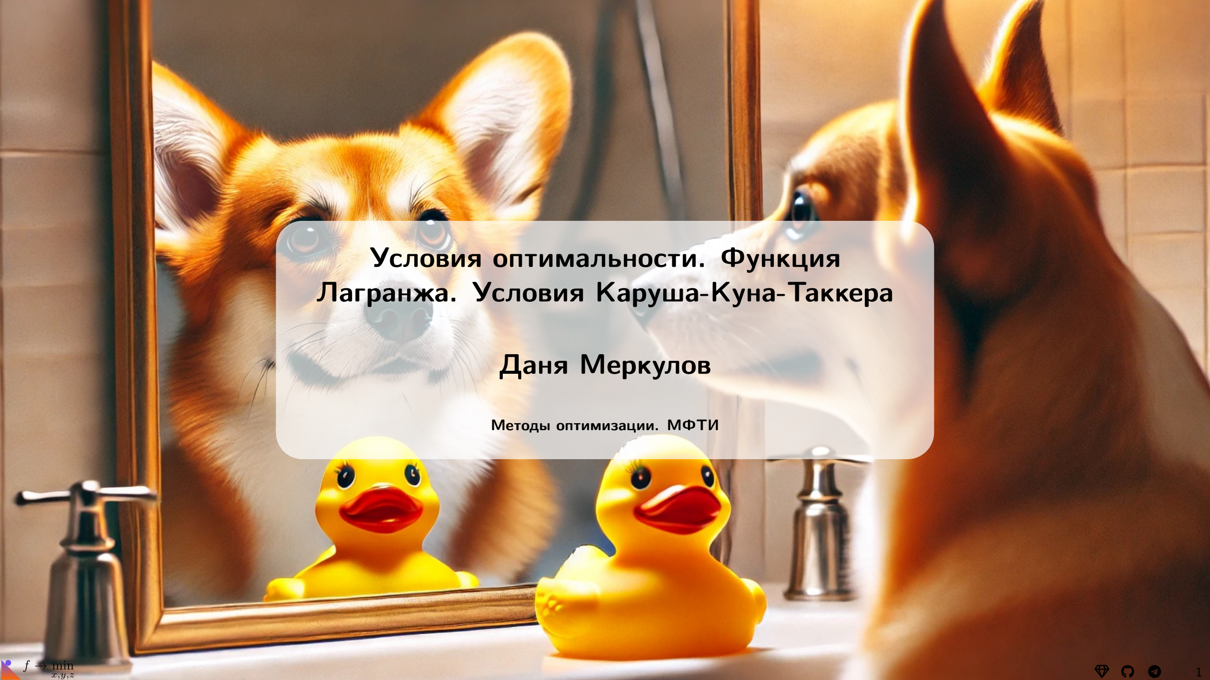


Условия оптимальности. Функция Лагранжа. Условия Каруша-Куна-Таккера

Даня Меркулов

Методы оптимизации. МФТИ

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

—Предисловие к "Аналитической механике"

Рис. 1: Жозеф Луи Лагранж

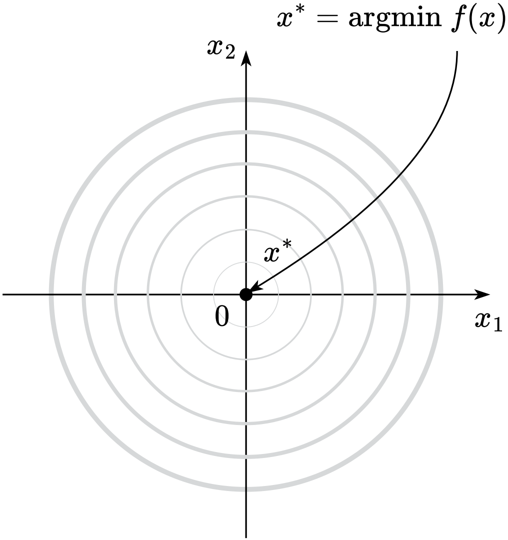
Задачи с ограничениями-неравенствами

Пример задачи с ограничениями-неравенствами

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

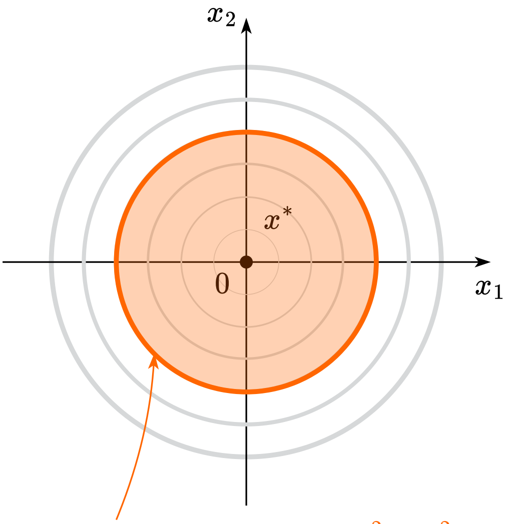
$$\begin{aligned} f(x) &\to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } g(x) &\leq 0 \end{aligned}$$

Задачи с ограничениями-неравенствами



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

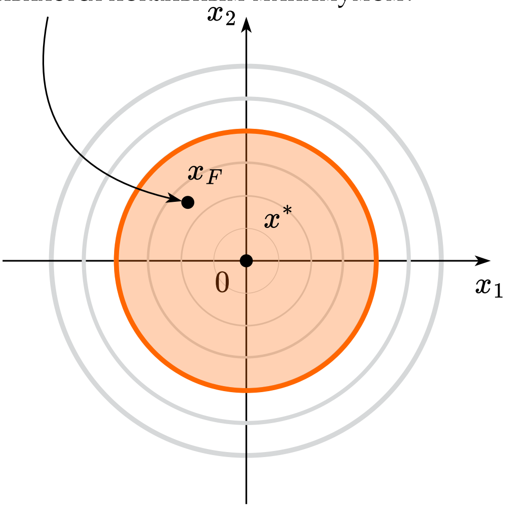
Задачи с ограничениями-неравенствами



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

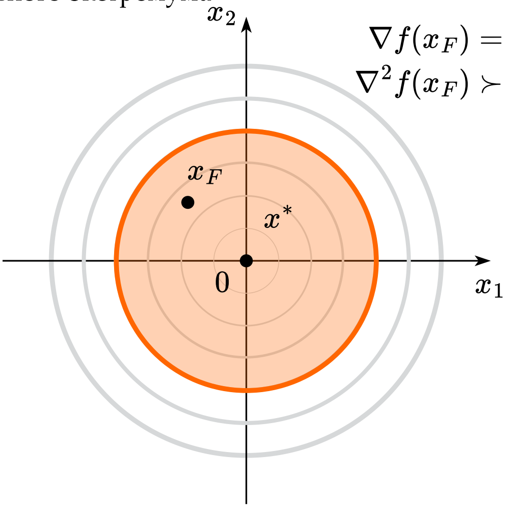
Задачи с ограничениями-неравенствами

Как понять, что некоторая допустимая точка является локальным минимумом?



Задачи с ограничениями-неравенствами

Просто! Проверим достаточные условия локального экстремума



$$\nabla f(x_F) = 0$$
$$\nabla^2 f(x_F) \succ 0$$

Задачи с ограничениями-неравенствами

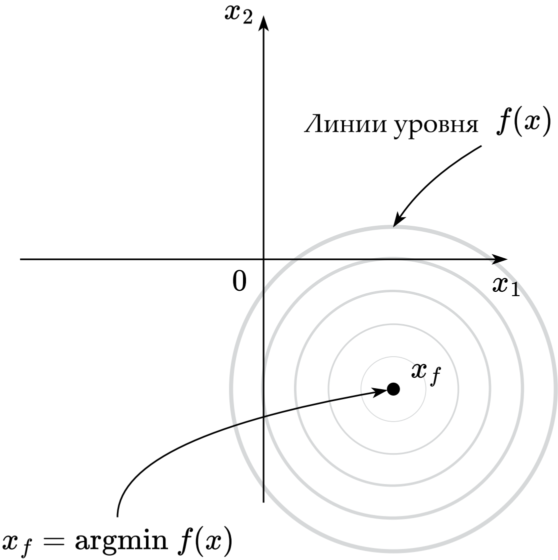
Таким образом, если ограничения типа неравенства неактивны в условной задаче, то мы можем решать задачу без ограничений. Однако так бывает не всегда. Рассмотрим второй простой пример.

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

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

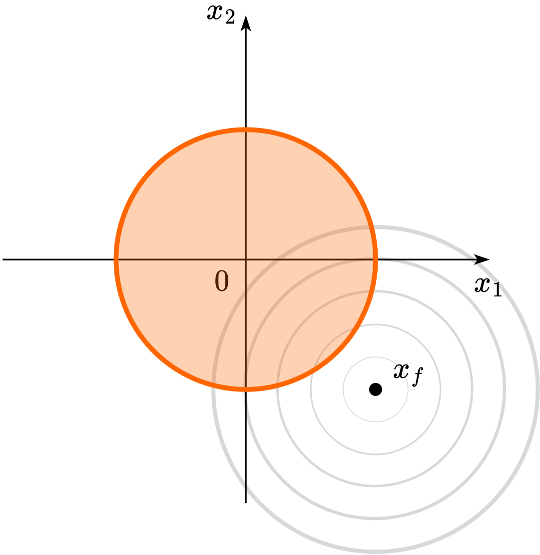
Задачи с ограничениями-неравенствами

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



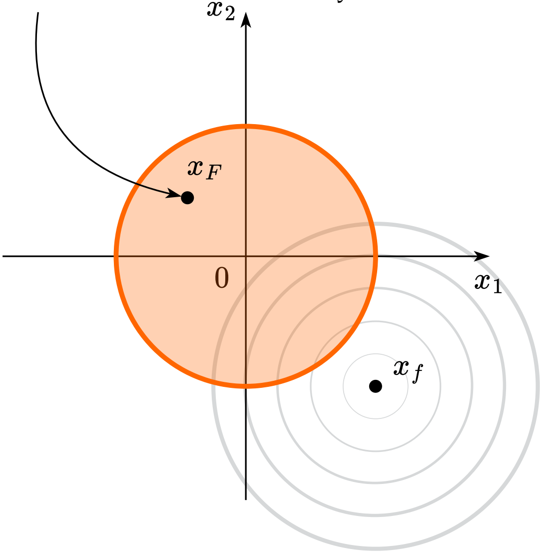
Задачи с ограничениями-неравенствами

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



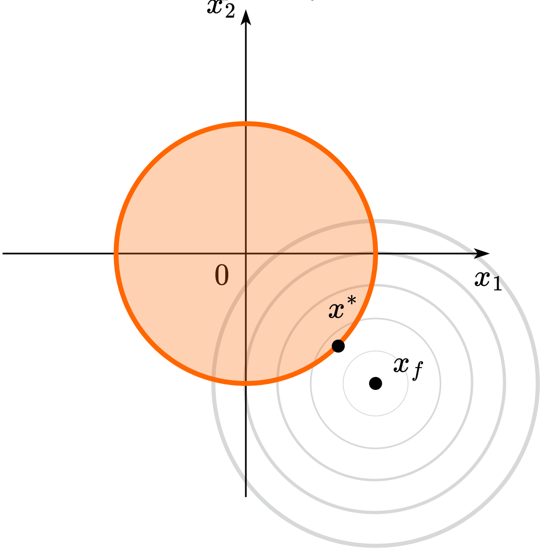
Задачи с ограничениями-неравенствами

Как понять, что некоторая допустимая точка является локальным минимумом?



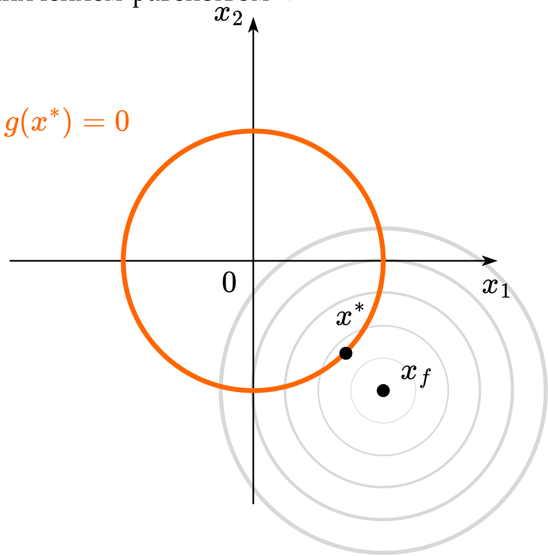
Задачи с ограничениями-неравенствами

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

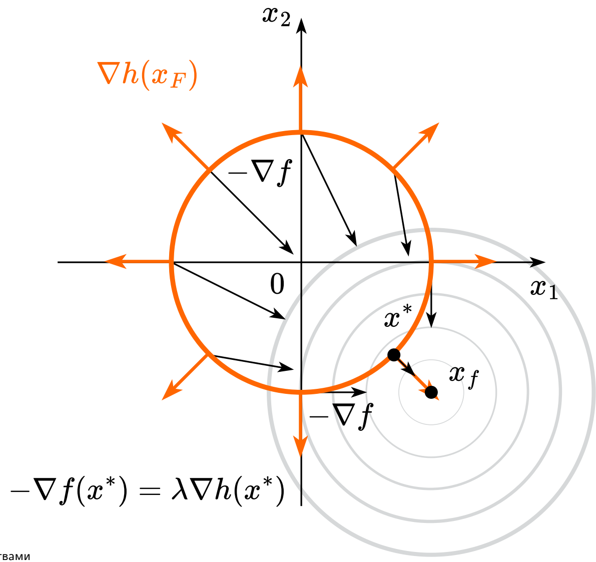


Задачи с ограничениями-неравенствами

Фактически имеем задачу
с ограничением-равенством 💡

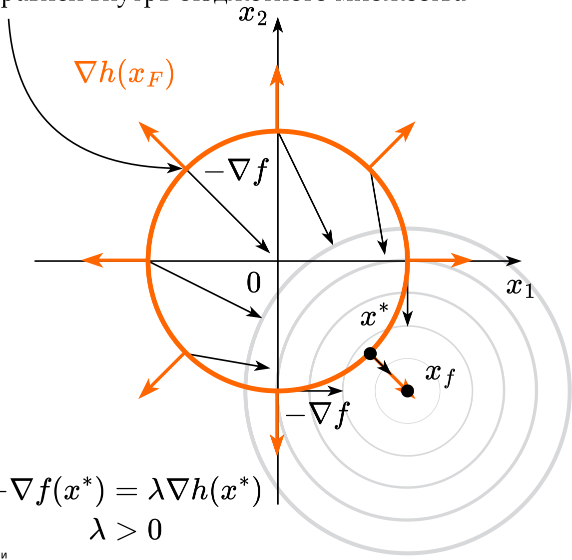


Задачи с ограничениями-неравенствами



Задачи с ограничениями-неравенствами

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



$$-\nabla f(x^*) = \lambda \nabla h(x^*)$$
$$\lambda > 0$$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

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

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

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

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

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

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

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

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

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

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

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

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

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

$g(x) \leq 0$ активно: $g(x^*) = 0$

- $g(x^*) = 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

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

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

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

$g(x) \leq 0$ активно: $g(x^*) = 0$

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

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

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

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

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

$g(x) \leq 0$ активно: $g(x^*) = 0$

- $g(x^*) = 0$
- Необходимые условия: $-\nabla f(x^*) = \lambda \nabla g(x^*)$, $\lambda > 0$
- Достаточные условия: $\langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0, \forall y \neq 0 \in \mathbb{R}^n : \nabla g(x^*)^\top y = 0$

Лагранжиан для задач с ограничениями-неравенствами

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

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

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.

Лагранжиан для задач с ограничениями-неравенствами

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

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

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:

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

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

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

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

$$(5)\ \forall y \in C(x^*) : \langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0$$

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

$I(x^*) = \{i \mid g_i(x^*) = 0\}$

KKT

Общая формулировка

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

Данная формулировка является общей задачей математического программирования.

Решение включает в себя построение лагранжиана:

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

Необходимые условия

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

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

Необходимые условия

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

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

Необходимые условия

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

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

Необходимые условия

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

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

Необходимые условия

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

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

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, если у нас есть регулярность, мы можем записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределенным* гессианом лагранжиана.

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

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, если у нас есть регулярность, мы можем записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y\rangle \geq 0$ с *полуопределенным* гессианом лагранжиана.

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

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, если у нас есть регулярность, мы можем записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределенным* гессианом лагранжиана.

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

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, если у нас есть регулярность, мы можем записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределенным* гессианом лагранжиана.

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

Доказательство в простом случае

Субдифференциальная форма ККТ

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

$$
\begin{split}
& f_0(x) \to \min_{x \in X} \\
\text{s.t. } & f_j(x) \leq 0, \; j = 1, \ldots, m
\end{split}
$$

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

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

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

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

Доказательство в простом случае

Доказательство

1. Рассмотрим функцию

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

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

Доказательство в простом случае

Доказательство

1. Рассмотрим функцию

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

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

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

Доказательство в простом случае

Доказательство

1. Рассмотрим функцию

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

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

2. Из теоремы Ферма в субдифференциальной форме

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

3. Из теоремы Дубовицкого-Милютина имеем

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

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

Доказательство в простом случае

Доказательство

1. Рассмотрим функцию

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

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

2. Из теоремы Ферма в субдифференциальной форме

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

3. Из теоремы Дубовицкого-Милютина имеем

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

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

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

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

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

Пример. Проекция на гиперплоскость

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

Пример. Проекция на гиперплоскость

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

Решение

Лагранжиан:

Пример. Проекция на гиперплоскость

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

Решение

Лагранжиан:

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

Пример. Проекция на гиперплоскость

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

Решение

Лагранжиан:

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

Производная L по $\mathbf{x}$:

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

Пример. Проекция на гиперплоскость

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

Решение

Лагранжиан:

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

Производная L по $\mathbf{x}$:

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

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

Пример. Проекция на гиперплоскость

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

Решение

Лагранжиан:

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

Производная L по $\mathbf{x}$:

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

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

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

Пример. Проекция на единичный симплекс

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

Пример. Проекция на единичный симплекс

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

Условия ККТ

Лагранжиан задается следующим образом:

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

Пример. Проекция на единичный симплекс

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

Условия ККТ

Лагранжиан задается следующим образом:

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

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

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

Пример. Проекция на единичный симплекс

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

Условия ККТ

Лагранжиан задается следующим образом:

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

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

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

Пример. Проекция на единичный симплекс

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

Условия ККТ

Лагранжиан задается следующим образом:

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

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

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

Пример. Проекция на единичный симплекс

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

Условия ККТ

Лагранжиан задается следующим образом:

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

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

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

Пример. Проекция на единичный симплекс

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

Условия ККТ

Лагранжиан задается следующим образом:

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

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

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

Пример. Проекция на единичный симплекс

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

Условия ККТ

Лагранжиан задается следующим образом:

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

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

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

Question

Решите систему выше за $O(n \log n)$.

Пример. Проекция на единичный симплекс

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

Условия ККТ

Лагранжиан задается следующим образом:

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

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

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

Question

Решите систему выше за $O(n \log n)$.

Question

Решите систему выше за $O(n)$.

Ссылки

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

Ссылки

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

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.