

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

Даня Меркулов

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”*



Рисунок 1: Жозеф Луи Лагранж

1 Условия оптимальности

1.1 Теория

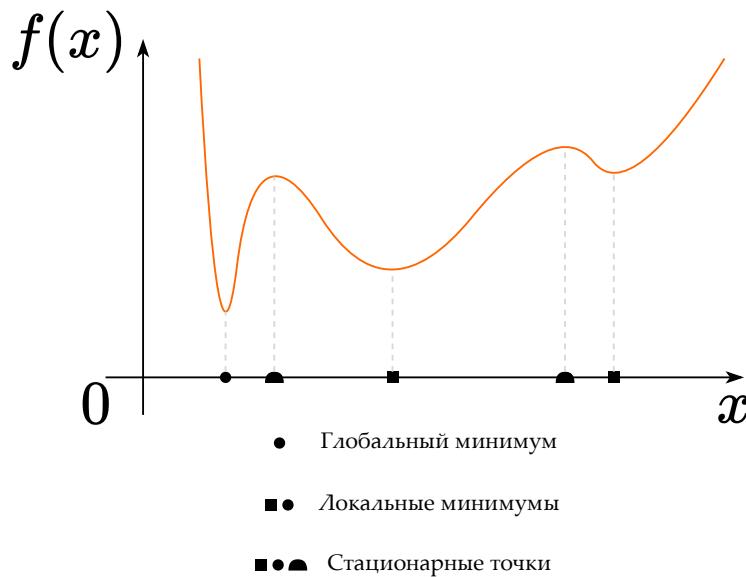


Рисунок 2: Иллюстрация различных стационарных (критических) точек

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

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.

1.2 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.

GOOD NEWS EVERYONE!



Рисунок 3: Многие практические задачи теоретически разрешимы

■ 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 $t \in (0, 1)$.

2 Безусловная оптимизация

2.1 Необходимые условия

■ Необходимое условие оптимальности первого порядка

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

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

Доказательство

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

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

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

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

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

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

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

2.2 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 поэтому

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

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.

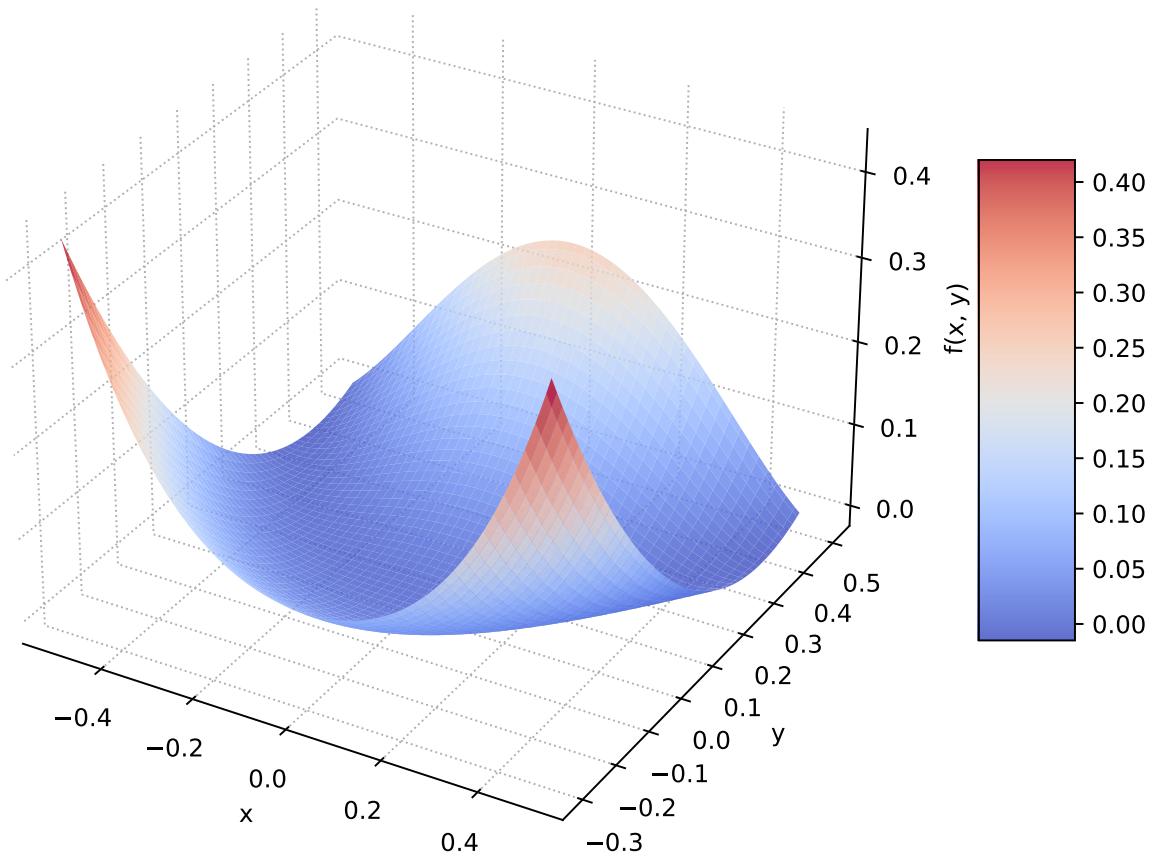
2.3 Контрпример Пеано

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 $y = tx$ 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



3 Условная оптимизация

3.1 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

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

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

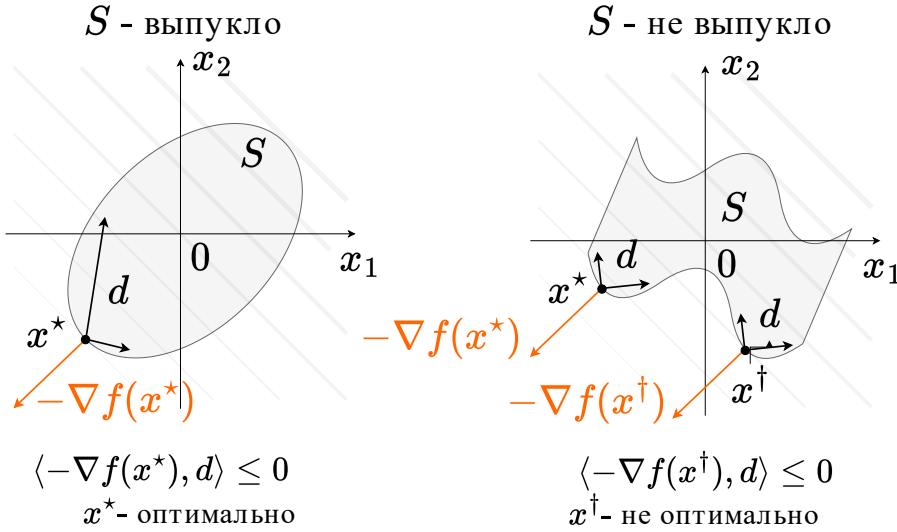


Рисунок 4: Общее условие локальной оптимальности первого порядка

3.2 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: $S^* = \{x^*\}$.

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

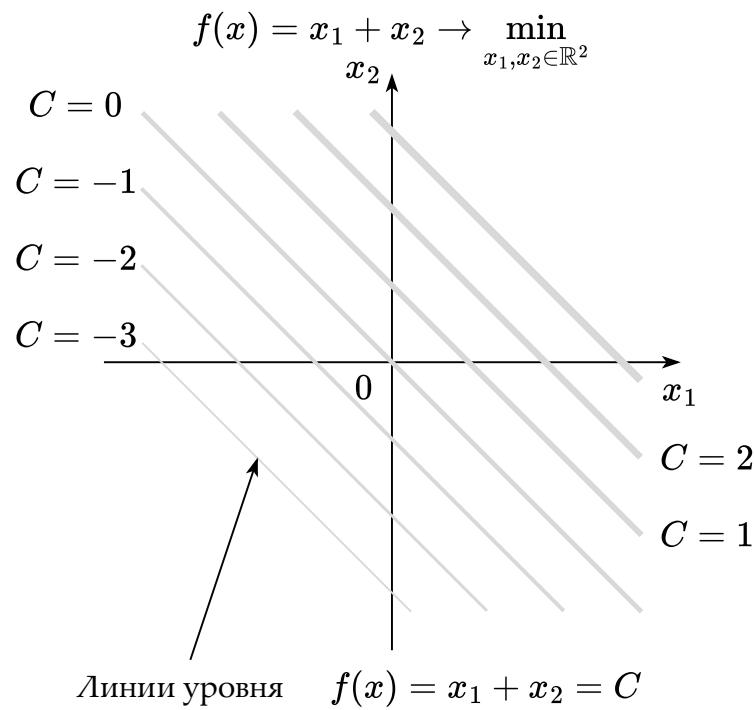


Рисунок 5: Иллюстрация ККТ

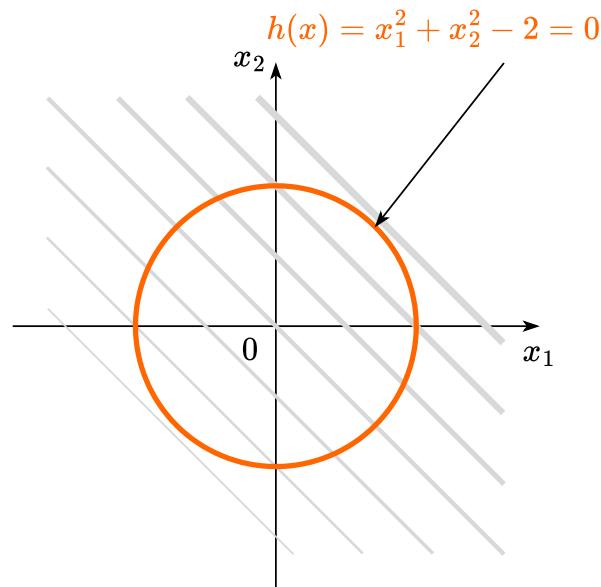


Рисунок 6: Иллюстрация ККТ

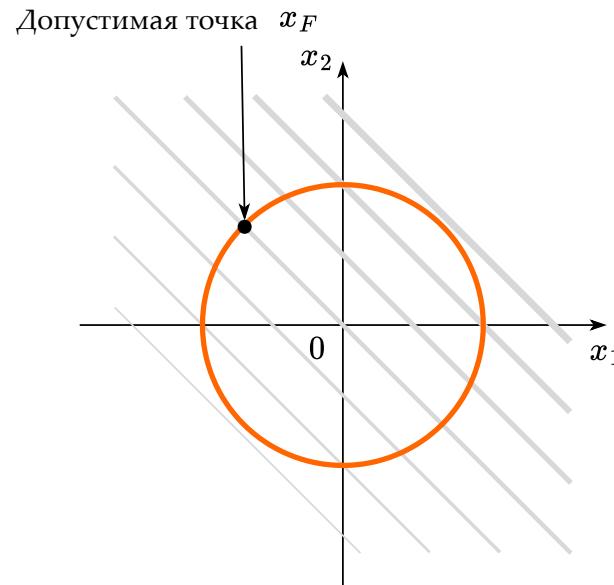


Рисунок 7: Иллюстрация ККТ

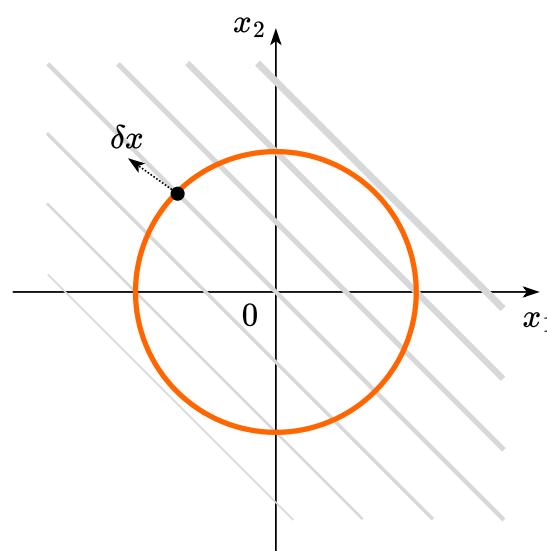


Рисунок 8: Иллюстрация ККТ

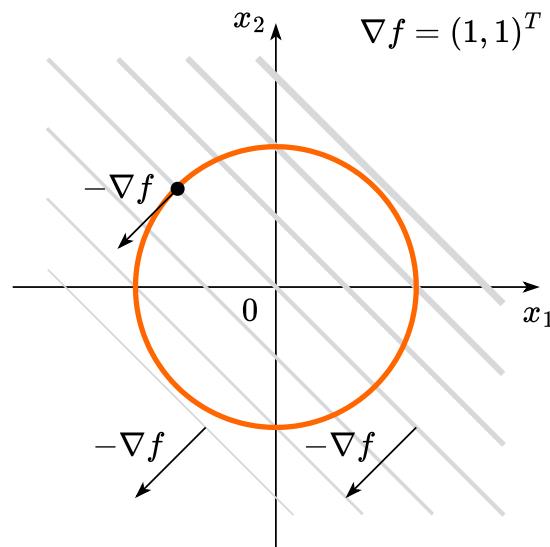


Рисунок 9: Иллюстрация ККТ

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

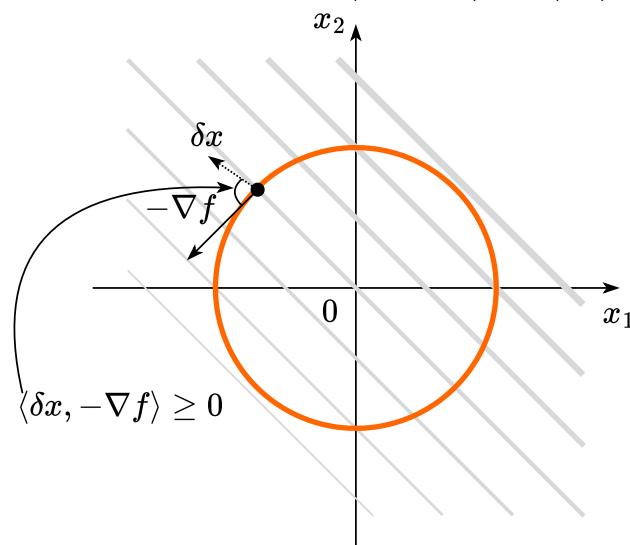


Рисунок 10: Иллюстрация ККТ

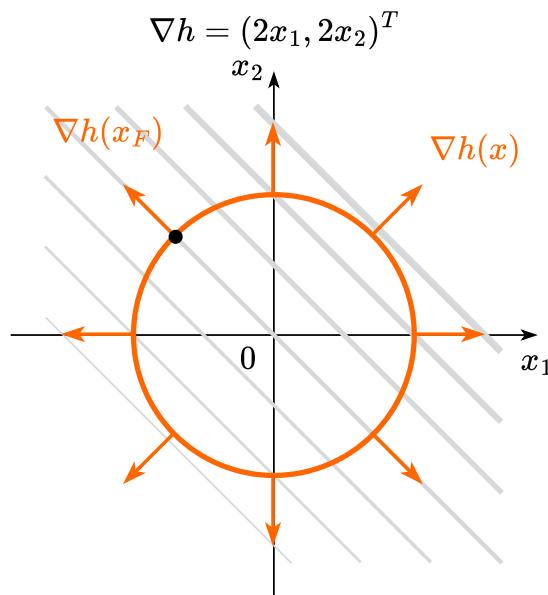


Рисунок 11: Иллюстрация ККТ

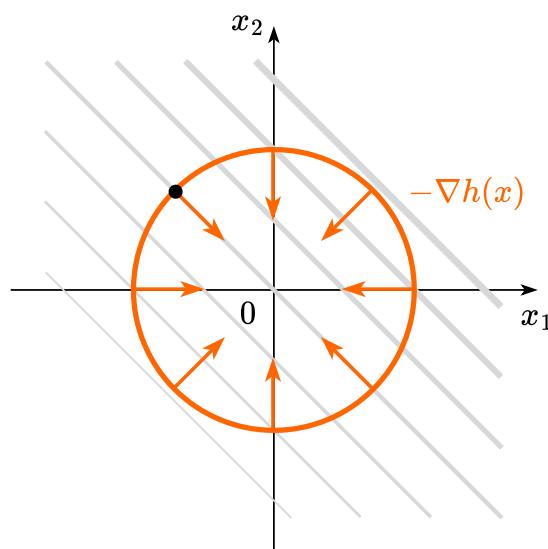


Рисунок 12: Иллюстрация ККТ

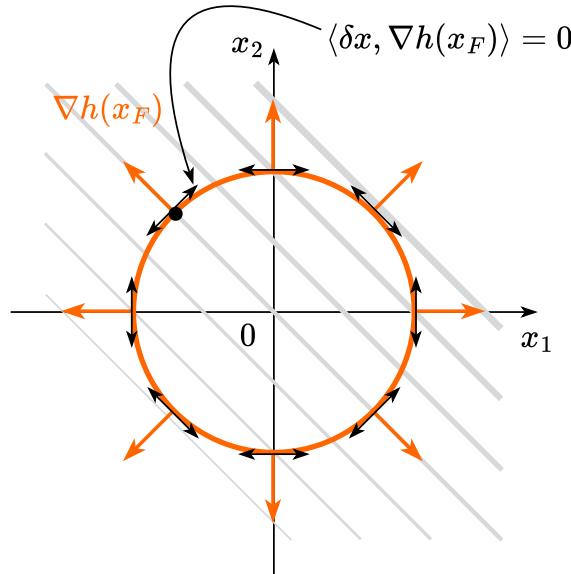


Рисунок 13: Иллюстрация ККТ

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.

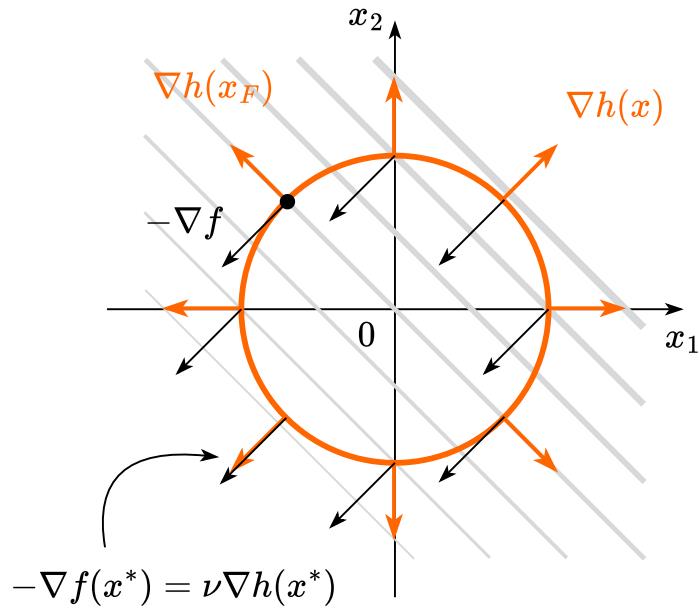


Рисунок 14: Иллюстрация ККТ

3.4 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 $L(x^*, \nu^*) = f(x^*)$.

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

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

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

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

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

3.5 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):

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

4 Задачи с ограничениями-неравенствами

4.1 Пример задачи с ограничениями-неравенствами

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

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

4.2 Задачи с ограничениями-неравенствами

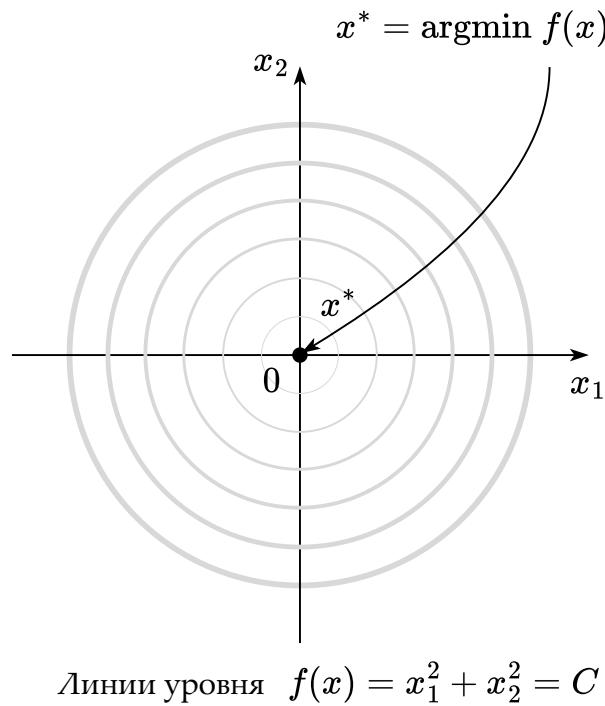


Рисунок 15: Иллюстрация ККТ (случай неравенства)

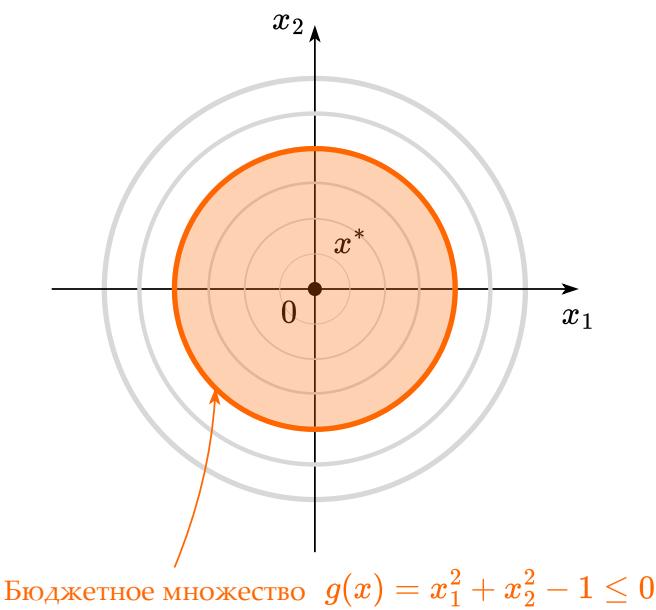


Рисунок 16: Иллюстрация ККТ (случай неравенства)

Как понять, что некоторая допустимая
точка является локальным минимумом?

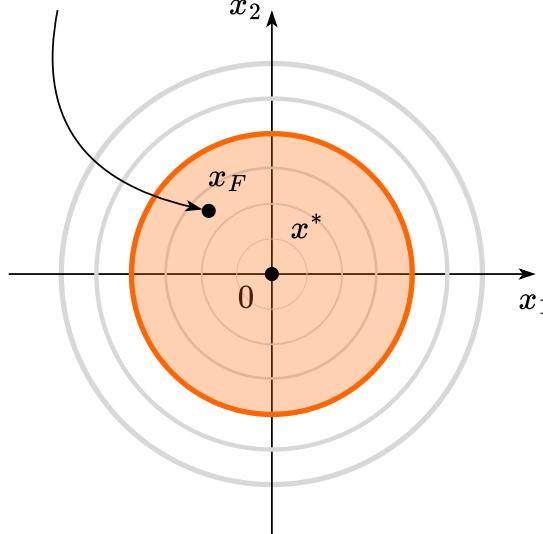


Рисунок 17: Иллюстрация ККТ (случай неравенства)

Просто! Проверим достаточные условия
локального экстремума

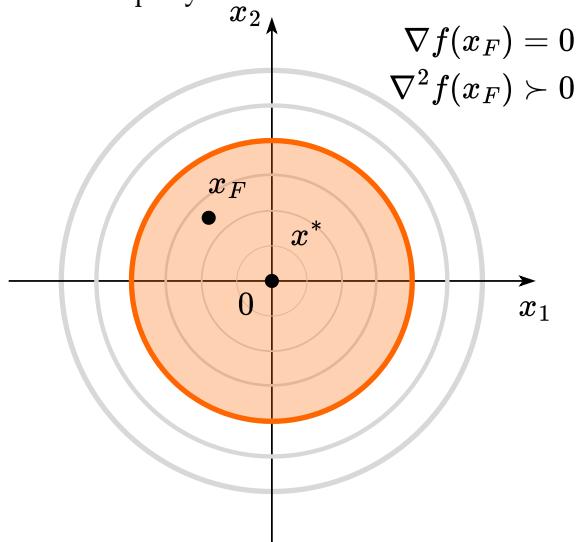


Рисунок 18: Иллюстрация ККТ (случай неравенства)

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.

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

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

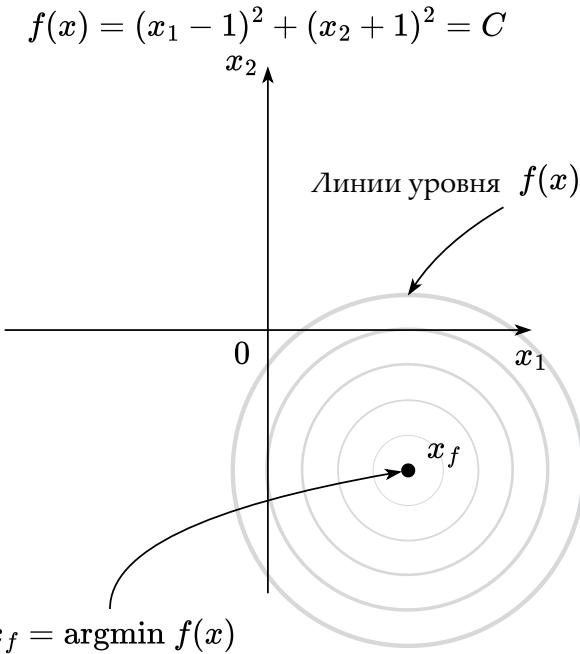


Рисунок 19: Иллюстрация ККТ (случай неравенства)

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

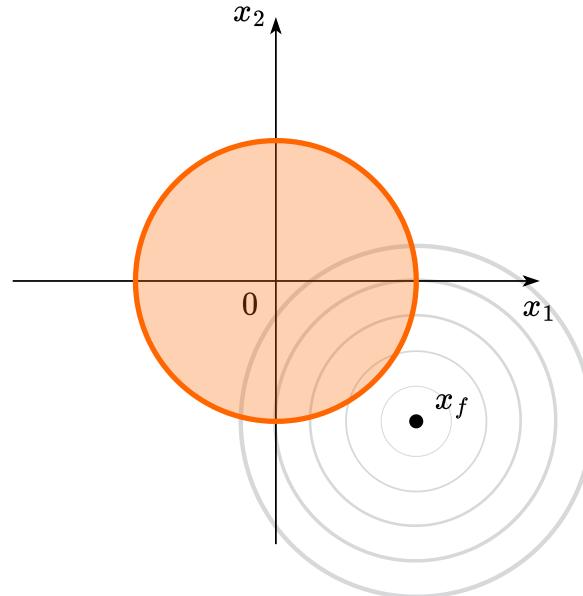


Рисунок 20: Иллюстрация ККТ (случай неравенства)

Как понять, что некоторая допустимая
точка является локальным минимумом?

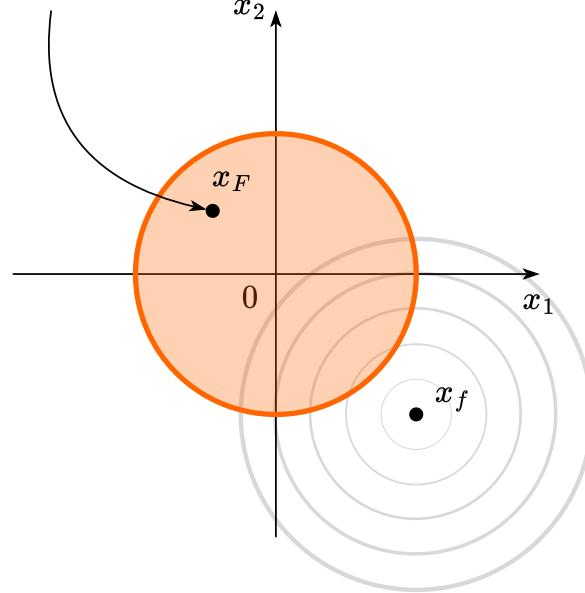


Рисунок 21: Иллюстрация ККТ (случай неравенства)

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

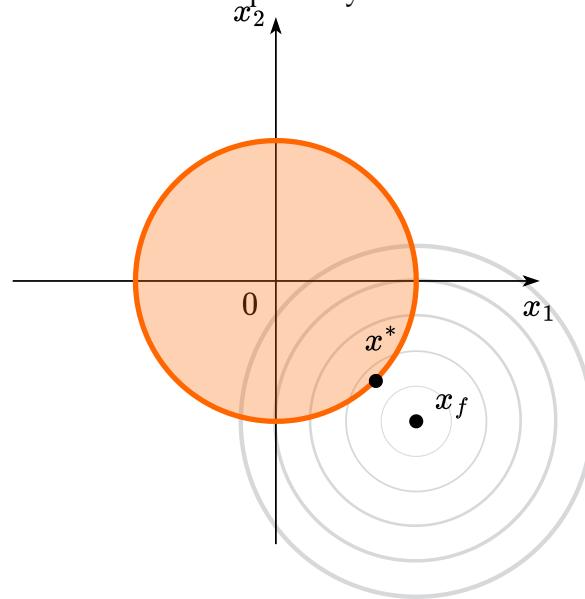


Рисунок 22: Иллюстрация ККТ (случай неравенства)

Фактически имеем задачу
с ограничением-равенством 

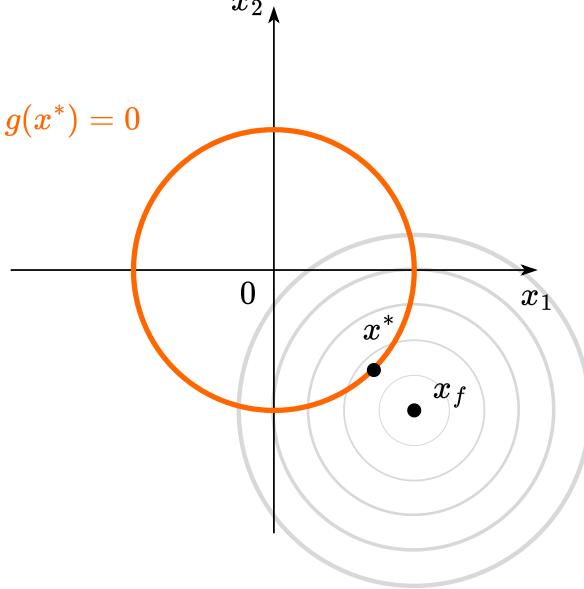


Рисунок 23: Иллюстрация ККТ (случай неравенства)

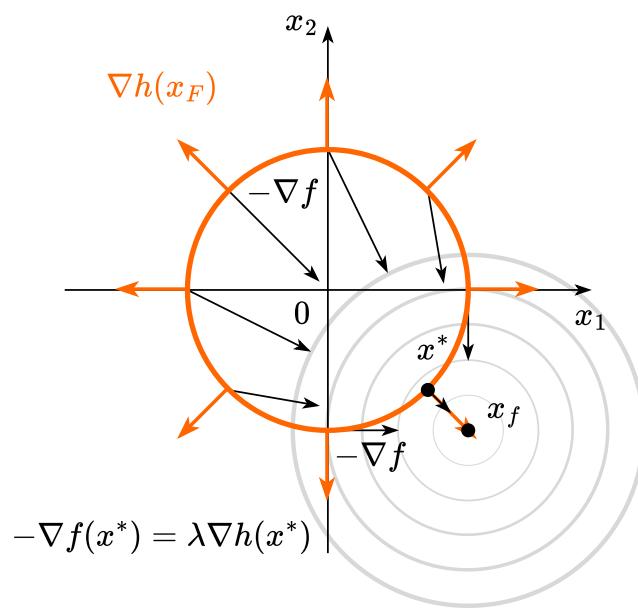


Рисунок 24: Иллюстрация ККТ (случай неравенства)

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

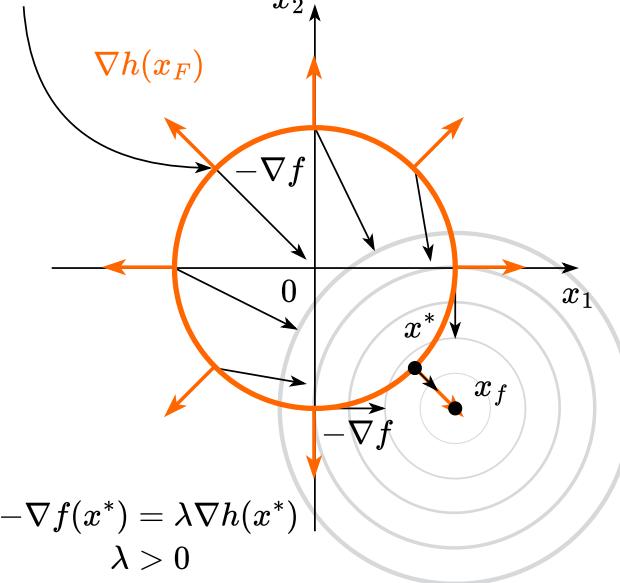


Рисунок 25: Иллюстрация ККТ (случай неравенства)

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

4.3 Лагранжиан для задач с ограничениями-неравенствами

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

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

Определим функцию Лагранжа:

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

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

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

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

4.4 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:

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

4.5 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.

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

4.6 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](#).

4.7 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} :

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

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

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

4.8 Проекция на единичный симплекс

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

4.8.0.1 Условия ККТ

Лагранжиан задается следующим образом:

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

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

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

Question

Решите систему выше за $O(n \log n)$.

Question

Решите систему выше за $O(n)$.

4.9 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.

5 Задачи

Задача 1

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

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

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

Найдите явное решение следующей задачи.

$$\begin{aligned} f(x, y) &= x + y \rightarrow \min \\ \text{s.t. } &x^2 + y^2 = 1 \end{aligned}$$

где $x, y \in \mathbb{R}$.

Задача 3

Найдите явное решение следующей задачи.

$$\begin{aligned} \langle c, x \rangle + \sum_{i=1}^n x_i \log x_i &\rightarrow \min_{x \in \mathbb{R}^n} \\ \text{s.t. } &\sum_{i=1}^n x_i = 1, \end{aligned}$$

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

Задача 4

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

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

Имеет единственное решение и найдите его.

Задача 5

Пусть $y \in \{-1, 1\}$, и $X \in \mathbb{R}^{n \times p}$, задача опорных векторов (Support Vector Machine) задается следующим образом:

$$\begin{aligned} \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \xi_i &\rightarrow \min_{w, w_0, \xi_i} \\ \text{s.t. } \xi_i &\geq 0, i = 1, \dots, n \\ y_i (x_i^T w + w_0) &\geq 1 - \xi_i, i = 1, \dots, n \end{aligned}$$

найдите условие стационарности ККТ.

Задача 6

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

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

$$C \in \mathbb{S}_{++}^n, a \neq 0$$

В ответе следует избежать явного обращения матрицы C .

Задача 7 (БОНУС)

Для некоторых $\Sigma, \Sigma_0 \in \mathbb{S}_{++}^n$ определим KL-расхождение 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.

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

Покажите, 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)

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

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

Имеет единственное решение I_n , и выведите неравенство Гильберта:

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

6 Задачи на дом

В этом разделе вы можете рассматривать произвольную норму или евклидову норму, если не указано иное.

1. Простой пример [10 баллов]

$$\begin{aligned} x^2 + 1 &\rightarrow \min_{x \in \mathbb{R}} \\ \text{s.t. } (x-2)(x-4) &\leq 0 \end{aligned}$$

- Найдите допустимое множество, оптимальное значение и оптимальное решение.
- Постройте график функции $x^2 + 1$ в зависимости от x . На том же графике покажите допустимое множество, оптимальную точку и значение, а также постройте график лагранжиана $L(x, \mu)$ в зависимости от x для нескольких положительных значений μ . Проверьте свойство нижней границы ($p^* \geq \inf_x L(x, \mu)$ для $\mu \geq 0$). Выведите и нарисуйте функцию Лагранжа g .
- Пусть $p^*(u)$ обозначает оптимальное значение следующей задачи:

$$\begin{aligned} x^2 + 1 &\rightarrow \min_{x \in \mathbb{R}} \\ \text{s.t. } (x-2)(x-4) &\leq u \end{aligned}$$

как функции параметра u . Постройте график $p^*(u)$. Проверьте, что $\frac{dp^*(0)}{du} = -\mu^*$

- Рассмотрим гладкую выпуклую функцию $f(x)$ в некоторой точке x_k . Её разложение Тейлора первого порядка имеет вид:

$$f_{x_k}^I(x) = f(x_k) + \nabla f(x_k)^\top (x - x_k),$$

где мы можем определить $\delta x = x - x_k$ и $g = \nabla f(x_k)$. Таким образом, разложение можно переписать как:

$$f_{x_k}^I(\delta x) = f(x_k) + g^\top \delta x.$$

Предположим, мы хотим построить семейство методов оптимизации, которое будет определяться следующим образом:

$$x_{k+1} = \arg \min_x \left\{ f_{x_k}^I(\delta x) + \frac{\lambda}{2} \|\delta x\|^2 \right\},$$

где $\lambda > 0$ является параметром.

- [5 баллов] Покажите, что этот метод эквивалентен методу градиентного спуска с выбором евклидовой нормы вектора $\|\delta x\| = \|\delta x\|_2$. Найдите соответствующий коэффициент обучения.

2. [5 баллов] Докажите, что следующее утверждение верно:

$$\arg \min_{\delta x \in \mathbb{R}^n} \left\{ g^T \delta x + \frac{\lambda}{2} \|\delta x\|^2 \right\} = -\frac{\|g\|_*}{\lambda} \arg \max_{\|t\|=1} \{t^T g\},$$

где $\|g\|_*$ является [двойственной нормой](#) g .

3. [3 балла] Рассмотрим другую векторную норму $\|\delta x\| = \|\delta x\|_\infty$. Запишите явное выражение для соответствующего метода.

4. [2 балла] Рассмотрим индуцированную операторную матричную норму для любой матрицы $W \in \mathbb{R}^{d_{out} \times d_{in}}$

$$\|W\|_{\alpha \rightarrow \beta} = \max_{x \in \mathbb{R}^{d_{in}}} \frac{\|Wx\|_\beta}{\|x\|_\alpha}.$$

Обычно, когда мы решаем оптимизационные задачи в глубоком обучении, мы складываем матрицы весов для всех слоев $l = [1, L]$ в один вектор.

$$w = \text{vec}(W_1, W_2, \dots, W_L) \in \mathbb{R}^n,$$

Можете ли вы записать явное выражение, которое связывает

$$\|w\|_\infty \quad \text{and} \quad \|W_l\|_{\alpha \rightarrow \beta}, \quad l = [1, L]?$$

3. [10 баллов] Найдите явное решение следующей задачи линейного программирования.

$$\begin{aligned} c^\top x &\rightarrow \min_{x \in \mathbb{R}^n} \\ \text{s.t. } 1^\top x &= 1, \\ x &\succeq 0 \end{aligned}$$

Эта задача может быть рассмотрена как самый простой пример задачи оптимизации портфеля.

4. [20 баллов] Покажите, что следующая задача имеет единственное решение и найдите его:

$$\begin{aligned} \langle C^{-1}, X \rangle - \log \det X &\rightarrow \min_{x \in \mathbb{R}^{n \times n}} \\ \text{s.t. } \langle Xa, a \rangle &\leq 1, \end{aligned}$$

где $C \in \mathbb{S}_{++}^n$, $a \in \mathbb{R}^n \neq 0$. Ответ не должен включать обращение матрицы C .

5. [20 баллов] Найдите явное решение следующей задачи квадратичного программирования.

$$\begin{aligned} c^\top x &\rightarrow \min_{x \in \mathbb{R}^n} \\ \text{s.t. } (x - x_c)^\top A(x - x_c) &\leq 1, \end{aligned}$$

где $A \in \mathbb{S}_{++}^n$, $c \neq 0$, $x_c \in \mathbb{R}^n$.

6. [10 баллов] Рассмотрим задачу наименьших квадратов с ограничениями равенства.

$$\begin{aligned} \|Ax - b\|_2^2 &\rightarrow \min_{x \in \mathbb{R}^n} \\ \text{s.t. } &Cx = d, \end{aligned}$$

где $A \in \mathbb{R}^{m \times n}$ с $\text{rank}A = n$, и $C \in \mathbb{R}^{k \times n}$ с $\text{rank}C = k$. Запишите условия ККТ, и выведите выражения для решения x^* .

7. **Интерпретация условий ККТ в терминах опорной гиперплоскости.** [10 баллов] Рассмотрим выпуклую задачу без ограничений равенства.

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

Предположим, что $\exists x^* \in \mathbb{R}^n, \mu^* \in \mathbb{R}^m$ удовлетворяют условиям ККТ

$$\begin{aligned} \nabla_x L(x^*, \mu^*) &= \nabla f_0(x^*) + \sum_{i=1}^m \mu_i^* \nabla f_i(x^*) = 0 \\ \mu_i^* &\geq 0, \quad i = [1, m] \\ \mu_i^* f_i(x^*) &= 0, \quad i = [1, m] \\ f_i(x^*) &\leq 0, \quad i = [1, m] \end{aligned}$$

Покажите, что

$$\nabla f_0(x^*)^\top (x - x^*) \geq 0$$

для всех допустимых x . Другими словами, условия ККТ подразумевают простой критерий оптимальности или $\nabla f_0(x^*)$ определяет опорную гиперплоскость к допустимому множеству в точке x^* .

8. **Метод штрафов для ограничений равенства.** [10 баллов] Рассмотрим задачу минимизации

$$\begin{aligned} f_0(x) &\rightarrow \min_{x \in \mathbb{R}^n} \\ \text{s.t. } &Ax = b, \end{aligned}$$

где $f_0(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ выпукла и дифференцируема, и $A \in \mathbb{R}^{m \times n}$ с $\text{rank}A = m$. В методе квадратичных штрафов мы формируем вспомогательную функцию

$$\phi(x) = f_0(x) + \alpha \|Ax - b\|_2^2,$$

где $\alpha > 0$ является параметром. Эта вспомогательная функция состоит из целевой функции плюс штрафное слагаемое $\alpha \|Ax - b\|_2^2$. Идея состоит в том, что минимизатор вспомогательной функции,

\tilde{x} , должен быть приближенным решением исходной задачи. Интуиция подсказывает, что чем больше вес штрафа α , тем лучше приближение \tilde{x} к решению исходной задачи. Предположим, что \tilde{x} является минимизатором $\phi(x)$. Найдите соответствующую нижнюю границу для оптимального значения исходной задачи.