

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"

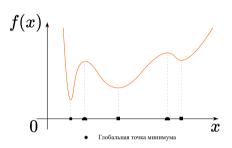


Рис. 1: Жозеф Луи Лагранж



Условия оптимальности



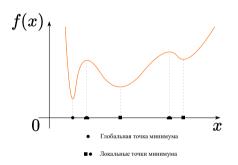


Локальные точки минимума Стационарные точки

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

Рис. 2: Иллюстрация различных стационарных (критических) точек

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

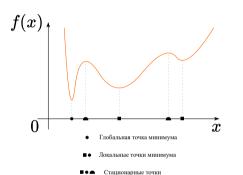


Стационарные точки

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

Рис. 2: Иллюстрация различных стационарных (критических) точек

⊕ 0 0



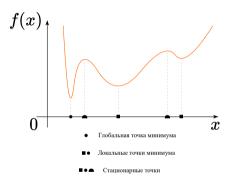
$$f(x) \to \min_{x \in S}$$

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

Рис. 2: Иллюстрация различных стационарных (критических) точек

⊕ ∩ **0**



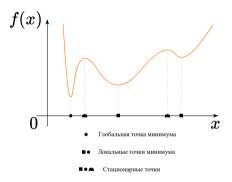
 $f(x) \to \min_{x \in S}$

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

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

Рис. 2: Иллюстрация различных стационарных (критических) точек



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

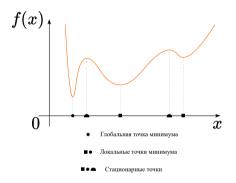
Множество S обычно называется допустимым множеством (или бюджетным множеством).

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 $f(x^*) \le f(x)$ 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 $x \in N \cap S$.

Рис. 2: Иллюстрация различных стационарных (критических) точек

⊕ n ø



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

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

Рис. 2: Иллюстрация различных стационарных (критических) точек

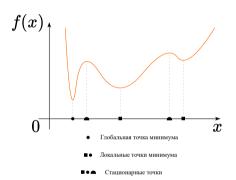


Рис. 2: Иллюстрация различных стационарных (критических) точек

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

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 $f(x^*) \le f(x)$ 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.

⊕ n ø

i 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.

i 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

i 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

і Теорема Тейлора

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

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

i 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

і Теорема Тейлора

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

$$f(x+p) = f(x) + \nabla f(x+tp)^T p$$
 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



1 Необходимое условие оптимальности первого порядка

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

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

1 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

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

🕯 Необходимое условие оптимальности первого порядка

Если x^st - 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

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

🕯 Необходимое условие оптимальности первого порядка

Если x^st - 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

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

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

$$f(x^*+ar t p)=f(x^*)+ar t\, p^T\,
abla f(x^*+tp),$$
 для некоторого $\,t\in(0,ar t)\,$

🕯 Необходимое условие оптимальности первого порядка

Если x^st - 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

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

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

$$f(x^*+ar t p)=f(x^*)+ar t\, p^T\,
abla f(x^*+tp),$$
 для некоторого $\,t\in(0,ar t)\,$

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.

1 Достаточные условия оптимальности второго порядка

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

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

Тогда x^{st} является строгим локальным минимумом функции f.

1 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 поэтому

1 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 поэтому

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

1 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 поэтому

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

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

🗓 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 поэтому

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

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

где $z = x^* + tp$ для некоторого $t \in (0,1)$. Поскольку $z \in B$, то $p^T \nabla^2 f(z) p > 0$, и поэтому $f(x^* + p) > f(x^*)$, что доказывает утверждение.

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.

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.

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

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.

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

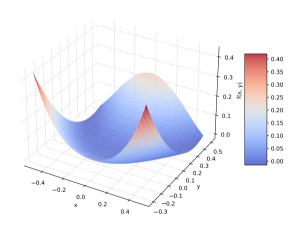
Хотя поверхность не имеет локального минимума в начале координат, ее пересечение с любой вертикальной плоскостью, проходящей через начало координат (плоскость с уравнением u = mx или x = 0) является кривой, которая имеет локальный минимум в начале координат. Другими словами, если точка начинает движение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

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 $u = \sqrt{2}x^2$. приведет к уменьшению значения функции.

Non-convex PL function



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.

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 fнад S, и предположим далее, что fнепрерывно дифференцируема в

окрестности x^* .

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 fнад S, и предположим далее, что fнепрерывно дифференцируема в окрестности x^* .

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

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 fнад S, и предположим далее, что fнепрерывно 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

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

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 fнад S, и предположим далее, что fнепрерывно 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

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

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 fнад S, и предположим далее, что fнепрерывно 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

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

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

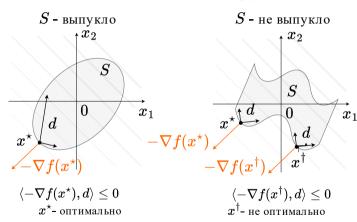


Рис. 4: Общее условие локальной оптимальности первого порядка

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





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 $f(x):S o\mathbb{R}$ выпуклая функция, определённая на выпуклом множестве S, то:

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 $f(x):S o\mathbb{R}$ выпуклая функция, определённая на выпуклом множестве S, то:

• Любой локальный минимум является глобальным.





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 $f(x):S o\mathbb{R}$ выпуклая функция, определённая на выпуклом множестве S, то:

- Любой локальный минимум является глобальным.
- Множество локальных минимумов S^* выпукло.

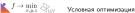




Следует отметить, что в выпуклом случае (то есть при выпуклых f и S) необходимое условие становится достаточным.

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

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



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

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

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

$$\mathrm{s.t.}\ h(x)=0$$

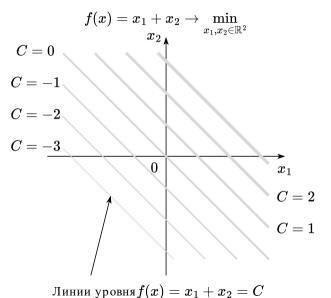


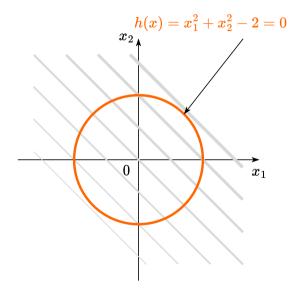


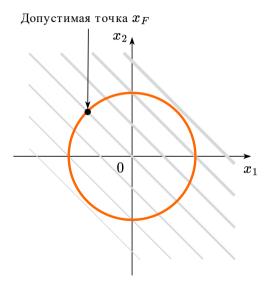
В задачах без ограничений всё довольно интуитивно. В этом разделе мы добавим одно ограничение-равенство, то есть:

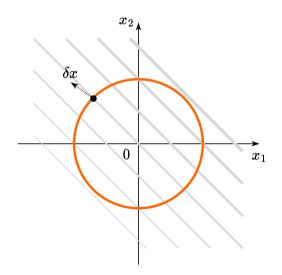
$$f(x) \to \min_{x \in \mathbb{R}^n}$$
 s.t. $h(x) = 0$

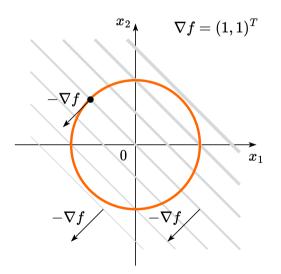
Мы попробуем проиллюстрировать подход к решению этой задачи через простой пример с $f(x)=x_1+x_2$ и $h(x) = x_1^2 + x_2^2 - 2$.

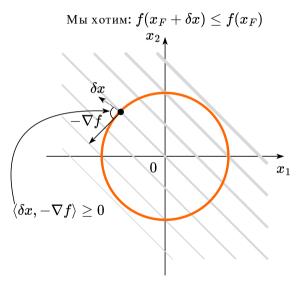


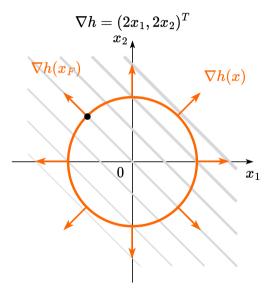


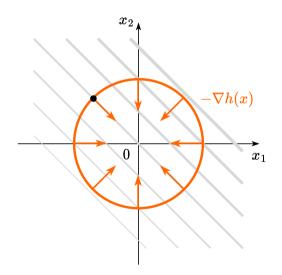


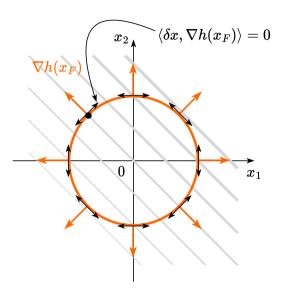












В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции. необходимо обеспечить два условия:



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

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



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

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

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



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



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

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



В общем случае, чтобы двигаться от x_E 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

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

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



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

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

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

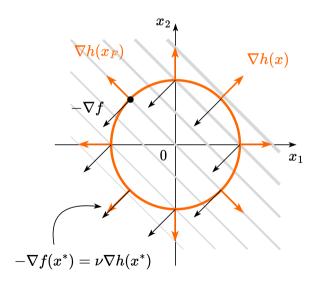
Предположим, что в процессе такого движения мы пришли в точку, где

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

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

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.





Давайте определим лагранжиан (для удобства):

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

Давайте определим лагранжиан (для удобства):

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

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





Давайте определим лагранжиан (для удобства):

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

Если задача $perynnersize{pr$ описанной выше задачи, то существует ν^* :

Необходимые условия

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



Давайте определим лагранжиан (для удобства):

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

Если задача $perynnersize{pr$ описанной выше задачи, то существует ν^* :

Необходимые условия

$$abla_x L(x^*,
u^*) = 0$$
 это мы уже написали выше

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



Давайте определим лагранжиан (для удобства):

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

Если задача $perynnersize{pr$ описанной выше задачи, то существует ν^* :

Необходимые условия

$$abla_x L(x^*,
u^*) = 0$$
 это мы уже написали выше

$$abla_{
u}L(x^*,
u^*)=0$$
 бюджетное ограничение

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



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

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

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

i Example

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

• *m* < *n*

Задача наименьших квадратов

i Example

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

- *m* < *n*
- \bullet m=n



Задача наименьших квадратов

i Example

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

- *m* < *n*
- \bullet m=n
- m > n





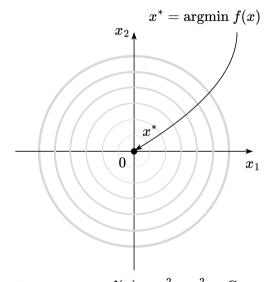
Пример задачи с ограничениями-неравенствами

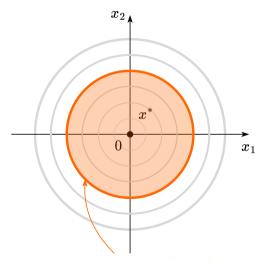
$$f(x) = x_1^2 + x_2^2$$
 $g(x) = x_1^2 + x_2^2 - 1$

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

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

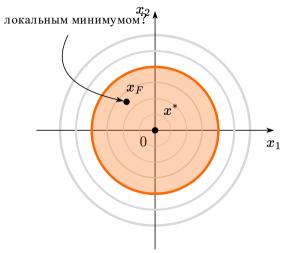




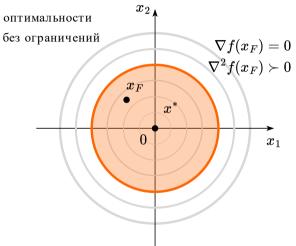


Допустимая область $g(x) = x_1^2 + x_2^2 - 1 \leq 0$

Как понять, что допустимая точка является



В этом случае легко - проверьте условия





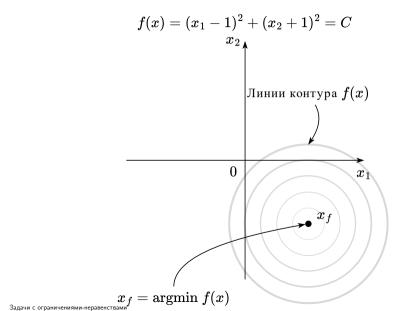
Таким образом, если ограничения типа неравенства неактивны в условной задаче, то мы можем решать задачу без ограничений. Однако так бывает не всегда. Рассмотрим второй простой пример.

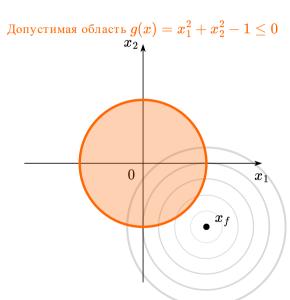
$$f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 \quad g(x) = x_1^2 + x_2^2 - 1$$

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

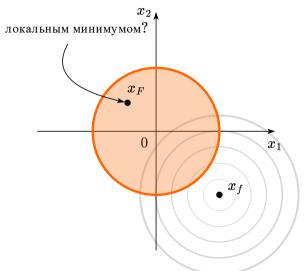
$$\text{s.t. } g(x) \leq 0$$





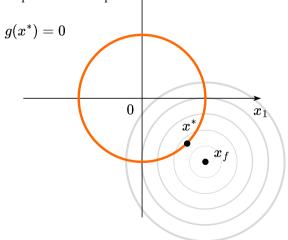


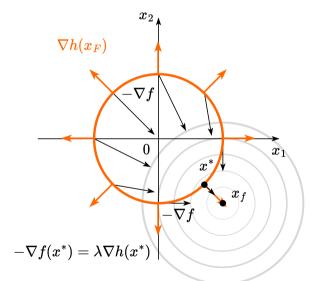
Как понять что допустимая точка является



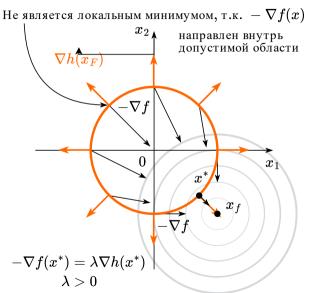
Не так просто в данном случае! Градиент $\neq 0$ в точке локального x_{2} минимума 😕

Фактически тут мы имеем задачу x_2 с ограничениями-равенствами!











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. $g(x^*) < 0$

• $g(x^*) < 0$

Итак, у нас есть задача:

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

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

$$g(x) \leq 0$$
 неактивно. $g(x^*) < 0$

•
$$g(x^*) < 0$$

•
$$\nabla f(x^*) = 0$$

Итак, у нас есть задача:

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

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

$$g(x) \leq 0$$
 неактивно. $g(x^*) < 0$

$$\overline{g}(x^*) < 0$$

•
$$\nabla^2 \hat{f}(x^*) > 0$$



Итак, у нас есть задача:

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

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

$$g(x) \leq 0$$
 неактивно. $g(x^*) < 0$

$$\overline{g}(x^*) < 0$$

•
$$\nabla^2 \hat{f}(x^*) > 0$$



Итак, у нас есть задача:

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

s.t. $g(x) \leq 0$

$$g(x) \leq 0$$
 неактивно. $g(x^*) < 0$

•
$$q(x^*) < 0$$

•
$$\nabla f(x^*) = 0$$

$$\nabla f(x') = 0$$

•
$$\nabla^2 \hat{f}(x^*) > 0$$

$$g(x) \leq 0$$
 активно. $g(x^*) = 0$

$$g(x^*) = 0$$



Итак, у нас есть задача:

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

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

$$g(x) \leq 0$$
 неактивно. $g(x^*) < 0$

$$q(x^*) < 0$$

•
$$\nabla f(x^*) = 0$$

•
$$\nabla^2 f(x^*) > 0$$

$$g(x) \leq 0$$
 активно. $g(x^*) = 0$

•
$$g(x^*) = 0$$

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





Лагранжиан для задач с ограничениями-неравенствами

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

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

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

Определим функцию Лагранжа:

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

Классические условия 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 $\,$ Если x^* является локальным минимумом для описанной выше задачи, можем записать общие условия для задачи:

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

s.t. $g(x) \leq 0$

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

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

то существует единственный множитель Лагранжа λ^* такой, что:

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

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

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

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



Общая формулировка

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

Данная формулировка является общей задачей математического программирования.

Решение включает в себя построение лагранжиана:

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

Пусть x^* . (λ^*, ν^*) является решением 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.

• $\nabla_{m}L(x^{*},\lambda^{*},\nu^{*})=0$



Пусть x^* . (λ^*, ν^*) является решением 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.

- $\nabla_{m}L(x^{*},\lambda^{*},\nu^{*})=0$
- $\nabla_{\cdot \cdot} L(x^*, \lambda^*, \nu^*) = 0$



Пусть x^* . (λ^*, ν^*) является решением 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.

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



Пусть x^* . (λ^*, ν^*) является решением 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.

- $\nabla_{m}L(x^{*},\lambda^{*},\nu^{*})=0$
- $\nabla_{\cdot \cdot} L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda_i^* > 0, i = 1, ..., m$
- $\lambda_i^* f_i(x^*) = 0, i = 1, ..., m$



Пусть x^* , (λ^*, ν^*) является решением 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.

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

•
$$\nabla_{\nu} L(x^*, \lambda^*, \nu^*) = 0$$

•
$$\lambda_i^* \geq 0, i = 1, ..., m$$

•
$$\lambda_i^* f_i(x^*) = 0, i = 1, ..., m$$

•
$$f_i(x^*) < 0, i = 1, ..., m$$





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

ullet 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 $f_{\epsilon}(x) < 0$ (существует строго допустимая точка), то зазор двойственности равен нулю, и условия Каруша—Куна—Таккера становятся необходимыми и достаточными.



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

- ullet **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 $f_{\epsilon}(x) < 0$ (существует строго допустимая точка), то зазор двойственности равен нулю, и условия Каруша—Куна—Таккера становятся необходимыми и достаточными.
- ullet **Условие линейной квалификации ограничений.** Если f_i и h_i являются аффинными функциями, то никаких других условий не требуется.



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

- ullet **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 $f_{\epsilon}(x) < 0$ (существует строго допустимая точка), то зазор двойственности равен нулю, и условия Каруша—Куна—Таккера становятся необходимыми и достаточными.
- ullet **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^* .



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

- ullet **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 $f_{\epsilon}(x) < 0$ (существует строго допустимая точка), то зазор двойственности равен нулю, и условия Каруша—Куна—Таккера становятся необходимыми и достаточными.
- ullet **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.



$$\min \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

$$\min \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Решение

Лагранжиан:

 $\min \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T \mathbf{x} = b.$

Решение

Лагранжиан:

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

 $\min \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T \mathbf{x} = b.$

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} :

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

 $\min \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T \mathbf{x} = b.$

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} :

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

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

 $\min \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T \mathbf{x} = b.$

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} :

$$\begin{split} \frac{\partial L}{\partial \mathbf{x}} &= \mathbf{x} - \mathbf{y} + \nu \mathbf{a} = 0, \qquad \mathbf{x} = \mathbf{y} - \nu \mathbf{a} \\ \mathbf{a}^T \mathbf{x} &= \mathbf{a}^T \mathbf{y} - \nu \mathbf{a}^T \mathbf{a} \qquad \nu = \frac{\mathbf{a}^T \mathbf{y} - b}{\|\mathbf{a}\|^2} \end{split}$$

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

$$\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

$$\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

Условия ККТ

Лагранжиан задается следующим образом:

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

$$\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

Условия ККТ

Лагранжиан задается следующим образом:

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

Взяв производную L по x_i и записав ККТ, мы получаем: • $\frac{\partial L}{\partial x_i}=x_i-y_i-\lambda_i+\nu=0$

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

$$\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

Условия ККТ

Лагранжиан задается следующим образом:

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

Взяв производную L по x_i и записав ККТ, мы получаем: • $\frac{\partial L}{\partial x_i}=x_i-y_i-\lambda_i+\nu=0$

- $\lambda_i \dot{x}_i = 0$

$$\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

Условия ККТ

Лагранжиан задается следующим образом:

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

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

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

$$\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

Условия ККТ

Лагранжиан задается следующим образом:

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

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

- $\frac{\partial L}{\partial x_i} = x_i y_i \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\begin{array}{ll} \bullet \ \lambda_i \geq 0 \\ \bullet \ x^\top 1 = 1, \quad x > 0 \end{array}$

$$\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

Условия ККТ

Лагранжиан задается следующим образом:

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

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

- $\frac{\partial L}{\partial x_i} = x_i y_i \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\begin{array}{ll} \bullet \ \lambda_i \geq 0 \\ \bullet \ x^\top 1 = 1, \quad x > 0 \end{array}$

$$\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

Условия ККТ

Лагранжиан задается следующим образом:

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

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

- $\frac{\partial L}{\partial x} = x_i y_i \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\begin{array}{ll} \bullet & \lambda_i \geq 0 \\ \bullet & x^\top 1 = 1, \quad x > 0 \end{array}$

i Question

Решите систему выше за $O(n \log n)$.



$$\min \frac{1}{2}\|x-y\|^2, \quad \text{s.t.} \quad x^\top 1 = 1, \quad x \geq 0.$$

Условия ККТ

Лагранжиан задается следующим образом:

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

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

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

$$\lambda_i x_i = 0$$

$$\lambda_{\underline{i}} \geq 0$$

$$\begin{array}{ll} \bullet & \lambda_i \geq 0 \\ \bullet & x^\top 1 = 1, \quad x \geq 0 \end{array}$$

i Question

Решите систему выше за $O(n \log n)$.

Решите систему выше за O(n).

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





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





- 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





- 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





- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе "Элементы статистического обучения" 0 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.



