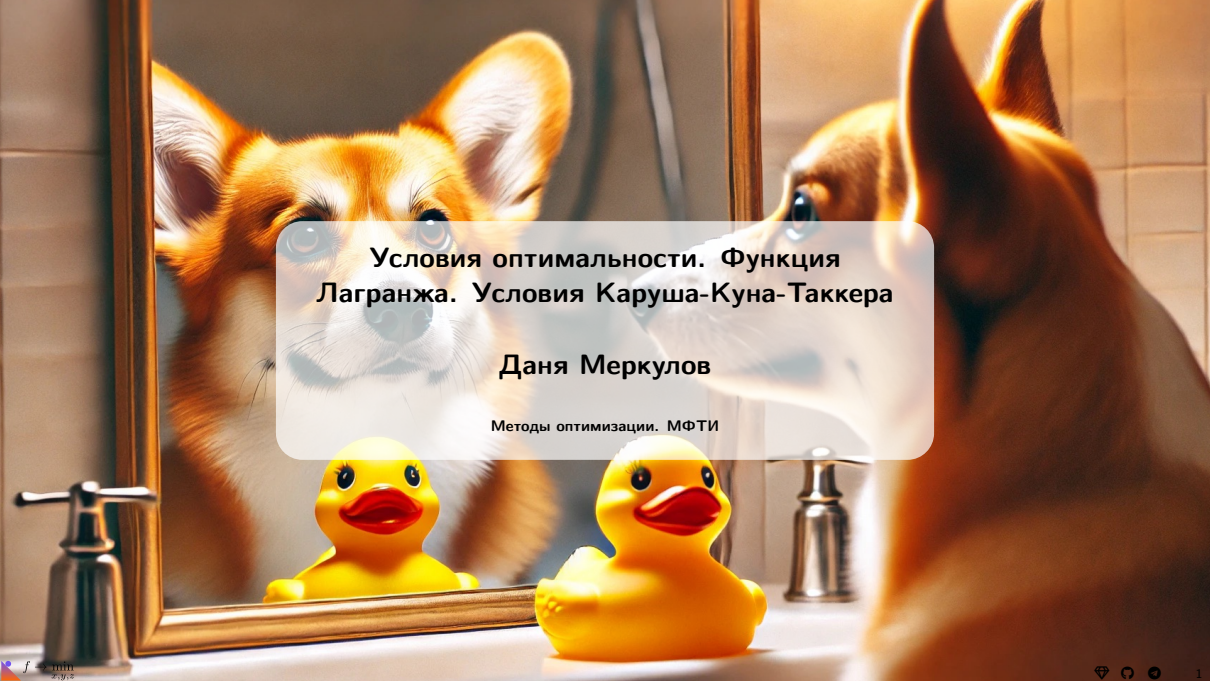


# Условия оптимальности. Функция Лагранжа. Условия Каруша-Куна-Таккера

Даня Меркулов

Методы оптимизации. МФТИ

*В этой работе совершенно отсутствуют какие бы то ни было чертежи. 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"*

Рис. 1: Жозеф Луи Лагранж

# Условия оптимальности

# Теория

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

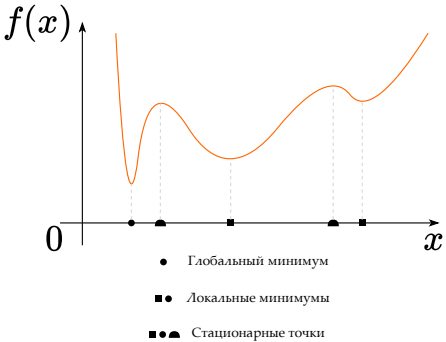



Рис. 2: Иллюстрация различных стационарных (критических) точек

# Теория

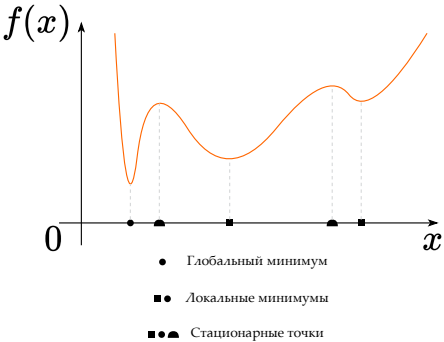



Рис. 2: Иллюстрация различных стационарных (критических) точек

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

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

# Теория

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

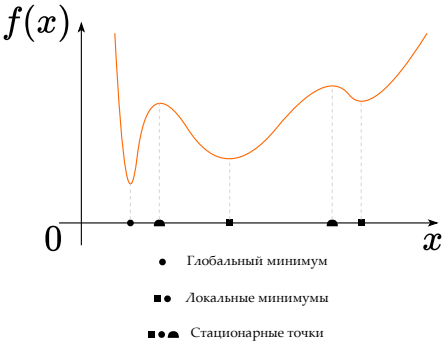



Рис. 2: Иллюстрация различных стационарных (критических) точек

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**: $x^* \in S$.

# Теория

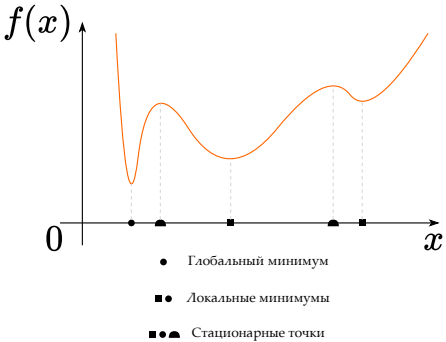



Рис. 2: Иллюстрация различных стационарных (критических) точек

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

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

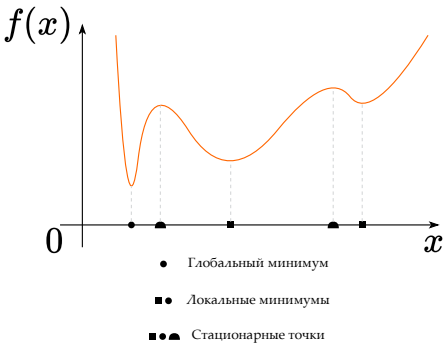



Рис. 2: Иллюстрация различных стационарных (критических) точек

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

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

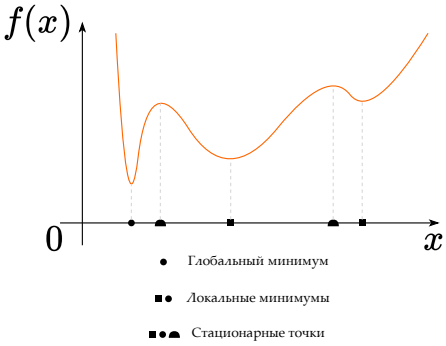



Рис. 2: Иллюстрация различных стационарных (критических) точек

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

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

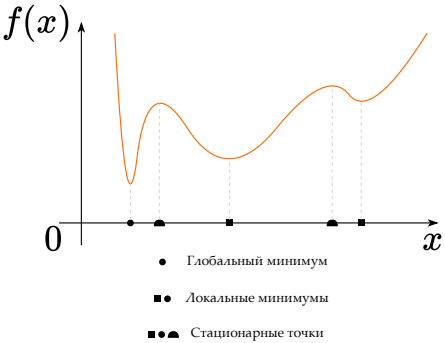



Рис. 2: Иллюстрация различных стационарных (критических) точек

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

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 $t \in (0,1)$.

# Безусловная оптимизация

# 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 $\nabla f$ 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 $\nabla f$ 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 $\nabla f$ 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

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

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 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.

# 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.

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

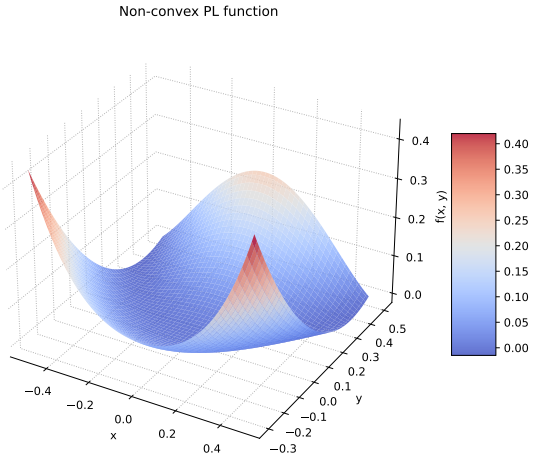
Хотя поверхность не имеет локального минимума в начале координат, ее пересечение с любой вертикальной плоскостью, проходящей через начало координат (плоскость с уравнением $y = mx$ или $x = 0$) является кривой, которая имеет локальный минимум в начале координат. Другими словами, если точка начинает движение 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.



# Условная оптимизация

# 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

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

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

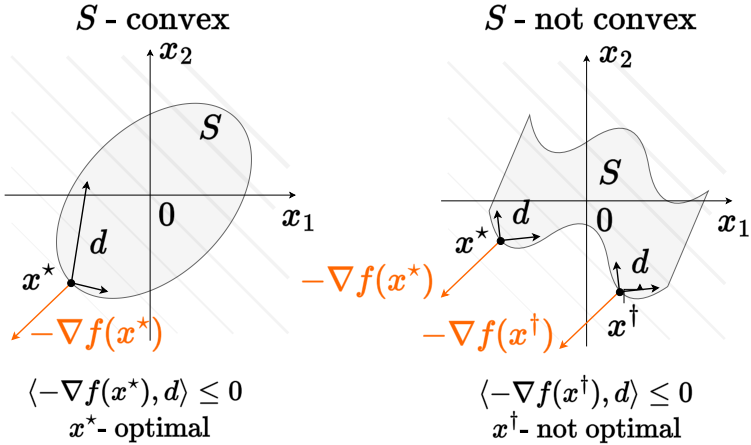



Рис. 4: Общее условие локальной оптимальности первого порядка

# 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:

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

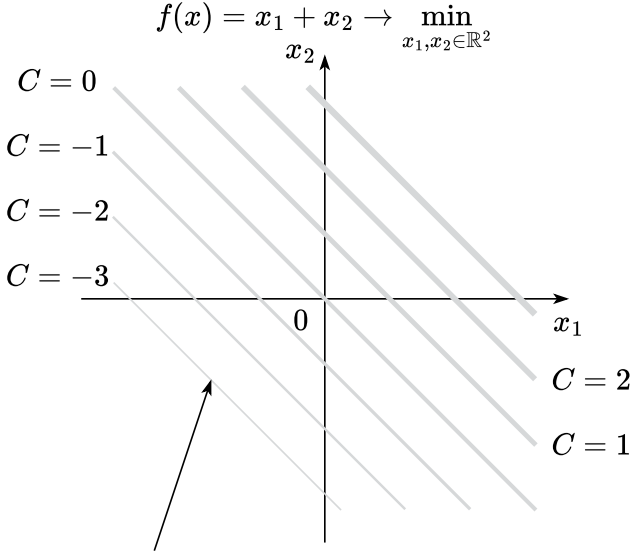
# Задачи с ограничениями-равенствами

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

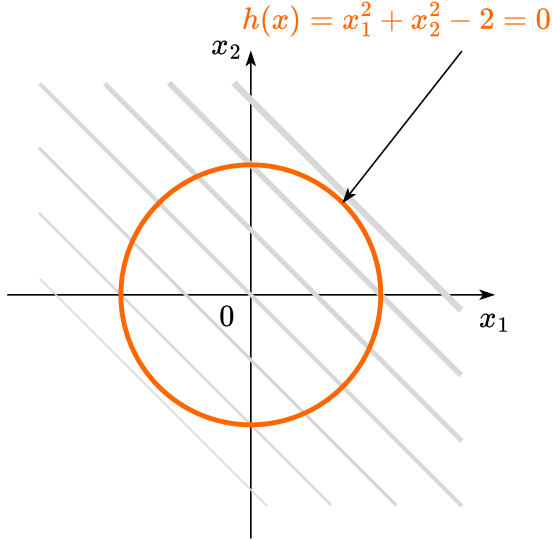
$$\begin{aligned} f(x) &\to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } h(x) &= 0 \end{aligned}$$

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

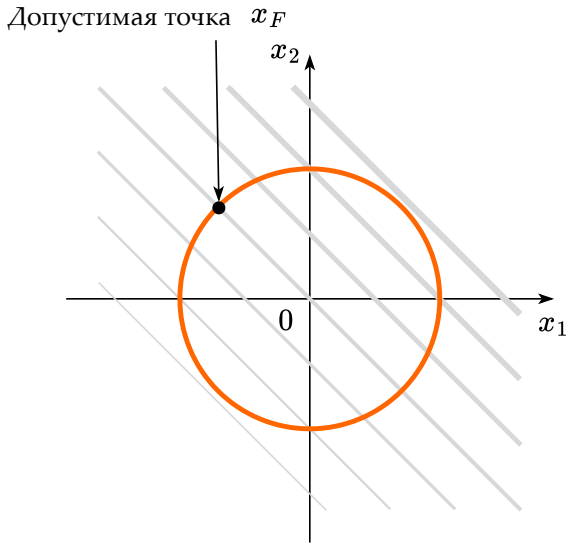
# Задачи с ограничениями-равенствами



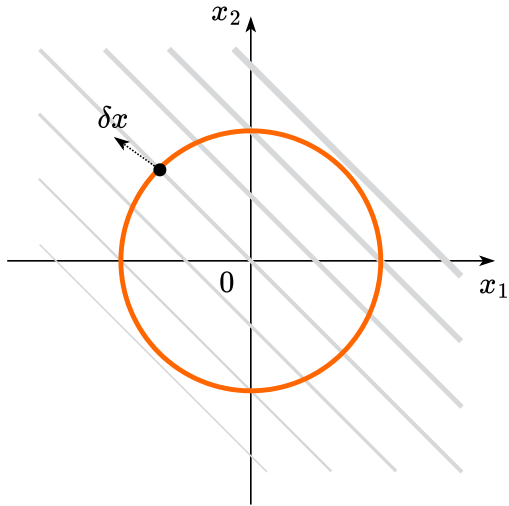
# Задачи с ограничениями-равенствами



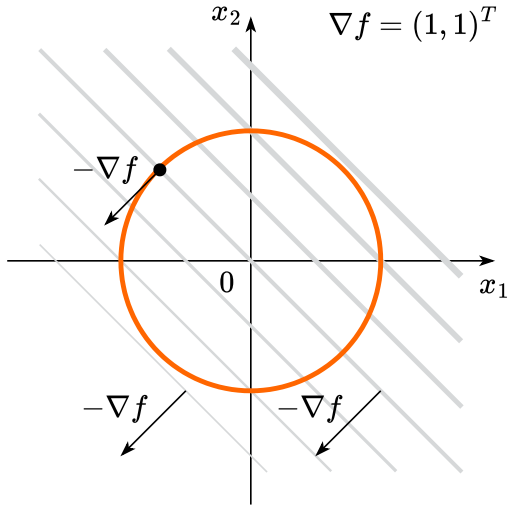
# Задачи с ограничениями-равенствами



# Задачи с ограничениями-равенствами

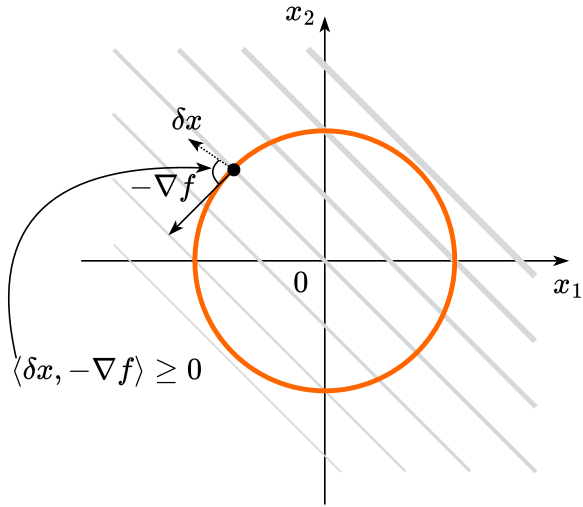


# Задачи с ограничениями-равенствами

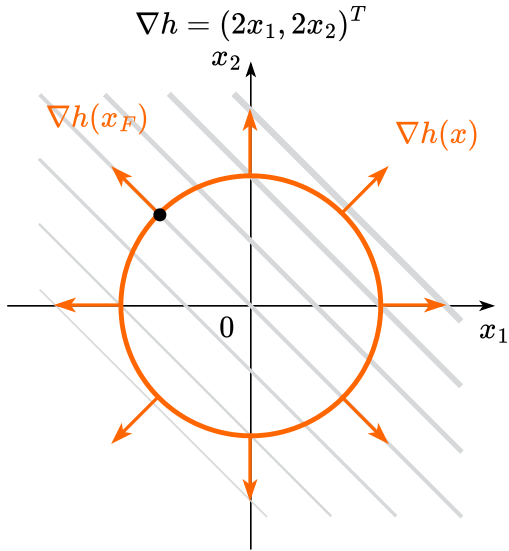


# Задачи с ограничениями-равенствами

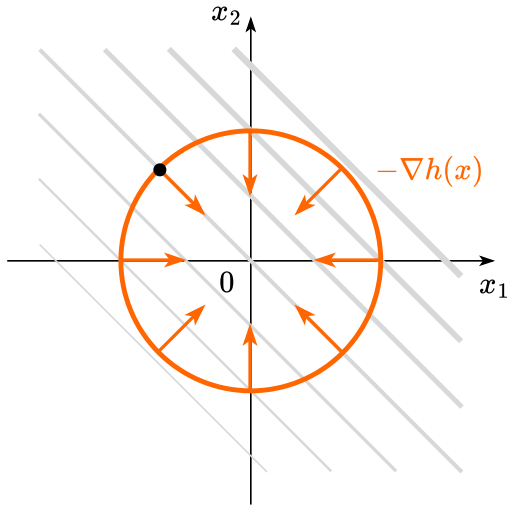
Мы хотим: $f(x_F + \delta x) \leq f(x_F)$



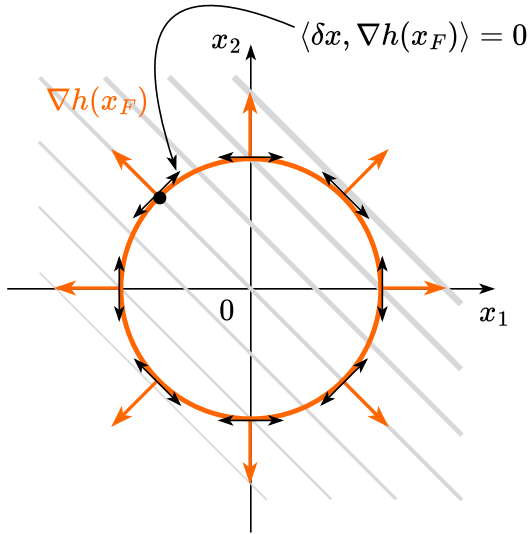
# Задачи с ограничениями-равенствами



# 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:

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

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

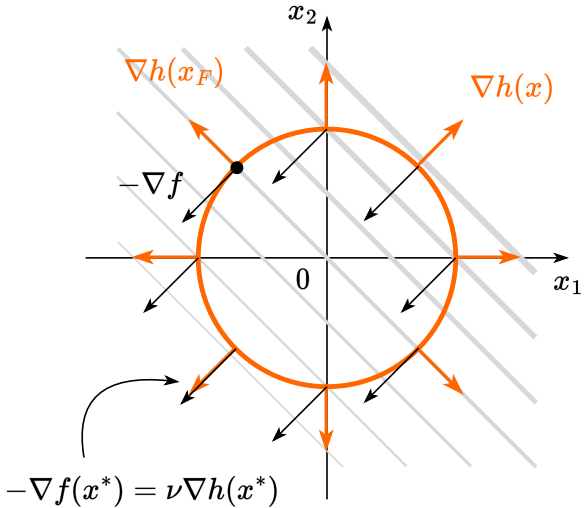
Предположим, что в процессе такого движения мы пришли в точку, где

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

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

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 $\nu^*$:

# 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 $\nu^*$:

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 $\nu^*$:

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 $\nu^*$:

Необходимые условия

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

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

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

# Задачи с ограничениями-равенствами

$$f(x) \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } h_i(x) = 0,\ i = 1, \dots, p \qquad \text{(ECP)}$$

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

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

**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):

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

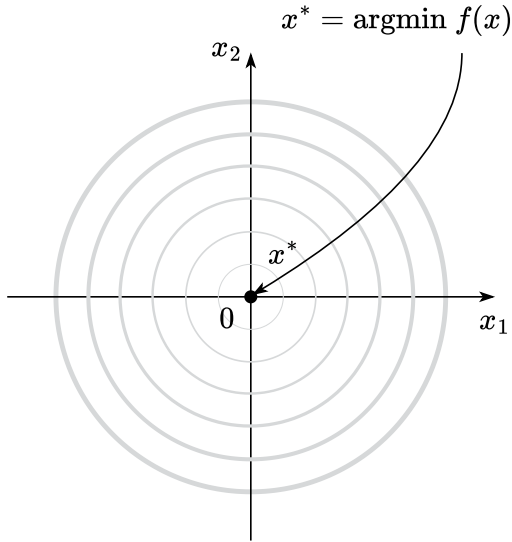
# Задачи с ограничениями-неравенствами

# Пример задачи с ограничениями-неравенствами

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

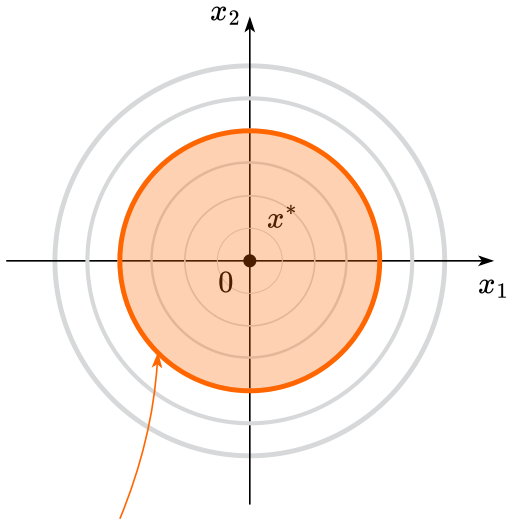
$$f(x) \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } g(x) \leq 0$$

# Задачи с ограничениями-неравенствами




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

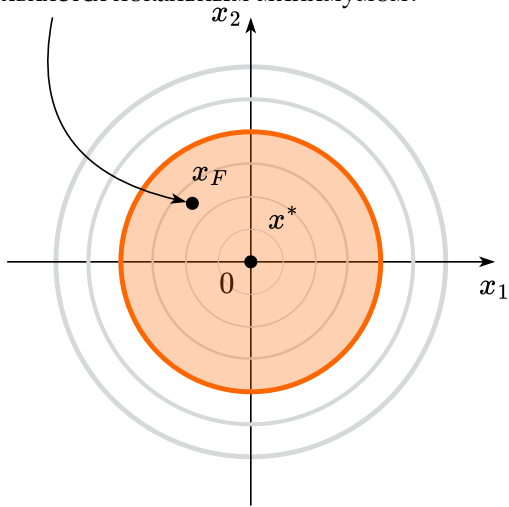
# Задачи с ограничениями-неравенствами




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

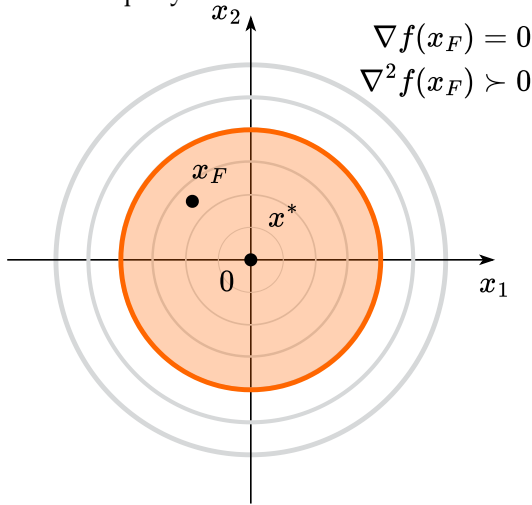
# Задачи с ограничениями-неравенствами

Как понять, что некоторая допустимая точка является локальным минимумом?



# Задачи с ограничениями-неравенствами

Просто! Проверим достаточные условия локального экстремума



# Задачи с ограничениями-неравенствами

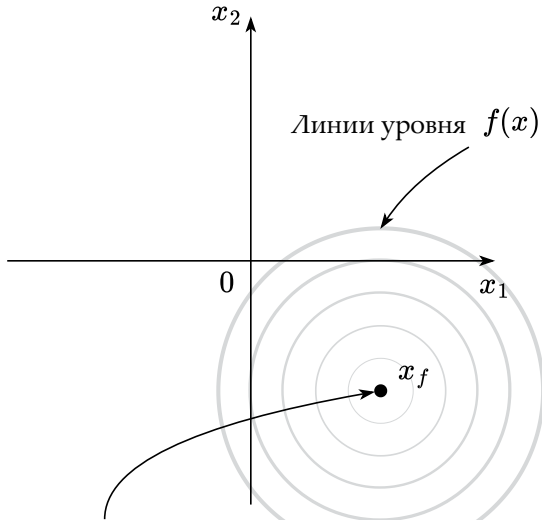
Таким образом, если ограничения типа неравенства неактивны в условной задаче, то мы можем решать задачу без ограничений. Однако так бывает не всегда. Рассмотрим второй простой пример.

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

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

# Задачи с ограничениями-неравенствами

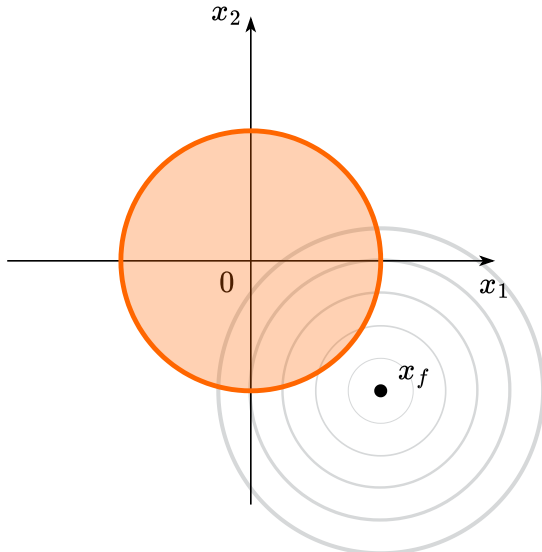
$$f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 = C$$




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

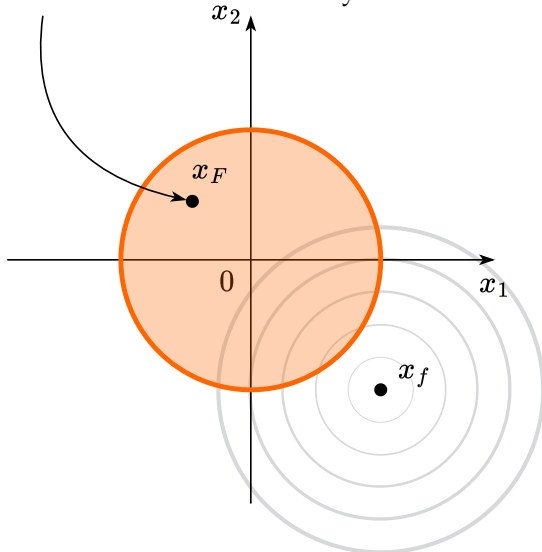
# Задачи с ограничениями-неравенствами

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



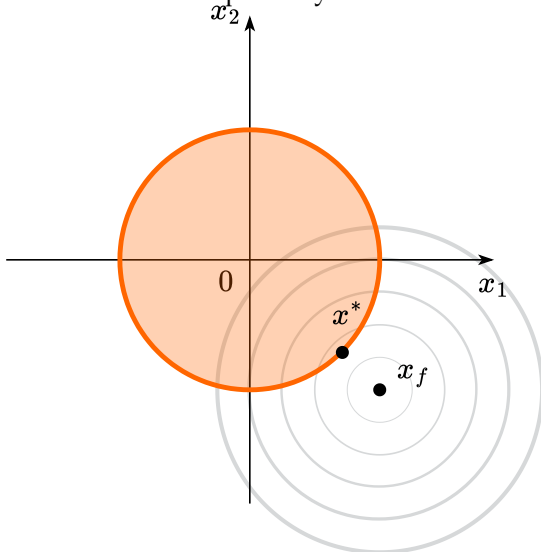
# Задачи с ограничениями-неравенствами

Как понять, что некоторая допустимая точка является локальным минимумом?



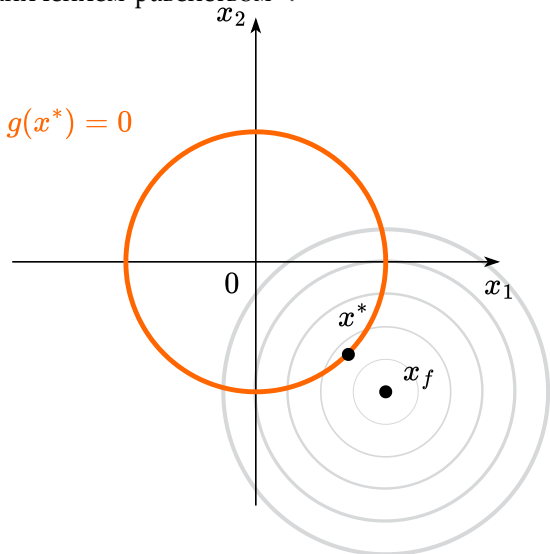
# Задачи с ограничениями-неравенствами

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

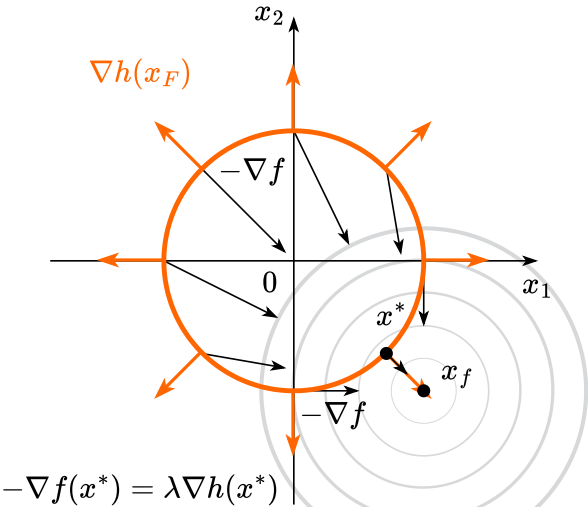


# Задачи с ограничениями-неравенствами

Фактически имеем задачу
с ограничением-равенством 💡



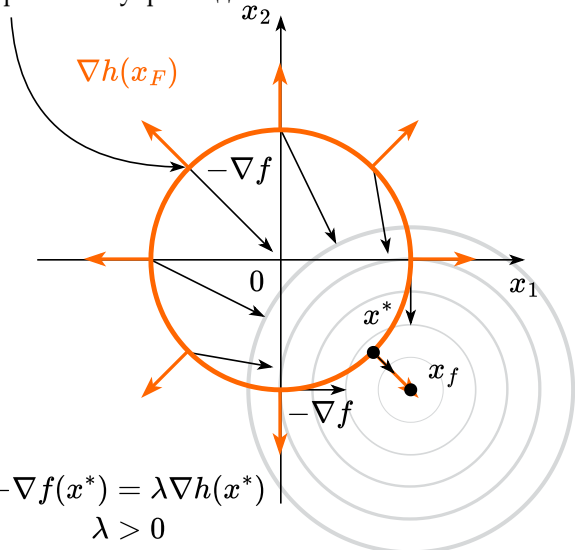
# Задачи с ограничениями-неравенствами




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

# 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:

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

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$

# 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: $g(x^*) < 0$

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

# 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 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 $\lambda^*$ такой, что:

(1) $\nabla_x L(x^*, \lambda^*) = 0$
(2) $\lambda^* \geq 0$
(3) $\lambda^* g(x^*) = 0$
(4) $g(x^*) \leq 0$
(5) $\forall y \in C(x^*) : \langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0$

where $C(x^*) = \{y \ \in \mathbb{R}^n | \nabla f(x^*)^\top y \leq 0$ and $\forall i \in I(x^*) : \nabla g_i(x^*)^T y \leq 0\}$ i
$I(x^*) = \{i \mid g_i(x^*) = 0\}$

# KKT

# Общая формулировка

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

Данная формулировка является общей задачей математического программирования.

Решение включает в себя построение лагранжиана:

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

# Необходимые условия

Пусть $x^*$, $(\lambda^*, \nu^*)$ 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 $x^*$, $(\lambda^*, \nu^*)$ 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 $x^*$, $(\lambda^*, \nu^*)$ 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.

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

# Необходимые условия

Пусть $x^*$, $(\lambda^*, \nu^*)$ 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

Пусть $x^*$, $(\lambda^*, \nu^*)$ 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.

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

# 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, если у нас есть регулярность, мы можем записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределенным* 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 зазор двойственности 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, если у нас есть регулярность, мы можем записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределенным* 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 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.

# 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, если у нас есть регулярность, мы можем записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределенным* 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 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). Кроме того, если у нас есть регулярность, мы можем записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределенным* 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 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.

# Доказательство в простом случае

**Субдифференциальная форма KKT**

Пусть $X$ - линейное нормированное пространство, а $f_j : X \to \mathbb{R}$, $j = 0, 1, \dots, m$ - выпуклая и не принимающая значения $-\infty$ и $\infty$. Рассмотрим задачу

$$f_0(x) \to \min_{x \in X}$$
$$\text{s.t. } f_j(x) \leq 0,\ j = 1, \dots, m$$

Пусть $x^* \in X$ - точка минимума, и функции $f_j$, $j = 0, 1, \dots, m$, непрерывны в $x^*$. Тогда существуют $\lambda_j \geq 0$, $j = 0, 1, \dots, m$ такие, что

$$\sum_{j=0}^{m} \lambda_j = 1,$$

$$\lambda_j f_j(x^*) = 0, \quad j = 1, \dots, m,$$

$$0 \in \sum_{j=0}^{m} \lambda_j \partial f_j(x^*).$$

# Доказательство в простом случае

**Доказательство**

1. Рассмотрим функцию

$$f(x) = \max\{f_0(x) - f_0(x^*), f_1(x), \dots, f_m(x)\}.$$

Точка $x^*$ - точка глобального минимума этой функции. Действительно, если в некоторой точке $x_e \in X$ выполнено неравенство $f(x_e) < 0$, то $f_0(x_e) < f_0(x^*)$ и $f_j(x_e) < 0$, $j = 1, \dots, m$, что противоречит тому, что точка $x^*$ является точкой минимума.

# Доказательство в простом случае

**Доказательство**

1. Рассмотрим функцию

$$f(x) = \max\{f_0(x) - f_0(x^*), f_1(x), \dots, f_m(x)\}.$$

Точка $x^*$ - точка глобального минимума этой функции. Действительно, если в некоторой точке $x_e \in X$ выполнено неравенство $f(x_e) < 0$, то $f_0(x_e) < f_0(x^*)$ и $f_j(x_e) < 0$, $j = 1, \dots, m$, что противоречит тому, что точка $x^*$ является точкой минимума.

2. Из теоремы Ферма в субдифференциальной форме

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

# Доказательство в простом случае

**Доказательство**

1. Рассмотрим функцию

$$f(x) = \max\{f_0(x) - f_0(x^*), f_1(x), \dots, f_m(x)\}.$$

Точка $x^*$ - точка глобального минимума этой функции. Действительно, если в некоторой точке $x_e \in X$ выполнено неравенство $f(x_e) < 0$, то $f_0(x_e) < f_0(x^*)$ и $f_j(x_e) < 0$, $j = 1, \dots, m$, что противоречит тому, что точка $x^*$ является точкой минимума.

2. Из теоремы Ферма в субдифференциальной форме

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

3. Из теоремы Дубовицкого-Милютина имеем

$$\partial f(x^*) = \text{conv} \left( \bigcup_{j \in I} \partial f_j(x^*) \right),$$

где $I = \{0\} \cup \{j : f_j(x^*) = 0, 1 \le j \le m\}$.

# Доказательство в простом случае

**Доказательство**

1. Рассмотрим функцию

$$f(x) = \max\{f_0(x) - f_0(x^*), f_1(x), \dots, f_m(x)\}.$$

Точка $x^*$ - точка глобального минимума этой функции. Действительно, если в некоторой точке $x_e \in X$ выполнено неравенство $f(x_e) < 0$, то $f_0(x_e) < f_0(x^*)$ и $f_j(x_e) < 0$, $j = 1, \dots, m$, что противоречит тому, что точка $x^*$ является точкой минимума.

2. Из теоремы Ферма в субдифференциальной форме

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

3. Из теоремы Дубовицкого-Милютина имеем

$$\partial f(x^*) = \text{conv}\left(\bigcup_{j \in I} \partial f_j(x^*)\right),$$

где $I = \{0\} \cup \{j : f_j(x^*) = 0, 1 \le j \le m\}$.

4. Следовательно, существует $g_j \in \partial f_j(x^*)$, $j \in I$ такая, что

$$\sum_{j \in I} \lambda_j g_j = 0, \quad \sum_{j \in I} \lambda_j = 1, \quad \lambda_j \ge 0, \quad j \in I.$$

Осталось установить $\lambda_j = 0$ for $j \notin I$.

# 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 KKT

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 KKT

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 KKT, 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 KKT

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 KKT, 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 KKT

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 KKT, 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 KKT

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 KKT, 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 KKT

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 KKT, 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 KKT

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 KKT, 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 KKT

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 KKT, 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 $O(n)$.

# Задачи

# Задача 1

**Question**

Функция $f : E \to \mathbb{R}$ определена как

$$f(x) = \ln\left(-Q(x)\right)$$

где $E = \{x \in \mathbb{R}^n : Q(x) < 0\}$ и

$$Q(x) = \frac{1}{2}x^\top Ax + b^\top x + c$$

с $A \in \mathbb{S}^n_{++}$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.
Найдите точку максимума $x^*$ функции $f$.

# Задача 2

**Question**

Найдите явное решение следующей задачи.

$$f(x, y) = x + y \to \min$$
$$\text{s.t. } x^2 + y^2 = 1$$

где $x, y \in \mathbb{R}$.

# Задача 3

**Question**

Найдите явное решение следующей задачи.

$$\langle c, x\rangle + \sum_{i=1}^{n} x_i \log x_i \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } \sum_{i=1}^{n} x_i = 1,$$

где $x \in \mathbb{R}^n_{++}, c \neq 0$.

# Задача 4

**Question**

Пусть $A \in \mathbb{S}^n_{++}, b > 0$ покажите, что:

$$\det(X) \to \max_{X \in \mathbb{S}^n_{++}} \text{s.t.} \langle A, X \rangle \leq b$$

имеет единственное решение и найдите его.

# Задача 5

**Question**

Даны $y \in \{-1, 1\}$, и $X \in \mathbb{R}^{n \times p}$, задача об опорных векторах:

$$\frac{1}{2}||w||_2^2 + C\sum_{i=1}^{n}\xi_i \to \min_{w, w_0, \xi_i}$$

$$\text{s.t. } \xi_i \geq 0, i = 1, \dots, n$$

$$y_i(x_i^T w + w_0) \geq 1 - \xi_i, i = 1, \dots, n$$

найдите условие стационарности KKT.

# Задача 6

**Question**

Покажите, что следующая задача оптимизации с ограничениями имеет единственное решение и найдите его.

$$\langle C^{-1}, X\rangle - \log\det(X) \to \min_{X \in \mathbb{S}^n_{++}} \text{s.t. } a^T X a \leq 1$$

$C \in \mathbb{S}^n_{++}, a \neq 0$
Вы должны избежать явного обратного матрицы $C$ в ответе.

# Задача 7 (БОНУС)

Для некоторых $\Sigma, \Sigma_0 \in \mathbb{S}^n_{++}$ определите расхождение Кульбака-Лейблера между двумя гауссовыми распределениями как:

$$D(\Sigma, \Sigma_0) = \frac{1}{2}(\langle \Sigma_0^{-1}, \Sigma \rangle - \log\det(\Sigma_0^{-1}\Sigma) - n)$$

Теперь пусть $H \in \mathbb{S}^n_{++}$ и $y, x \in \mathbb{R}^n : \langle y, s \rangle > 0$

Мы хотим решить следующую задачу минимизации с ограничениями.

$$\min_{X \in \mathbb{S}^n_{++}} \{D(X^{-1}, H^{-1}) | Xy = s\}$$

Докажите, что она имеет единственное решение и оно равно:

$$(I_n - \frac{sy^T}{y^T s})H(I_n - \frac{ys^T}{y^T s}) + \frac{ss^T}{y^T s}$$

# Задача 8 (БОНУС)

**Question**

Пусть $e_1, \dots, e_n$ будет стандартным базисом в $\mathbb{R}^n$. Покажите, что:

$$\max_{X \in \mathbb{S}^n_{++}} \det(X) : ||Xe_i|| \leq 1 \forall i \in 1, \dots, n$$

имеет единственное решение $I_n$, и выведите неравенство Гильберта:

$$\det(X) \leq \prod_{i=1}^{n} ||Xe_i|| \forall X \in \mathbb{S}^n_{++}$$

# Приложения

# Адверсариальные атаки

Определение: Адверсариальные атаки используются для обмана моделей DL путем добавления небольших возмущений к входным данным. Мы можем сформулировать это как задачу оптимизации с ограничениями, где целью является минимизация/максимизация функции потерь при сохранении возмущения в определенных пределах (ограничение нормы).

Метод FGSM (быстрого знака градиента) является самым простым таким методом, который генерирует adversarial examples путем применения небольшого возмущения в направлении градиента функции потерь. Формально:

$$x' = x + \varepsilon \cdot \text{sgn}(\nabla_x L(x, y)), \text{s.t. } ||x - x'|| \leq \varepsilon$$

Таким образом, мы выполняем градиентный подъем на изображении (== максимизация потерь по отношению к этому изображению).

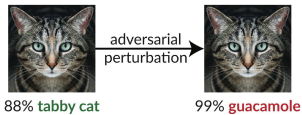



Рис. 26: Иллюстрация

# Ссылки

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе “Элементы статистического обучения” @ КТН.

# Ссылки

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе “Элементы статистического обучения” @ КТН.
- Однострочное доказательство ККТ

# Ссылки

- Лекция по условиям KKT (очень интуитивное объяснение) в курсе “Элементы статистического обучения” @ КТН.
- Однострочное доказательство KKT
- О втором порядке оптимальности для задач оптимизации с ограничениями неравенства

# Ссылки

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

# Ссылки

- Лекция по условиям KKT (очень интуитивное объяснение) в курсе “Элементы статистического обучения” @ КТН.
- Однострочное доказательство KKT
- 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.