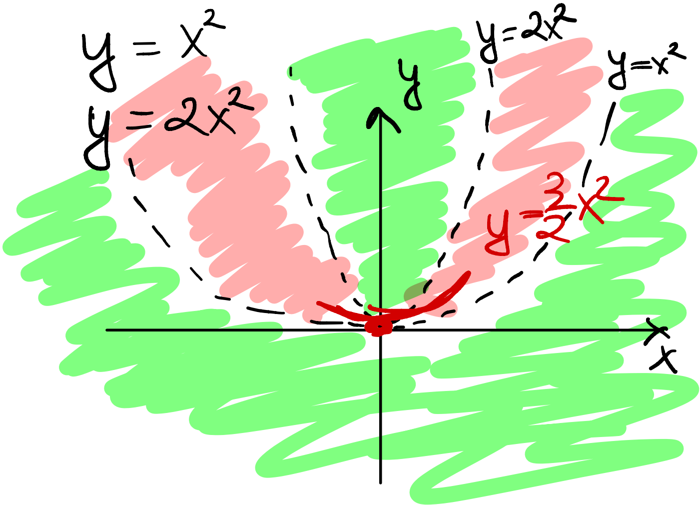
$\forall x \in \mathbb{R}^n$

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

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.

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



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.

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

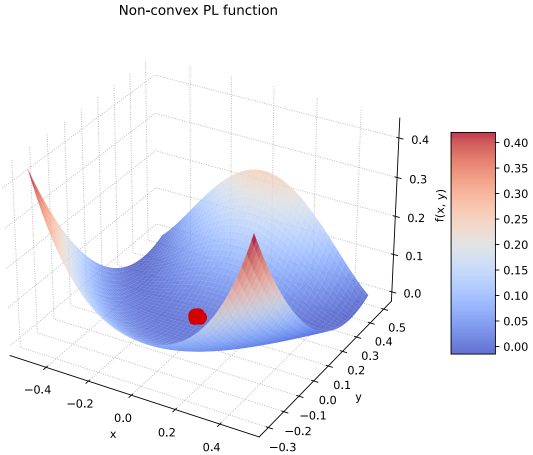
Хотя поверхность не имеет локального минимума в начале координат, ее пересечение с любой вертикальной плоскостью, проходящей через начало координат (плоскость с уравнением $y = mx$ или $x = 0$) является кривой, которая имеет локальный минимум в начале координат. Другими словами, если точка начинает движение в начале 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 $y = \sqrt{2}x^2$, приведет 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, ее 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 в начале 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 $y = \sqrt{2}x^2$, приведет к уменьшению значения функции.



Безуслов.

$\nabla f(x^*) = 0 \Leftarrow x^*$ - лок. экстр.

$\nabla f(x^*) = 0$
$\nabla^2 f(x^*) \succ 0$ $\Rightarrow x^*$ - лок. экстр. мин. (макс.)

Условная оптимизация

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

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.

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.

Пусть $S \subseteq \mathbb{R}^n$ и функция $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума для f над S, и предположим далее, 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.

Пусть $S \subseteq \mathbb{R}^n$ и функция $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума для f над S, и предположим далее, 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.

Пусть $S \subseteq \mathbb{R}^n$ и функция $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума для f над S, и предположим далее, 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.

Пусть $S \subseteq \mathbb{R}^n$ и функция $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума для f над S, и предположим далее, 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.

Пусть $S \subseteq \mathbb{R}^n$ и функция $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума для f над S, и предположим далее, 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

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

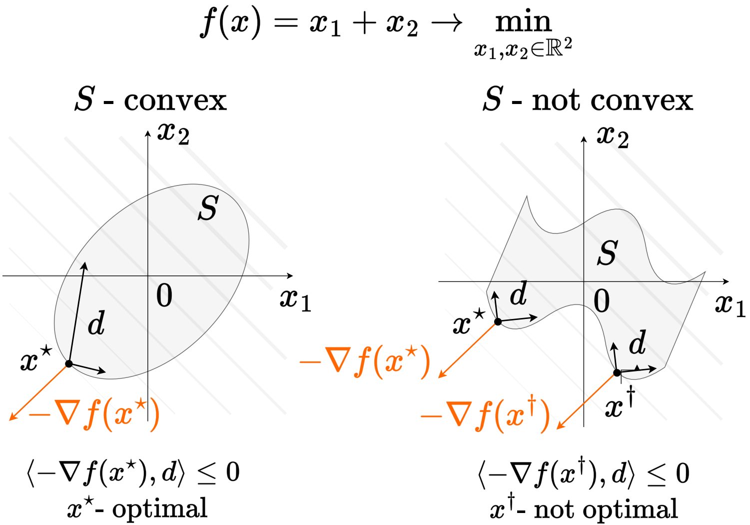


Рис. 4: Общее условие локальной оптимальности первого порядка

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.

Еще 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.

Еще 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.

Еще 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.

Еще 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

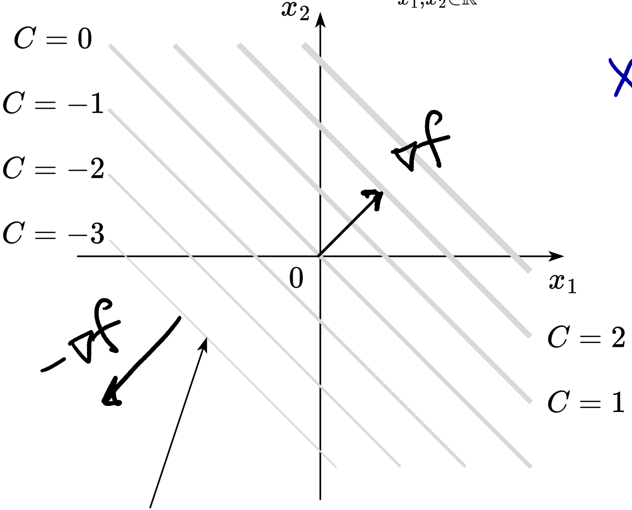
$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \nabla f$

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

$f: \mathbb{R}^2 \to \mathbb{R}$

$x_1 + x_2 = C$

$x_2 = C - x_1$

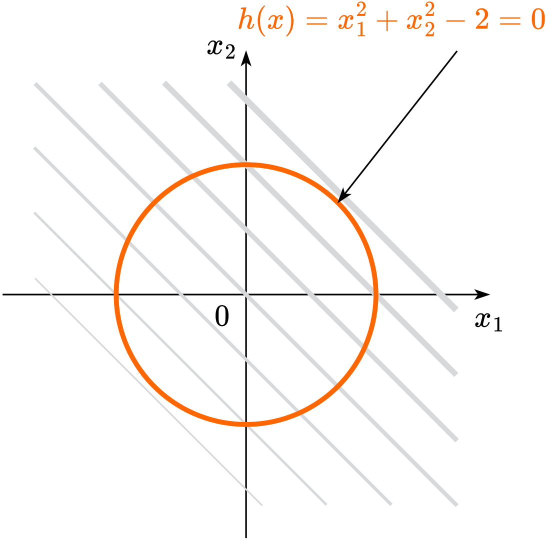


Линии уровня $f(x) = x_1 + x_2 = C$

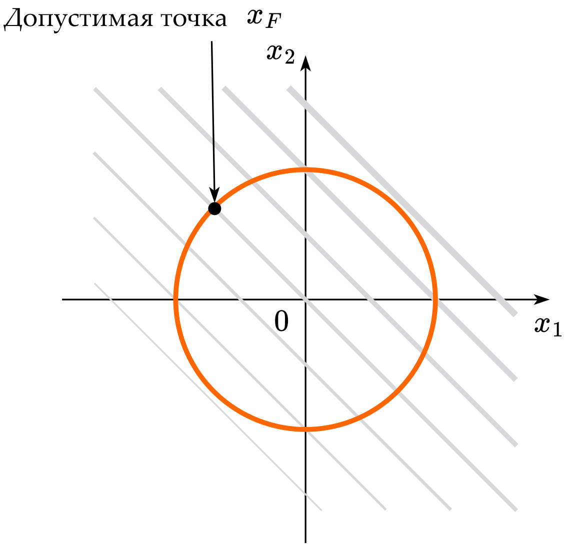
Задачи с ограничениями-равенствами

$h(x) = 0$

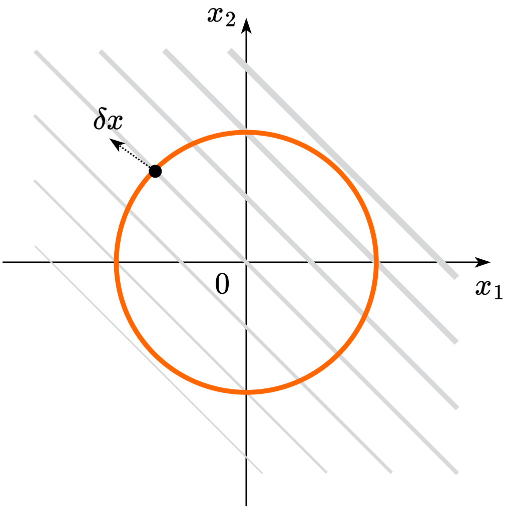
$x_1^2 + x_2^2 = 2$



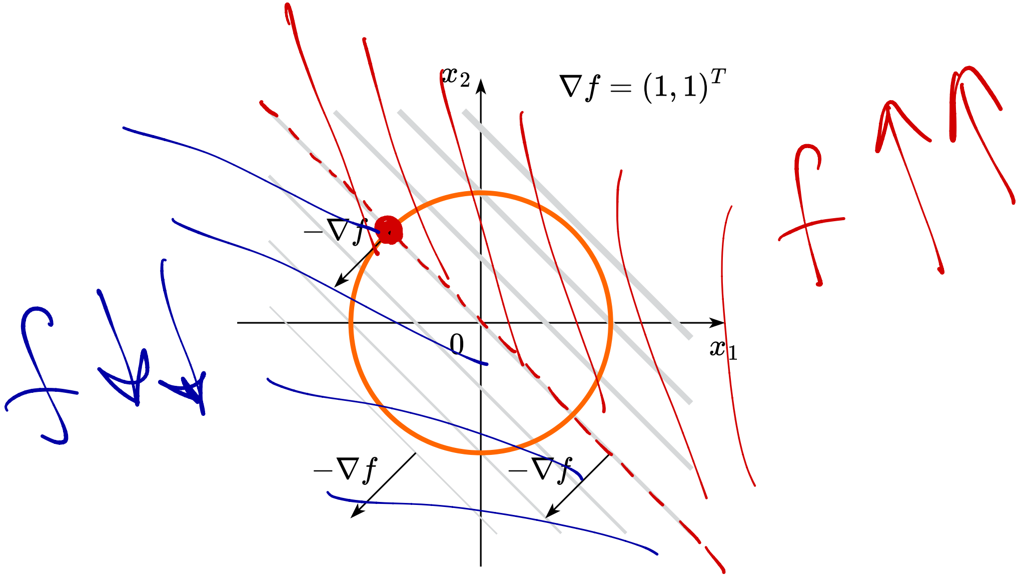
Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами

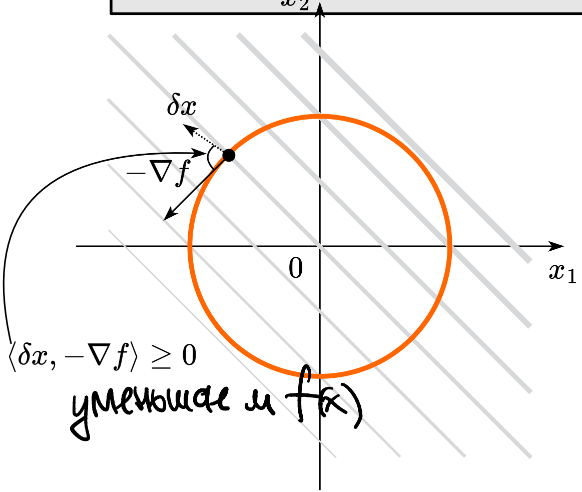


Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами

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



уменьшаем $f(x)$

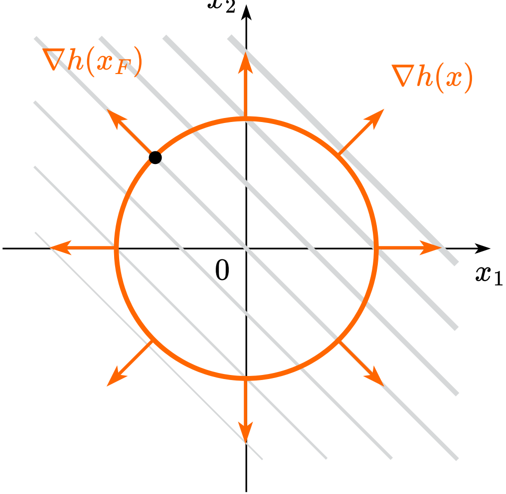
$\langle \delta x, -\nabla f(x) \rangle \leq$

увеличиваем

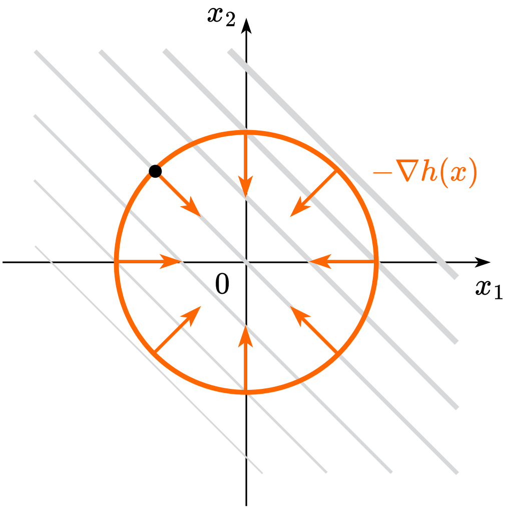
Задачи с ограничениями-равенствами

$h(x) = x_1^2 + x_2^2 - 2$

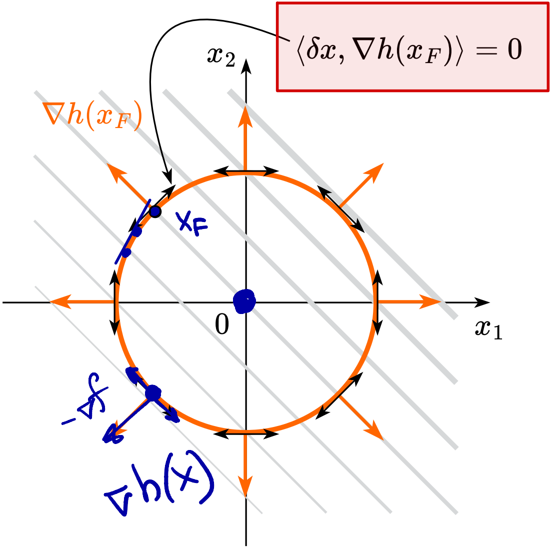
$$\nabla h = (2x_1, 2x_2)^T$$



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:

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

бюджетное ограничение

Задачи с ограничениями-равенствами

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

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

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

уменьшение знач. $f(x)$

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

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

ν – скаляр

$$\boxed{-\nabla f(x) = \nu \nabla h(x)}$$

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:

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

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

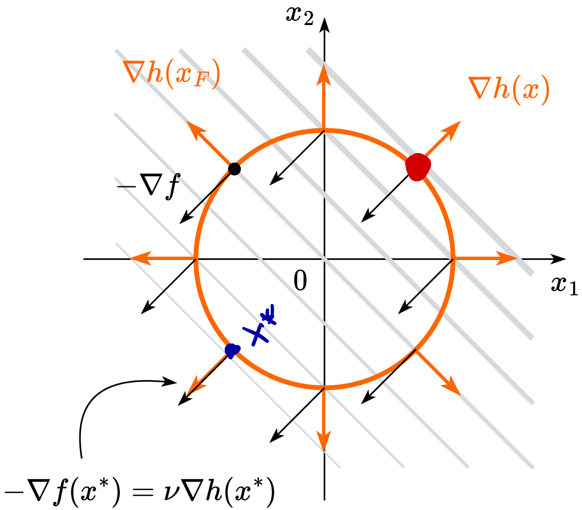
Предположим, что в процессе такого движения мы пришли в точку, где

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

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

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):

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

функция Лагранжа

$$\nabla_x L = 0 \qquad \nabla_\nu L = \boxed{h(x) = 0}$$

$$\nabla f(x) + \nu \cdot \nabla h(x) = 0$$

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

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

множитель Лагранжа

$\min f(x)$

$h(x) = 0$

Давайте определим лагранжиан (для удобства):

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

$$L(x, \lambda) = f(x) + \lambda \cdot h(x)$$

$$\frac{\partial L}{\partial x} = 0 \; ; \; \frac{\partial L}{\partial \lambda} = 0$$

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

$$\nabla L = 0$$

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

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

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

Задача наименьших квадратов

Example

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

- $m < n$

Задача наименьших квадратов

Example

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

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

Задача наименьших квадратов

~~$x = \frac{b}{A}$~~

$Ax = b$

ВОТ ОН

$x = ?$

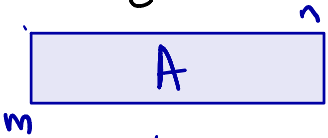
Example

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

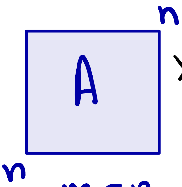
- $m < n$
- $m = n$
- $m > n$

$x = A^{-1} b$

$x + y = 0$

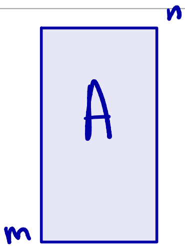


$m < n$



$x^* = A^{-1} b$

$m = n$



Задача наименьших квадратов

A

m

$m < n$

недоопределённая

∞ решений

ВЫБЕРЕМ

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

Example

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

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

Лагранжиан: $L(x, \nu) = \frac{1}{2}\|x\|_2^2 + \nu^T(Ax - b) =$

$$\min \frac{1}{2}\|x\|^2$$
$$Ax = b$$

$$h(x) = 0$$
$$Ax - b = 0$$

$$= \frac{1}{2}\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{m} \nu_i (Ax - b)_i$$

Задача наименьших квадратов

$$L(x,\nu) = \frac{1}{2}\|x\|_2^2 + \nu^T(Ax-b) =$$
$$= \frac{1}{2}\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{m} \nu_i (Ax-b)_i$$

$$dL = d\left(\frac{1}{2}\langle x, x\rangle + \langle \nu, Ax-b\rangle\right) =$$
$$= \frac{1}{2}\cdot 2 \cdot \langle x, dx\rangle + \langle \nu, Adx\rangle =$$
$$= \langle x, dx\rangle + \langle A^T\nu, dx\rangle$$
$$= \langle x + A^T\nu, dx\rangle$$

$$\nabla_x L = 0$$

Example

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

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

$$\Rightarrow \boxed{\nabla_x L = A^T\nu + x} = 0$$

$$\nabla_\nu L = Ax - b = 0$$

$$\begin{cases} A^T\nu + x = 0 \\ Ax - b = 0 \end{cases} \qquad \begin{cases} x = -A^T\nu \\ A(-A^T\nu) - b = 0 \end{cases} \qquad \begin{cases} x = -A^T\left(-(AA^T)^{-1}b\right) \\ -\underbrace{AA^T}_{m\times m}\nu = b \end{cases}$$

$$\boxed{\nu = -(AA^T)^{-1}b}$$

Задача наименьших квадратов

$$\boxed{x^* = A^T(AA^T)^{-1}b}$$

Example

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

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

Задачи с ограничениями-неравенствами

Пример задачи с ограничениями-неравенствами

$f : \mathbb{R}^2 \to \mathbb{R}$

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

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

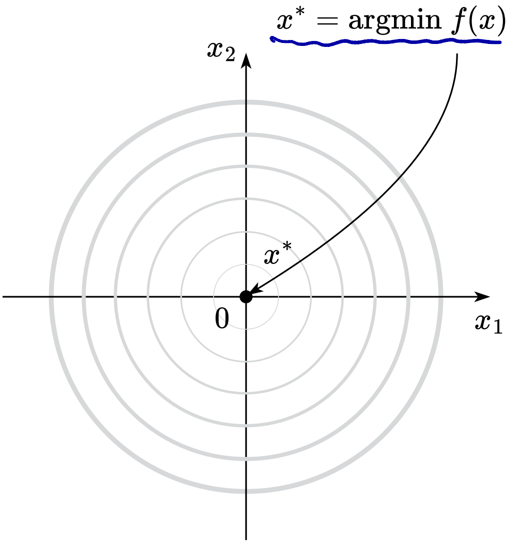
$$
\text{s.t. } g(x) \leq 0
$$

$$
\boxed{\begin{array}{l} \min\limits_{x \in \mathbb{R}^n} f(x) \\ h(x) = 0 \end{array}}
$$

множители Лагранжа

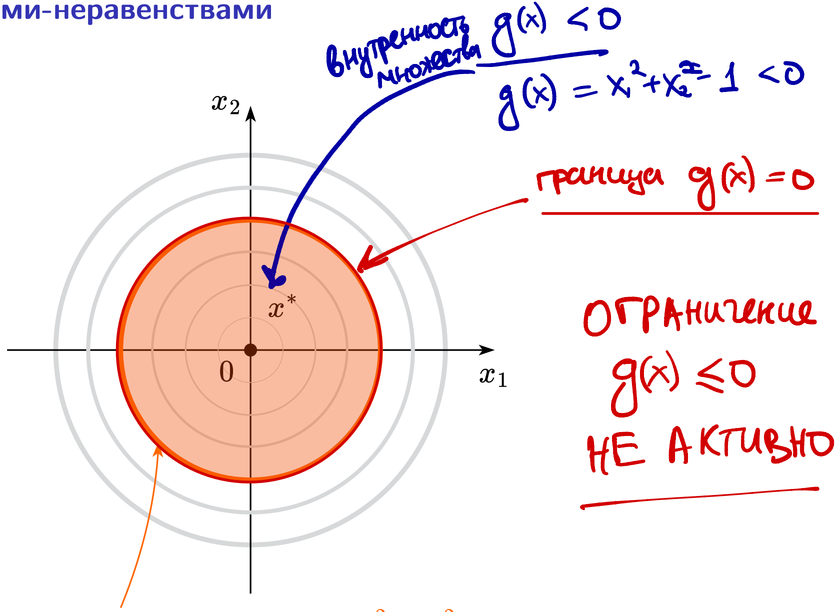
$L(x, \nu) \to \min\limits_{x \in \mathbb{R}^n}$

Задачи с ограничениями-неравенствами



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

Задачи с ограничениями-неравенствами

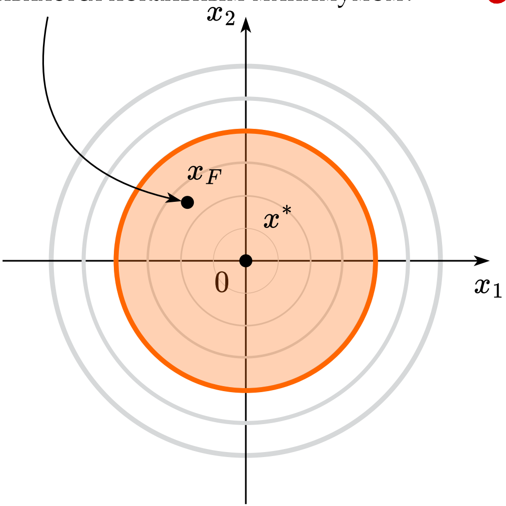


Бюджетное множество $g(x) = x_1^2 + x_2^2 - 1 \leq 0$

Задачи с ограничениями-неравенствами

Как понять, что некоторая допустимая точка является локальным минимумом?

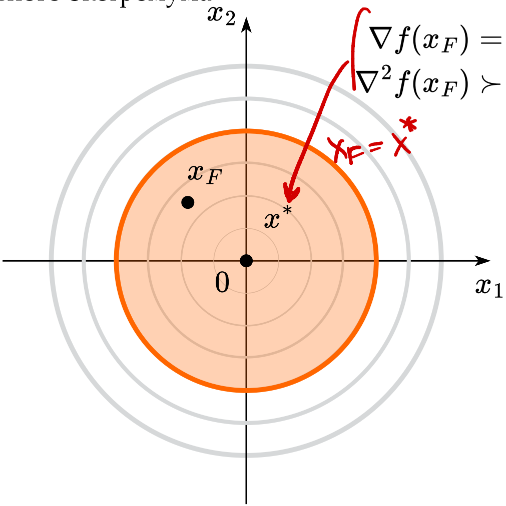
$g(x_F) \leq 0$



Задачи с ограничениями-неравенствами

Просто! Проверим достаточные условия локального экстремума

$$
\begin{cases}
\nabla f(x_F) = 0 \checkmark \\
\nabla^2 f(x_F) \succ 0 \checkmark
\end{cases}
$$



$x_F = x^*$

проверка допустимости

$g(x^*) \leq 0$

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.

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

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

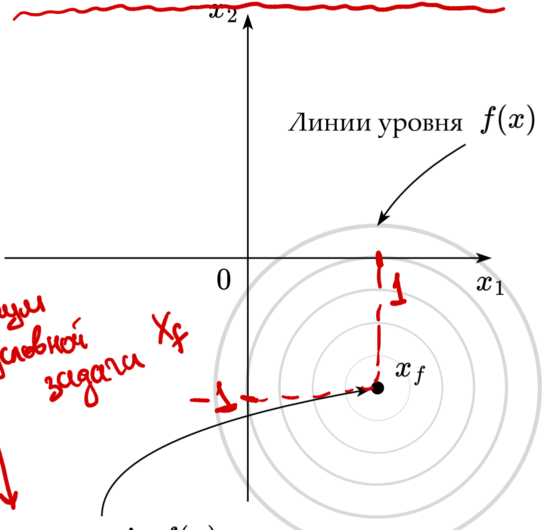
$$\text{s.t. } g(x) \leq 0$$

Задачи с ограничениями-неравенствами

другой пример:

$x_1^c = 1 \qquad x_2^c = -1$

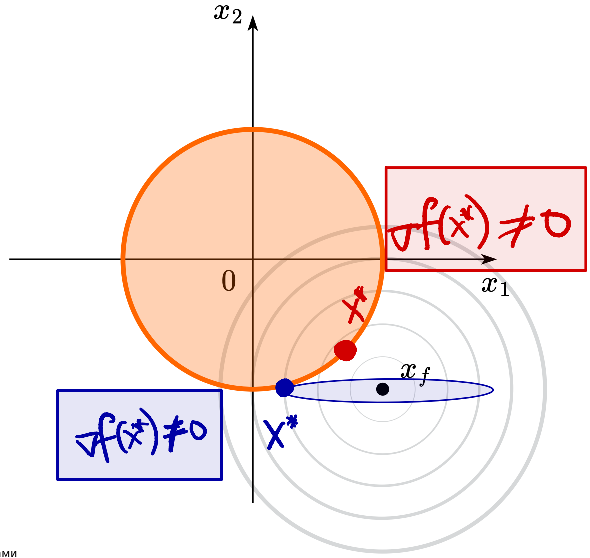
$$f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 = C$$



$$x_f = \text{argmin } f(x)$$

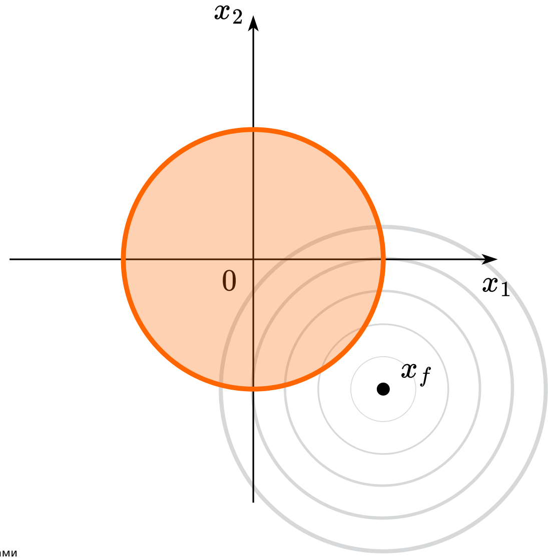
Задачи с ограничениями-неравенствами

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



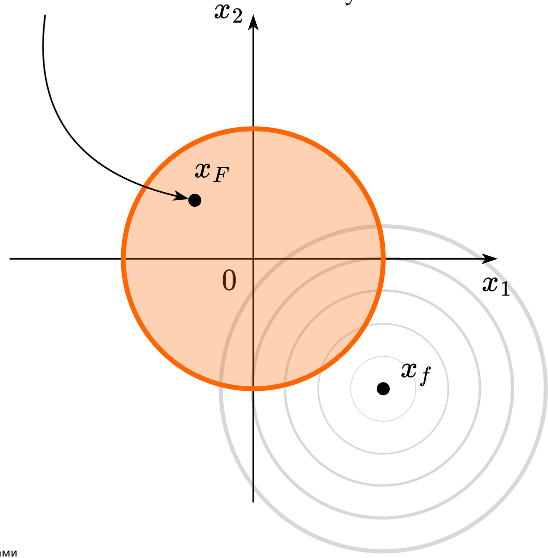
Задачи с ограничениями-неравенствами

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



Задачи с ограничениями-неравенствами

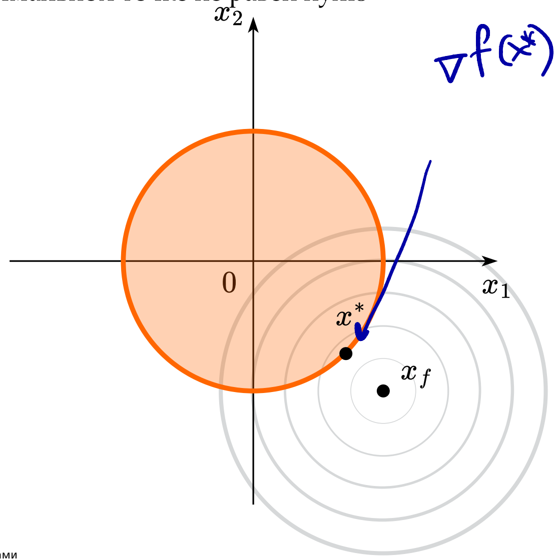
Как понять, что некоторая допустимая точка является локальным минимумом?



Задачи с ограничениями-неравенствами

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

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

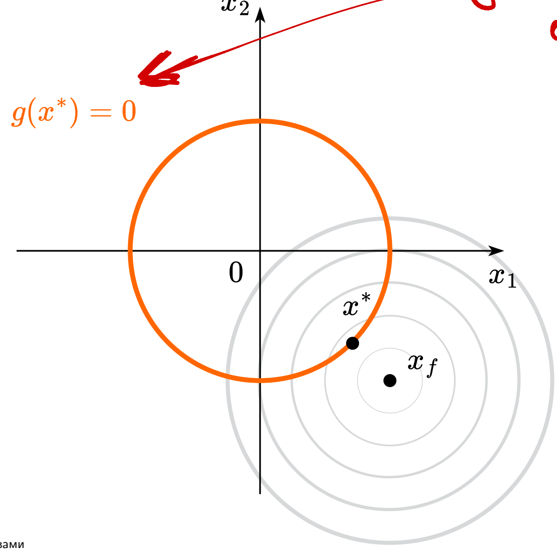


Задачи с ограничениями-неравенствами

Фактически имеем задачу
с ограничением-равенством

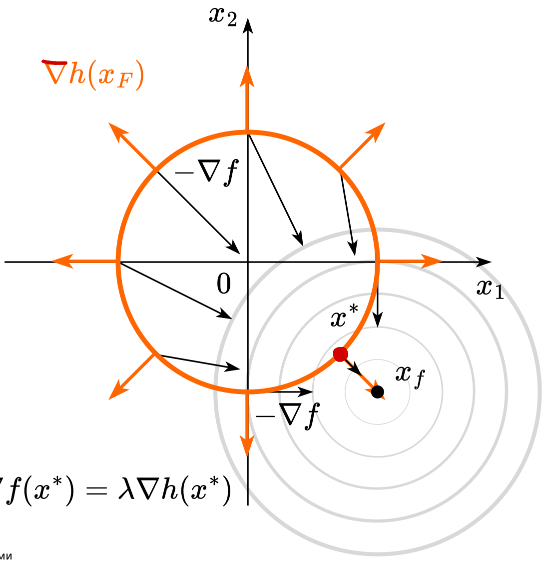
$g(x^*) = 0$

ограничение - равенство!



Задачи с ограничениями-неравенствами

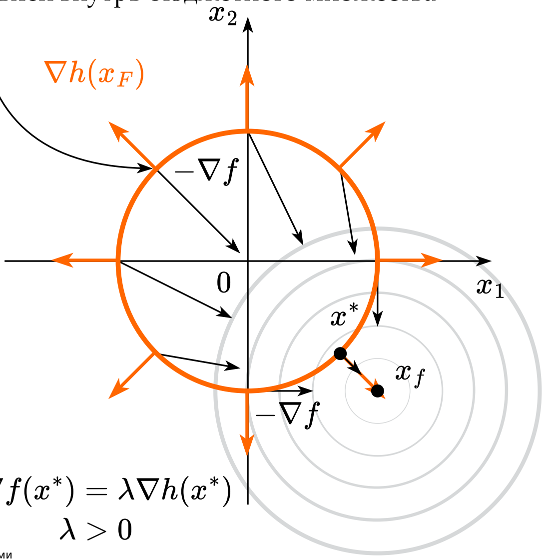
$h(x) = g(x) = x_1^2 + x_2^2 - 1 = 0$



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

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:

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

Два возможных случая:

$g(x) \leq 0$ неактивно: $\underline{g(x^*) < 0}$

- $g(x^*) < 0$

x^* лежит строго внутри бюджетного множества

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: $g(x^*) < 0$

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

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:

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

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

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

Сводится
к безусловной
задаче

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

решение x^* лежит НА ГРАНИЦЕ

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: $\boxed{g(x^*) = 0}$

- $g(x^*) = 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

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

сводится к ограничению-равенству

Лагранжиан для задач с ограничениями-неравенствами

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

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

Определим функцию Лагранжа:

$$\boxed{L(x, \lambda) = f(x) + \lambda g(x)}$$ Лагранжиан

Классические условия 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:

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

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

Определим функцию Лагранжа:

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

Классические условия 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:

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

$$(2)\ \lambda^* \geq 0$$

$$(3)\ \lambda^* g(x^*) = 0$$

$$(4)\ g(x^*) \leq 0$$

бюджетное ограничение

$$\nabla_x \left(f(x) + \lambda g(x) \right) = \nabla f(x) + \lambda \cdot \nabla g(x) = 0$$

$$-\nabla f(x) = \lambda \nabla g(x)$$

Пусть $g(x^*) < 0$ (ограничение НЕ АКТИВНО) $\Rightarrow \lambda^* = 0$

Пусть $g(x^*) = 0$ (ограничение активно) $\lambda^* \geq 0$ (НО ЧАСТО $\lambda^* > 0$)

Общая формулировка

ОБЩАЯ ЗАДАЧА МАТ. ПРОГРАММИРОВАНИЯ

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

Данная формулировка является общей задачей математического программирования.

Решение включает в себя построение лагранжиана:

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

МНОЖИТЕЛИ ЛАГРАНЖА

Необходимые условия

опт. решение исх. задачи

множители Лагранжа

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

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

Необходимые условия

$$L(x,\lambda,\nu) = f_0(x) + \lambda^T f(x) + \nu^T h(x)$$

$\min f_0(x)$

$f_i(x) \le 0$

$h_i(x) = 0$

$\sum\limits_{i=1}^n a_i b_i = a^T b$

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

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$ ← оптимальность
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$ ← $h_i(x) = 0$ бюджетное огран.
- $\lambda_i^* \geq 0, i = 1, \ldots, m$ $\quad f_i(x) = 0 \Rightarrow \ldots$ огр. активно
- $\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m$ $\quad f_i(x) < 0 \Rightarrow \ldots$ огр. неактивно
- $f_i(x^*) \leq 0, i = 1, \ldots, m$ ← бюджетное огран.

условия ККТ

Некоторые условия регулярности

$\min f_0(x)$

$f_i(x) \le 0$

$h_i(x) = 0$

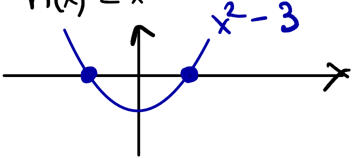
Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера).

- **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 зазор двойственности равен нулю, и условия Каруша—Куна—Таккера становятся необходимыми и достаточными.

пример, когда $h(x) = 0$ — НЕ выпукло е мн-во

$h(x)$ — выпуклая функция

$h(x) = x^2$



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).

- **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 зазор двойственности 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, то ККТ являются необходимыми условиями

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).

- **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 зазор двойственности 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

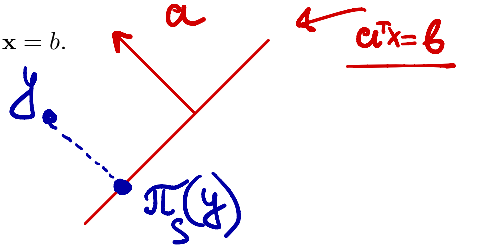
Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера).

- **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 зазор двойственности 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

$h(x) = 0$

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



Найти среди всех $x : a^Tx = b$ ближайшую к y

Решение:

1) $L(x, \lambda) = \frac{1}{2}\|x - y\|_2^2 + \lambda \cdot (a^Tx - b)$

2) ККТ: $\nabla_x L = 0$; $dL = d\left(\frac{1}{2}\langle x-y, x-y\rangle + \lambda(a^Tx - b)\right) =$

$= \frac{1}{2} \cdot 2\langle x - y, d(x-y)\rangle + \lambda \cdot d(\langle a, x\rangle) - 0 =$

$= \langle x - y, dx\rangle + \langle \lambda \cdot a, dx\rangle = \langle x - y + \lambda \cdot a, dx\rangle = 0$

$\Rightarrow \nabla_x L = x - y + \lambda \cdot a$

Пример. Проекция на гиперплоскость

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

$$x - y + \nu \cdot a = 0$$

$$\nabla_\nu L = 0$$

$$a^T x = b$$

$$\begin{cases} x - y + \nu \cdot a = 0 \\ a^T x = b \end{cases} \quad \begin{cases} x = y - \nu \cdot a \\ a^T(y - \nu a) = b \end{cases} \quad \begin{cases} x = y - \nu \cdot a \\ a^T y - \nu \cdot a^T a = b \end{cases}$$

$$\begin{cases} x = y - \nu \cdot a \\ \nu \cdot a^T a = a^T y - b \end{cases} \quad \begin{cases} x = y - \nu a \\ \nu = \dfrac{a^T y - b}{a^T a} \end{cases} \rightarrow x = y - \frac{a^T y - b}{a^T a} \cdot a$$

$$\boxed{x = y + \frac{b - \langle a, y \rangle}{\|a\|_2^2} \cdot a}$$

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:

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

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

Афинно

$f_1(x) \leq 0$

$-x \leq 0$

строго внутр. точка:

$$\left(\frac{1}{n}, \frac{1}{n}, \ldots\ldots \frac{1}{n}\right)$$

Решение:

1) Задача удовл. усл. Слейтера $\Rightarrow$ ККТ - необх. и дост.

2) Лагранжиан: $L(x, \nu, \lambda) = \frac{1}{2}\|x-y\|_2^2 + \nu(1^T x - 1) + \lambda^T(-x)$

3) ККТ: $\bullet\ \nabla_x L = 0$

$$\nabla_x L = x - y + \nu \cdot \underset{\mathbb{R}^n}{\underset{\cap}{1}} - \lambda = 0$$

- $\nabla_\nu L = 0 \Rightarrow 1^T x = 1$
- $x \geq 0$
- $\lambda \geq 0$
- $\lambda_i (-x_i) = 0 \quad , i = 1, \ldots, n$

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 задается 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 задается 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 задается 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 задается 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 задается 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:

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

К сожалению, нет аналитического решения

3n ур-ий + 1 нер-во

2n + 1 неизвестных

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 задается 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 задается 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 задается 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

$$\min_{x_1, x_2 \in \mathbb{R}^2} (x_1 - 1)^2 + (3x_2 + 1)^2$$
$$x_1 \geq 0$$
$$x_2 \geq 0$$

шаг 1:

выполнено условие слейтера

$x_1 = \frac{1}{2}$; $x_2 = \frac{1}{2}$

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

=> ККТ - необх. и дост.

$$L(x, \lambda) = (x_1 - 1)^2 + (3x_2 + 1)^2 - \lambda_1 x_1 - \lambda_2 x_2$$

1	2
—	—
+	—
—	+
+	+

$$\text{ККТ} \begin{cases} \dfrac{\partial L}{\partial x_1} = 0 \\ \dfrac{\partial L}{\partial x_2} = 0 \\ x_1 \geq 0 \\ x_2 \geq 0 \\ \lambda_1 x_1 = 0, \ \lambda_1 \geq 0 \\ \lambda_2 x_2 = 0, \ \lambda_2 \geq 0 \end{cases} \qquad \begin{cases} 2(x_1 - 1) - \lambda_1 = 0 \ ✓ \\ 2(3x_2 + 1) \cdot 3 - \lambda_2 = 0 \ ✓ \\ x_1 \geq 0 \ ✓ \\ x_2 \geq 0 \ ✓ \\ \lambda_1 x_1 = 0, \ \lambda_1 \geq 0 \ ✓ \\ \lambda_2 x_2 = 0, \ \lambda_2 \geq 0 \ ✓ \end{cases}$$

Предположим 1 не активно

$x_1 > 0$

2 активно

$x_2 = 0$

$$6 - \lambda_2 = 0 \Rightarrow \lambda_2 = 6 > 0 \ +$$

$$\lambda_1 = 0 \quad \boxed{x_1 = 1} \quad \boxed{x_2 = 0}$$

ответ

$$\min_{x_1, x_2 \in \mathbb{R}^2} (x_1-1)^2 + (3x_2+1)^2$$

$$x_1 \geq 0$$
$$x_2 \geq 0$$
$$x_1 + x_2 = 5$$

$$L(x_1, x_2, \nu, \lambda_1, \lambda_2) = (x_1-1)^2 + (3x_2+1)^2 + \nu(x_1+x_2-5) - \lambda_1 x_1 - \lambda_2 x$$

$$\text{KKT:} \begin{cases} \dfrac{\partial L}{\partial x_1} = 0 \\ \dfrac{\partial L}{\partial x_2} = 0 \\ \dfrac{\partial L}{\partial \nu} = 0 \\ x_1, x_2 \geq 0 \\ \lambda_1, \lambda_2 \geq 0 \\ \lambda_1 \cdot (-x_1) = 0 \\ \lambda_2 \cdot (-x_2) = 0 \end{cases} \quad \begin{cases} 2(x_1-1) + \nu - \lambda_1 = 0 \\ 6(3x_2+1) + \nu - \lambda_2 = 0 \\ x_1 + x_2 - 5 = 0 \\ x_1 \geq 0 \\ x_2 \geq 0 \\ \lambda_1 \geq 0 \\ \lambda_2 \geq 0 \\ \lambda_1 x_1 = 0 \\ \lambda_2 x_2 = 0 \end{cases} \quad \begin{cases} -2 + \nu - \lambda_1 = 0 \\ 18x_2 + 6 + \nu - \lambda_2 = 0 \\ x_2 - 5 = 0 \\ x_1 \geq 0 \\ x_2 \geq 0 \\ \lambda_1 \geq 0 \\ \lambda_2 \geq 0 \\ \lambda_1 x_1 = 0 \\ \lambda_2 = 0 \end{cases}$$

Пусть 1 АКТИВНО $x_1 = 0$

Пусть 2 НЕАКТИВНО $x_2 > 0$

НЕ УГАДАЛИ

$$\begin{cases} -2 + \nu - \lambda_1 = 0 \\ 18x_2 + 6 + \nu = 0 \\ x_2 = 5 \\ \lambda_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

$$\lambda_1 = \nu - 2 = -98$$ 

$$96 + \nu = 0 \rightarrow \nu = -96$$

$$\min_{x_1, x_2 \in \mathbb{R}^2} (x_1-1)^2 + (3x_2+1)^2$$

$x_1 \geq 0$

$x_2 \geq 0$

$x_1 + x_2 = 5$

$$L(x_1, x_2, \nu, \lambda_1, \lambda_2) = (x_1-1)^2 + (3x_2+1)^2 + \nu(x_1+x_2-5) - \lambda_1 x_1 - \lambda_2 x$$

KKT:
$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 0 \\ \dfrac{\partial L}{\partial x_2} = 0 \\ \dfrac{\partial L}{\partial \nu} = 0 \\ x_1, x_2 \geq 0 \\ \lambda_1, \lambda_2 \geq 0 \\ \lambda_1 \cdot (-x_1) = 0 \\ \lambda_2 \cdot (-x_2) = 0 \end{cases} \qquad \begin{cases} 2(x_1-1) + \nu - \lambda_1 = 0 \\ 6(3x_2+1) + \nu - \lambda_2 = 0 \\ x_1 + x_2 - 5 = 0 \\ x_1 \geq 0 \\ x_2 \geq 0 \\ \lambda_1 \geq 0 \\ \lambda_2 \geq 0 \\ \lambda_1 x_1 = 0 \\ \lambda_2 x_2 = 0 \end{cases}$$

Пусть 1 НЕ АКТИВНО $x_1 > 0$
$\lambda = 0$

Пусть 2 НЕ АКТИВНО $x_2 > 0$
$\lambda_2 = 0$

$$\begin{cases} 2x_1 - 2 + \nu = 0 & (1) \\ 18x_2 + 6 + \nu = 0 & (2) \\ x_1 + x_2 = 5 & (3) \end{cases}$$

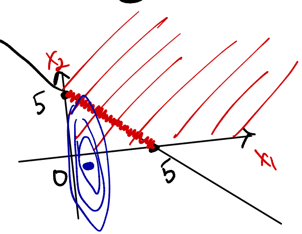
$(2) - (1)$:

$$\begin{cases} 18x_2 - 2x_1 + 8 = 0 & (4) \\ x_1 + x_2 = 5 & (5) \end{cases}$$

$(4) + 2 \cdot (5)$:

$$18x_2 - \cancel{2x_1} + 8 + \cancel{2x_1} + 2x_2 = 10$$

$$20x_2 = 2 \Rightarrow \boxed{x_2 = \frac{1}{10}} > 0$$



(5): $\boxed{x_1 = 5 - x_2 = 4.9}$

Ответ: $\begin{pmatrix} 4.9 \\ 0.1 \end{pmatrix}$

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.