



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

Даня Меркулов

Оптимизация для всех! ЦУ

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

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



Рис. 2: Иллюстрация различных стационарных (критических) точек

Теория

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

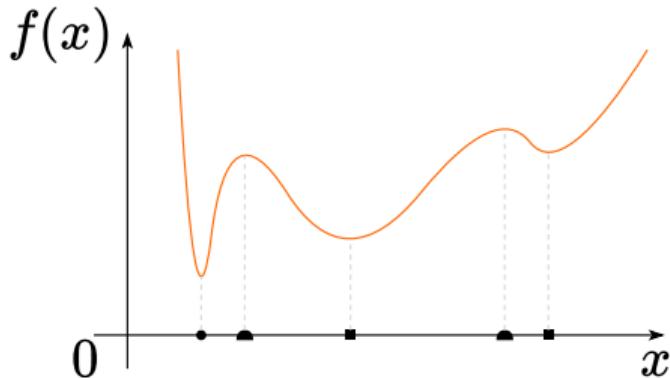
Множество S обычно называется **допустимым множеством** (или **бюджетным множеством**).



Рис. 2: Иллюстрация различных стационарных (критических) точек

Теория

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



- Глобальный минимум
- ● Локальные минимумы
- ● ■ Стационарные точки

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

Теория

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



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 $x \in S$.

Рис. 2: Иллюстрация различных стационарных (критических) точек

Теория

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



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 $x \in N \cap S$.

Рис. 2: Иллюстрация различных стационарных (критических) точек

Теория

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



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

- 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

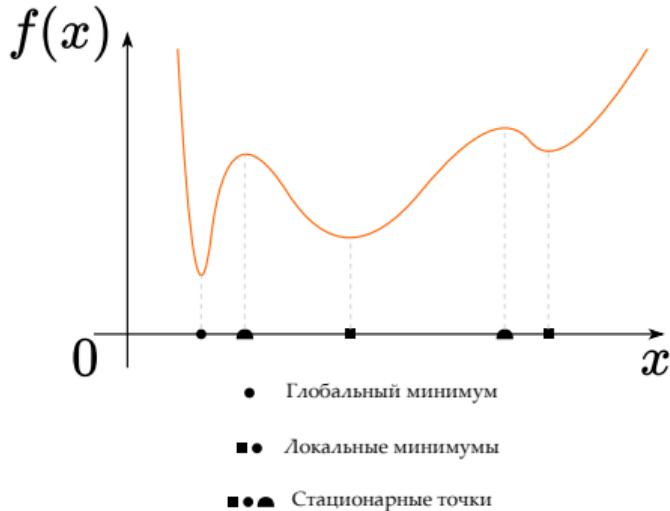


Рис. 2: Иллюстрация различных стационарных (критических) точек

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

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

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.

Теорема Вейерштрасса об экстремальных значениях

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.

GOOD NEWS EVERYONE!



Рис. 3: Многие практические задачи теоретически разрешимы

Теорема Вейерштрасса об экстремальных значениях

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.

GOOD NEWS EVERYONE!



i Теорема Тейлора

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

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

Рис. 3: Многие практические задачи теоретически разрешимы

Теорема Вейерштрасса об экстремальных значениях

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.

GOOD NEWS EVERYONE!



Рис. 3: Многие практические задачи теоретически разрешимы

i Теорема Тейлора

Пусть $f : \mathbb{R}^n \rightarrow \mathbb{R}$ - 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

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

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 .

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.

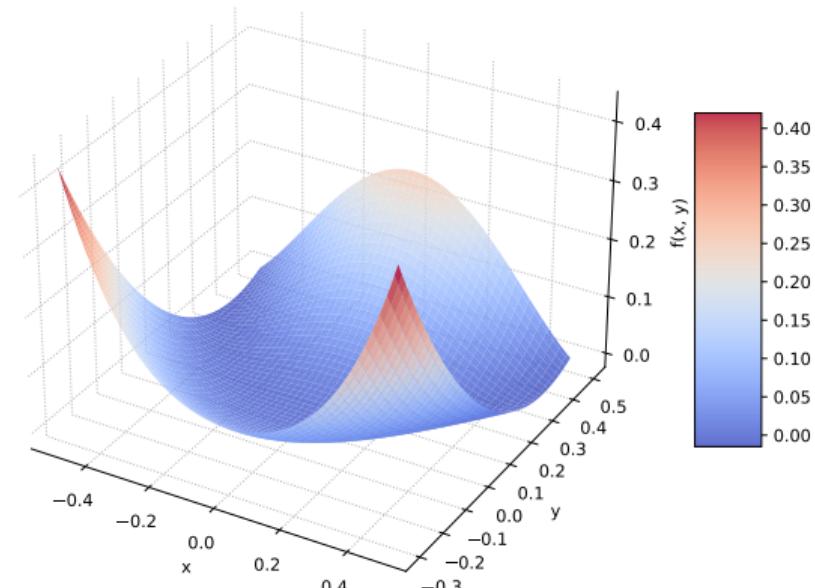
Контрпример Пеано

Заметим, что если $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succeq 0$ (гессиан положительно полуопределен), 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.

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 .

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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Предположим, что $x^* \in S$ является точкой локального минимума для f над S , и предположим далее, что f непрерывно дифференцируема в окрестности x^* .

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

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



Рис. 4: Общее условие локальной оптимальности первого порядка

Выпуклый случай

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

Выпуклый случай

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

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

- Любой локальный минимум является глобальным.
- Множество локальных 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 $f(x) : S \rightarrow \mathbb{R}$ — 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: $S^* = \{x^*\}$.

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:

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

Задачи с ограничениями-равенствами

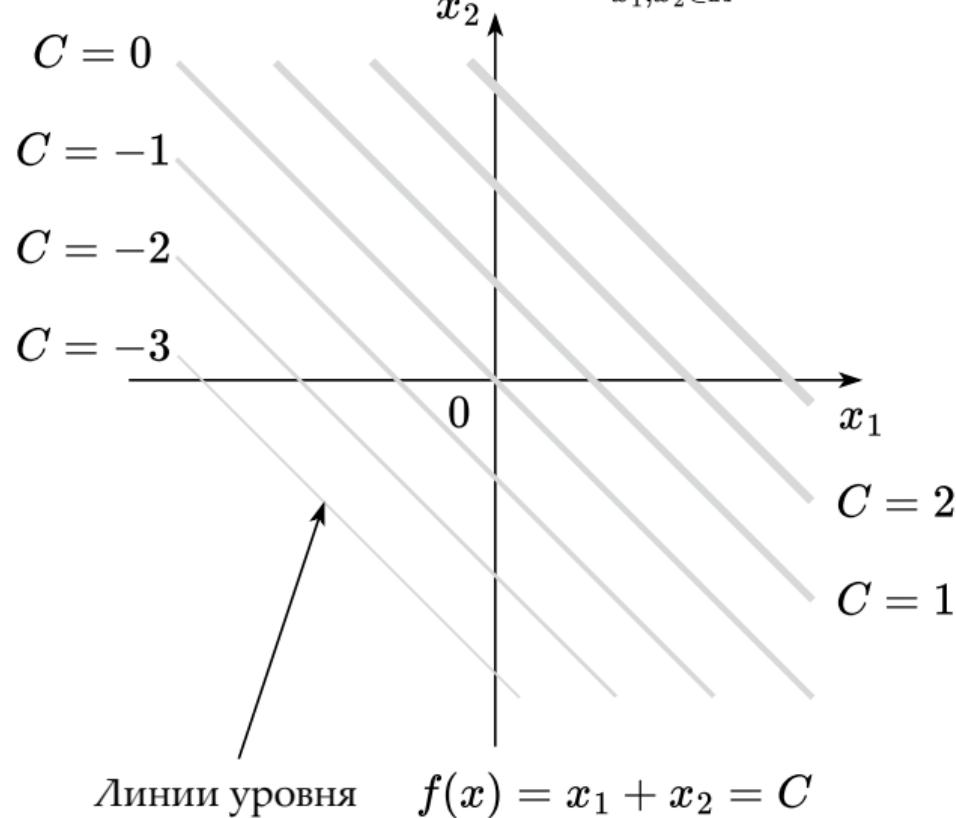
В задачах без ограничений всё довольно интуитивно. В этом разделе мы добавим одно ограничение-равенство, то есть:

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

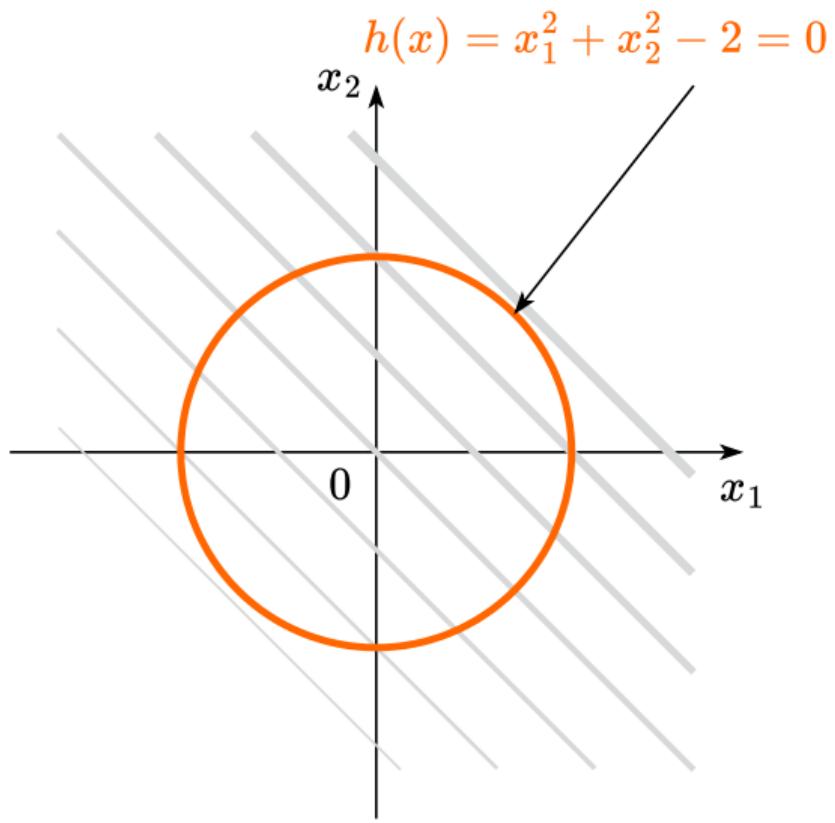
Мы попробуем проиллюстрировать подход к решению этой задачи через простой пример с $f(x) = x_1 + x_2$ и $h(x) = x_1^2 + x_2^2 - 2$.

Задачи с ограничениями-равенствами

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

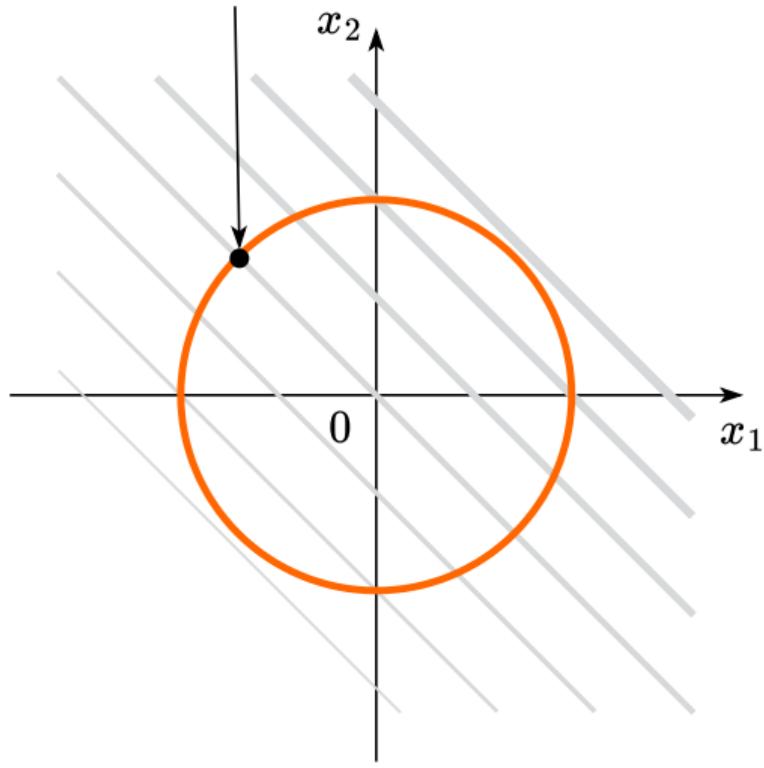


Задачи с ограничениями-равенствами

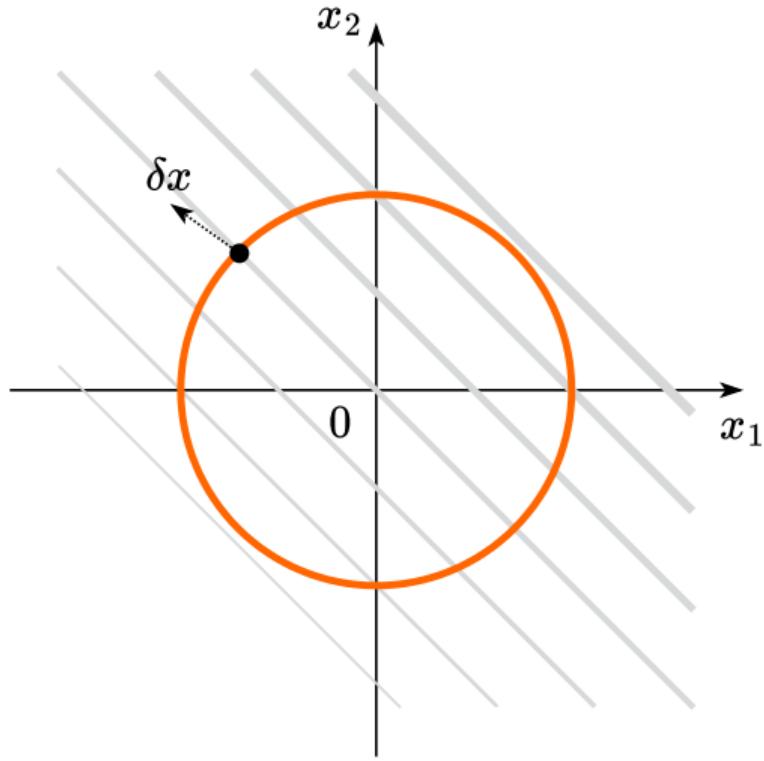


Задачи с ограничениями-равенствами

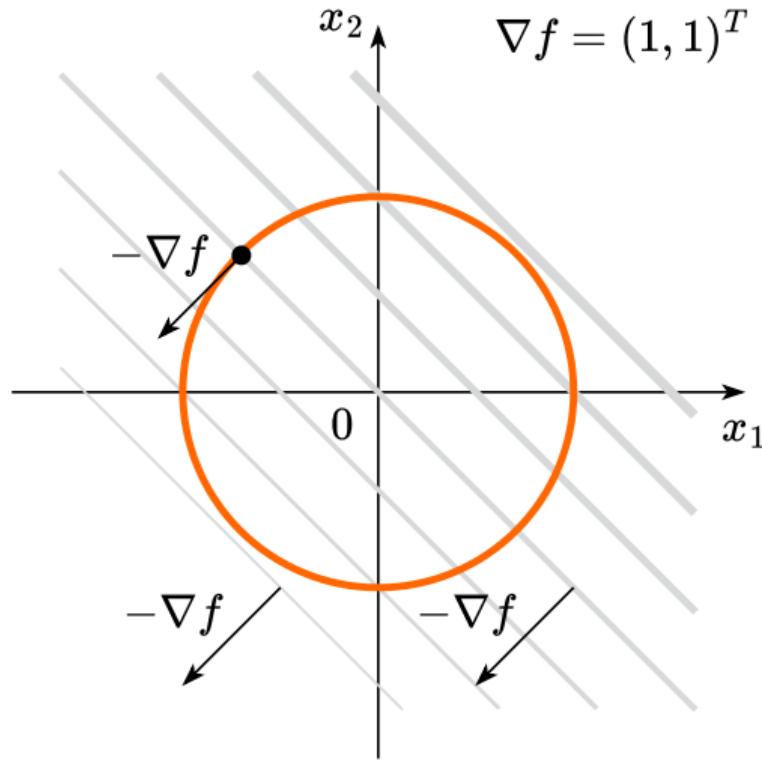
Допустимая точка x_F



Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами

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



Задачи с ограничениями-равенствами

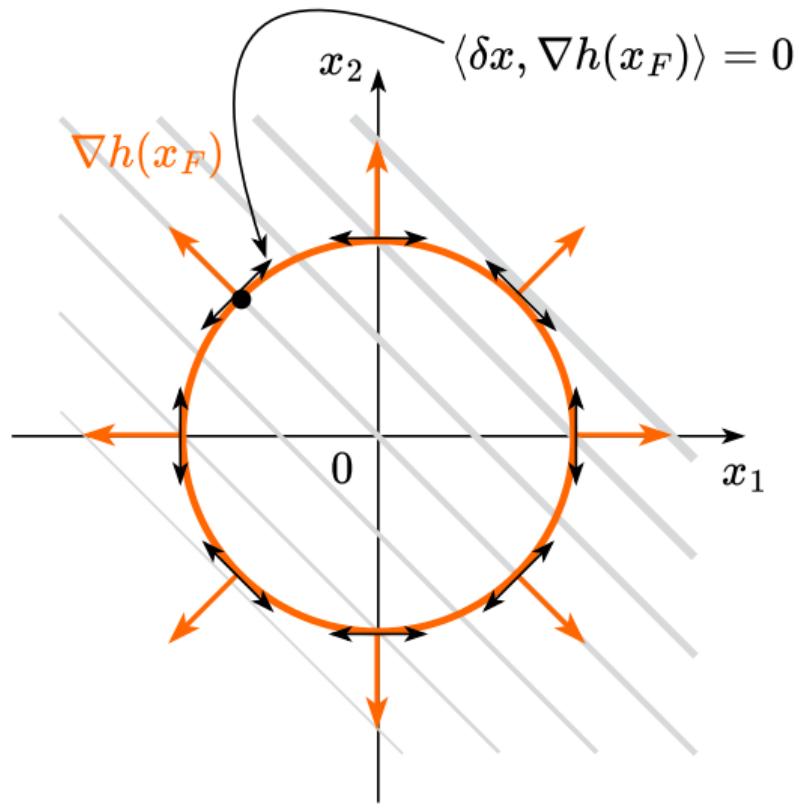
$$\nabla h = (2x_1, 2x_2)^T$$



Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами

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

Задачи с ограничениями-равенствами

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

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

Задачи с ограничениями-равенствами

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

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

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

Задачи с ограничениями-равенствами

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

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

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

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

Задачи с ограничениями-равенствами

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

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

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

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

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

Задачи с ограничениями-равенствами

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

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

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

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

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

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

Задачи с ограничениями-равенствами

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

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

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

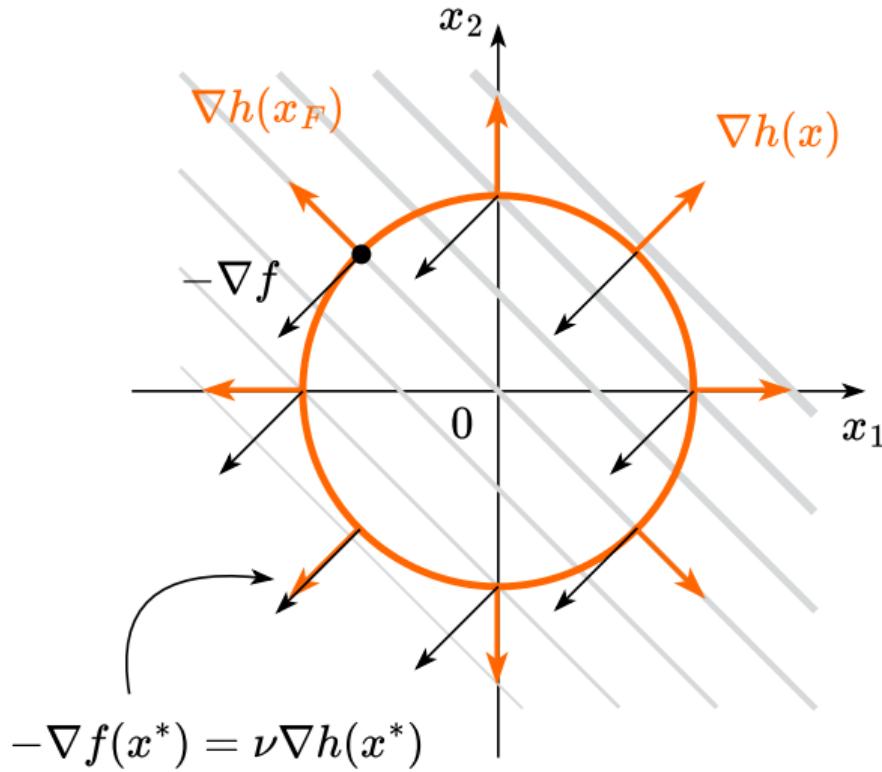
Предположим, что в процессе такого движения мы пришли в точку, где

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

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

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

Задачи с ограничениями-равенствами



Лагранжиан

Давайте определим лагранжиан (для удобства):

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

Лагранжиан

Давайте определим лагранжиан (для удобства):

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

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

Лагранжиан

Давайте определим лагранжиан (для удобства):

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

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

Необходимые условия

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

Лагранжиан

Давайте определим лагранжиан (для удобства):

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

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

Необходимые условия

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

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

Лагранжиан

Давайте определим лагранжиан (для удобства):

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

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

Необходимые условия

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

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

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

Задачи с ограничениями-равенствами

$$\begin{aligned} f(x) &\rightarrow \min_{x \in \mathbb{R}^n} \\ \text{s.t. } h_i(x) &= 0, \quad i = 1, \dots, p \end{aligned} \tag{ECP}$$

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

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

Необходимые условия

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

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

Задача наименьших квадратов

Example

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

- $m < n$

Задача наименьших квадратов

Example

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

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

Задача наименьших квадратов

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):

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

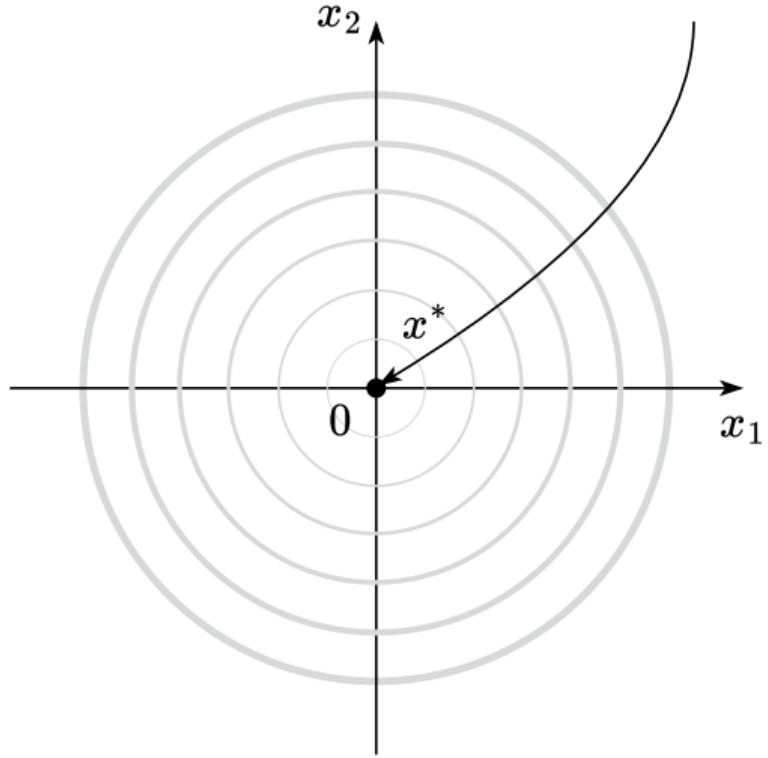
Пример задачи с ограничениями-неравенствами

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

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

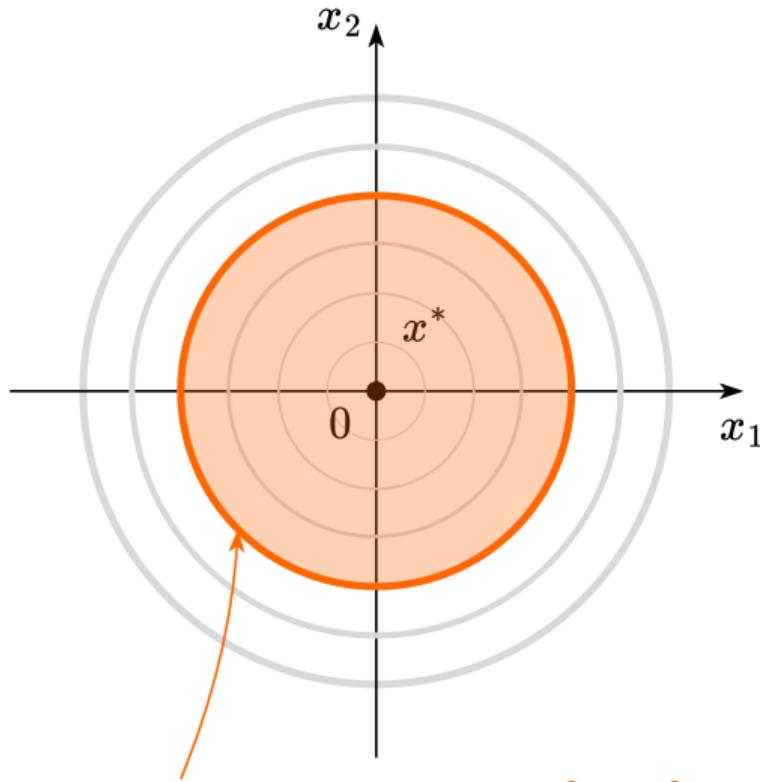
Задачи с ограничениями-неравенствами

$$x^* = \operatorname{argmin} f(x)$$



$$\text{Линии уровня } f(x) = x_1^2 + x_2^2 = C$$

Задачи с ограничениями-неравенствами



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

Задачи с ограничениями-неравенствами

Как понять, что некоторая допустимая
точка является локальным минимумом?



Задачи с ограничениями-неравенствами

Просто! Проверим достаточные условия
локального экстремума



Задачи с ограничениями-неравенствами

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

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

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

Задачи с ограничениями-неравенствами

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



Задачи с ограничениями-неравенствами

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



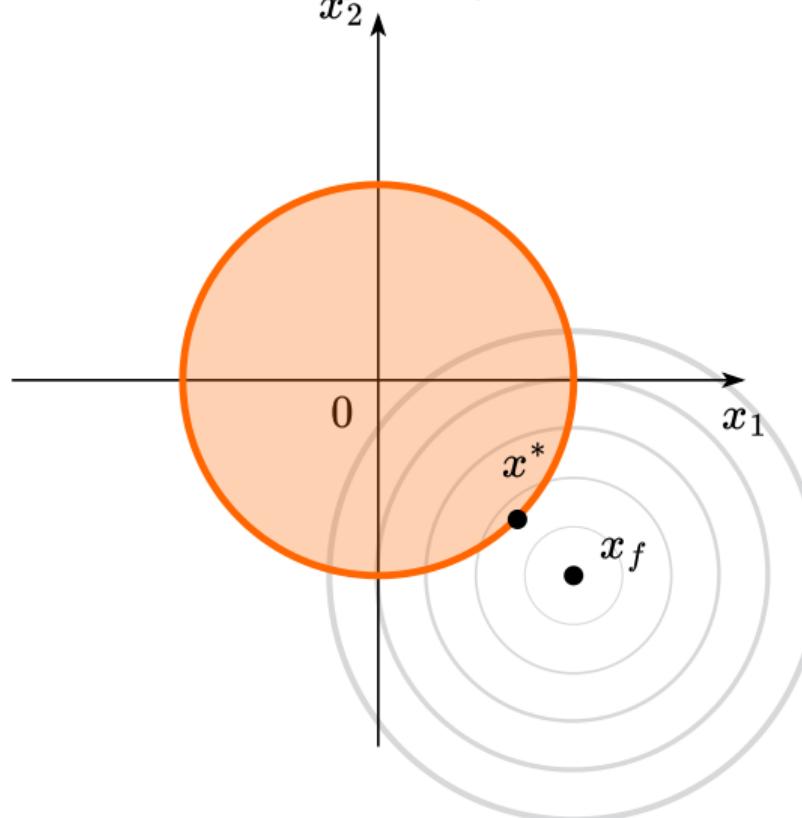
Задачи с ограничениями-неравенствами

Как понять, что некоторая допустимая
точка является локальным минимумом?



Задачи с ограничениями-неравенствами

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



Задачи с ограничениями-неравенствами

Фактически имеем задачу
с ограничением-равенством

x_2

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

0

x_1

x^*

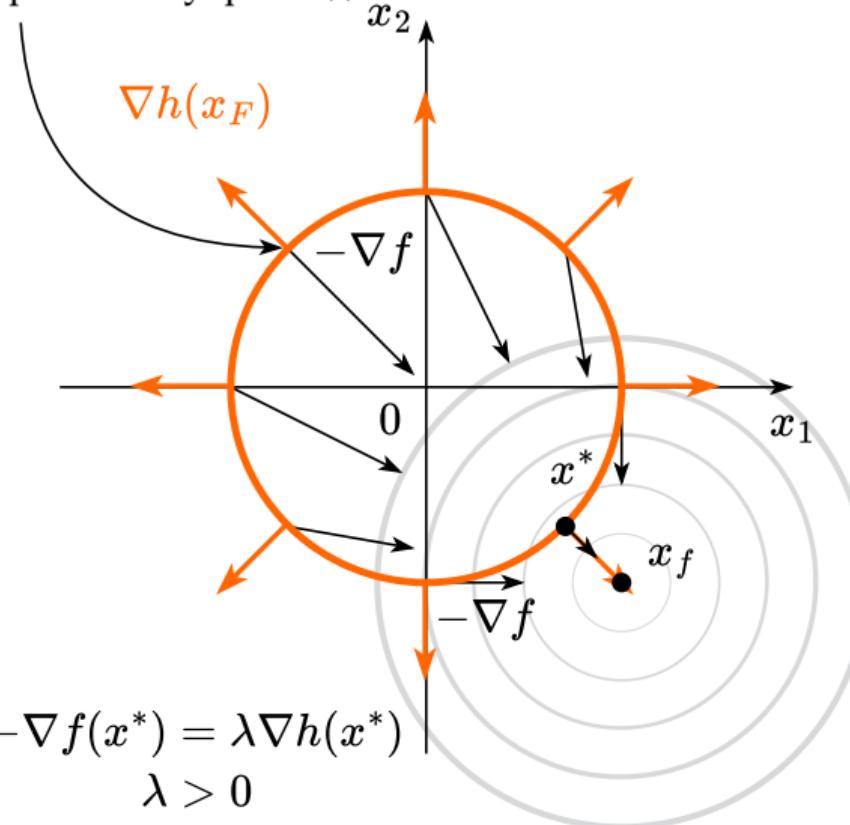
x_f

Задачи с ограничениями-неравенствами



Задачи с ограничениями-неравенствами

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



Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

Два возможных случая:

$g(x) \leq 0$ неактивно. $g(x^*) < 0$

- $g(x^*) < 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

Два возможных случая:

$g(x) \leq 0$ неактивно. $g(x^*) < 0$

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

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

Два возможных случая:

$g(x) \leq 0$ неактивно. $g(x^*) < 0$

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

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

Два возможных случая:

$g(x) \leq 0$ неактивно. $g(x^*) < 0$

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

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

Два возможных случая:

$g(x) \leq 0$ неактивно. $g(x^*) < 0$

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

$g(x) \leq 0$ активно. $g(x^*) = 0$

- $g(x^*) = 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

Два возможных случая:

$g(x) \leq 0$ неактивно. $g(x^*) < 0$

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

$g(x) \leq 0$ 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:

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

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

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:

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

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

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

- (1) $\nabla_x L(x^*, \lambda^*) = 0$
- (2) $\lambda^* \geq 0$
- (3) $\lambda^* g(x^*) = 0$
- (4) $g(x^*) \leq 0$

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.

Общая формулировка

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

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

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

Некоторые условия регулярности

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).

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

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

Некоторые условия регулярности

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

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

Пример. Проекция на гиперплоскость

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

Пример. Проекция на гиперплоскость

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

Решение

Лагранжиан:

Пример. Проекция на гиперплоскость

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

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

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

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} :

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

Пример. Проекция на гиперплоскость

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

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} :

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

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

Пример. Проекция на гиперплоскость

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

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} :

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

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

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

Пример. Проекция на единичный симплекс

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

Пример. Проекция на единичный симплекс

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

Условия ККТ

Лагранжиан задается следующим образом:

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

Пример. Проекция на единичный симплекс

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

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:

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

Пример. Проекция на единичный симплекс

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

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:

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

Пример. Проекция на единичный симплекс

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

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:

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

Пример. Проекция на единичный симплекс

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

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:

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

Пример. Проекция на единичный симплекс

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

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:

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

Пример. Проекция на единичный симплекс

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

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:

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

Question

Решите систему выше за $O(n \log n)$.

Пример. Проекция на единичный симплекс

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

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:

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

Question

Решите систему выше за $O(n \log n)$.

Question

Решите систему выше за $O(n)$.

Ссылки

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

Ссылки

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

Ссылки

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

Ссылки

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

Ссылки

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