Семинар по теме «Оптимизация»

Материал: Панин Никита (telegram: @nickzooot)
Конспект: Иванов Егор (telegram: @e1vanov)

5 семестр

Содержание

1 Наводящие соображения

Постановка задачи минимизации выглядит следующим образом:

$$f(\mathbf{x}) \to \inf_{X},\ X \subseteq \mathbb{R}^n$$

Задачи оптимизации можно разделить на 2 типа:

- *Безусловная* оптимизация: $X = \mathbb{R}^n$
- *Условная* оптимизация: $X \subset \mathbb{R}^n$

Пусть $\mathbf{v}$ – точка локального минимума, тогда возможны 2 случая:

- Если $\mathbf{v}$ – внутренняя точка X ($\mathbf{v} \in \operatorname{int} X$), то можем выписать необходимые условия локального минимума:
$$\begin{cases} \nabla f(\mathbf{v}) = 0 \\ \nabla^2 f(\mathbf{v}) \geq 0 \end{cases}$$
- Если же $\mathbf{v}$ – граничная точка X ($\mathbf{v} \in \partial X$), то все становится сложнее:

  Пример. $f(x) = -x^2$, $X = [1, 2]$. Минимум достигается в точке $x = 2$, при этом производная $f'(x) = -2x$ нигде на X в ноль не обращается.

  Отсюда делаем вывод, что задача условной оптимизации требует более глубокого анализа.

2 Необходимое и достаточное условия локального минимума в задачах условной минимизации

Выпишем постановку задачи

$$f(x) \to \inf_X, \tag{2.1}$$
$$X = \{\mathbf{x} \in X_0 \subseteq \mathbb{R}^n : \underbrace{g_1(\mathbf{x}) \leq 0, \ldots, g_m(\mathbf{x}) \leq 0}_{\text{ограничения типа неравенств}}, \underbrace{g_{m+1}(\mathbf{x}) = 0, \ldots, g_s(\mathbf{x}) = 0}_{\text{ограничения типа равенств}}\}$$

Замечание 1. $\operatorname{dom} f = \operatorname{dom} g_i = X_0$
Замечание 2. Допускаются следующие ситуации:

- $m = 0$: отсутствуют ограничения типа неравенств
- $m = s$: отсутствуют ограничения типа равенств
- $m = s = 0$: задача безусловной оптимизации

Замечание 3. Будем далее везде считать, что X_0 – выпуклое множество
Замечание 4. Говорят, что задача (2.1) записана в *каноническом* виде

Определение. *Функцией Лагранжа* задачи (2.1) называется следующий функционал:

$$\mathscr{L}(\mathbf{x}, \boldsymbol{\lambda}) = \lambda_0 f(\mathbf{x}) + \sum_{j=1}^{s} \lambda_j g_j(\mathbf{x}),$$

где $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_s) \in \mathbb{R}^{s+1}$, $\lambda_i \geq 0 \ \forall\, i = 1, \ldots, m$, $\mathbf{x} \in X_0$.

2.1 Необходимое условие

Теорема 2.1 (Правило множителей Лагранжа в общем виде, [1] – стр. 211)**.** Пусть $\mathbf{x}_*$ – точка локального минимума функции на X из задачи (2.1) и

1. $f, g_1, \ldots, g_m$ дифференцируемы в $\mathbf{x}_*$
2. $g_{m+1}, \ldots, g_s$ непрерывно дифференцируемы в $B(\mathbf{x}_*, \varepsilon) \cap X_0$ – некоторой окрестности точки $\mathbf{x}_*$, пересеченной с X_0

Тогда существуют $\boldsymbol{\lambda}^* = (\lambda_0^*, \ldots, \lambda_s^*)$ такие, что:

(a) $\boldsymbol{\lambda}^* \neq 0$

(b) $\lambda_i^* \geq 0,\ i = 0, 1 \ldots, m$

(c) $\langle \nabla_{\mathbf{x}} \mathscr{L}(\mathbf{x}_*, \boldsymbol{\lambda}^*), \mathbf{x} - \mathbf{x}_* \rangle \geq 0\ \forall\, \mathbf{x} \in X_0$

(d) $\lambda_i^* g_i(\mathbf{x}_*) = 0,\ i = 1, 2, \ldots, m$

Определение. $\lambda_0^*, \ldots, \lambda_s^*$ называются множителями Лагранжа, если они удовлетворяют условиям (a) - (d).

Определение. Условие пункта (c) называется *вариационным неравенством.*

Определение. Условия пункта (d) называются *условиями дополняющей нежесткости.*

Замечание 1. Точками локального минимума функции f на X могут быть лишь те точки $\mathbf{v} \in X_0$, для которых существуют множители Лагранжа $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_s)$ такие, что пара $(\mathbf{v}, \boldsymbol{\lambda})$ является решением следующей системы (в условиях, взятых из теоремы 2.1 следует заменить $\mathbf{x}_*$ на $\mathbf{v}$ и $\boldsymbol{\lambda}^*$ на $\boldsymbol{\lambda}$):

$$\begin{cases} \text{Выполнены условия (a), (b), (c), (d) теоремы 2.1} \\ \text{(e) } g_i(\mathbf{v}) \leq 0,\ i = 1, 2, \ldots, m \\ \text{(f) } g_i(\mathbf{v}) = 0,\ i = m+1, \ldots, s \end{cases} \tag{2.2}$$

Замечание 2. Если $\mathbf{x}_*$ – внутренняя точка X_0, то вариационное неравенство равносильно условию

$$\nabla_{\mathbf{x}} \mathscr{L}(\mathbf{x}_*, \boldsymbol{\lambda}^*) = 0$$

В частности, если X_0 – открытое множество, то это выполнено автоматически.

Пусть $\mathbf{v}$ – точка локального минимума f на X. Множество всех таких $\boldsymbol{\lambda}$, при которых $(\mathbf{v}, \boldsymbol{\lambda})$ – решение системы (2.2) будем обозначать $\Lambda(\mathbf{v})$.
Замечание 3. $\Lambda(\mathbf{v})$ – конус, так как верна следующая импликация:

$$\begin{cases} (\mathbf{v}, \boldsymbol{\lambda}) \in \Lambda(\mathbf{v}) \\ 0 < \alpha \in \mathbb{R} \end{cases} \Rightarrow (\mathbf{v}, \alpha\boldsymbol{\lambda}) \in \Lambda(\mathbf{v})$$

Этот конус будем называть *конусом Лагранжа* точки $\mathbf{v}$. Кроме того, это наблюдение позволяет, не умаляя общности, при условии положительности λ_0 считать его равным 1.

Определение. Ограничение $g_i(\mathbf{x}) \leq 0$ называется *активным* в точке $\mathbf{v}$, если $g_i(\mathbf{v}) = 0$, и *пассивным (или неактивным)* в противном случае.

Замечание 4. Ограничения типа равенств активны в любой точке X.

2.2 Достаточное условие

Теорема 2.2 (достаточное условие оптимальности, [2] – стр. 47)**.** Пусть в задаче (2.1)

1. $f, g_1, \ldots, g_s$ – непрерывно дифференцируемы в некоторой окрестности точки $\mathbf{x}_* \in X$

2. $f, g_1, \ldots, g_s$ – дважды дифференцируемы в $\mathbf{x}_*$

3. $(\mathbf{x}_*, \boldsymbol{\lambda}^*)$ – решение системы (2.2)

4. $\langle \nabla^2_{\mathbf{xx}} \mathscr{L}(\mathbf{x}_*, \boldsymbol{\lambda}^*)\,\mathbf{h}, \mathbf{h}\rangle > 0 \ \ \forall\, \mathbf{h} \neq 0 \in \mathbb{R}^n$, удовлетворяющих следующим условиям:

   4.a $\lambda_0^* \langle \nabla f(\mathbf{x}_*), \mathbf{h}\rangle \leq 0$

   4.b $\langle \nabla g_j(\mathbf{x}_*), \mathbf{h}\rangle \leq 0$ для всех активных в точке $\mathbf{x}_*$ ограничений

   4.c $\langle \nabla g_j(\mathbf{x}_*), \mathbf{h}\rangle = 0 \ \forall\, j = m+1, \ldots, s$

Тогда в $\mathbf{x}_*$ реализуется строгий локальный минимум.

Замечание 1. Если $X_0 = \mathbb{R}^n$ и $m = 0$, то условия 4.a и 4.b проверять не нужно.
Замечание 2. Для исследования на локальный максимум во всех сформулированных выше утверждениях нужно заменить условие $\lambda_0 \geq 0$ на $\lambda_0 \leq 0$. Это очевидным образом следует из того, что точки максимума для $f(x)$ являются точками минимума $-f(x)$.

3 Примеры решения задач

Алгоритм решения задач на поиск экстремумов:

1. проверить, что задача записана в каноническом виде

2. выписать функцию Лагранжа

3. найти подозрительные на экстремум точки с помощью теоремы 2.1

4. перебирая $\lambda_0 = -1, 0, 1$, пытаться воспользоваться теоремой 2.2, если не получается, то пытаться на основе неравенств доказать является ли точка локальным минимумом или локальным максимумом $\big($то есть, если локальный минимум(локальный максимум), то для некоторой окрестности должно быть выполнено

   $f\big($подозрительная на минимум(максимум) точка$) \leq (\geq)\ f($точки некоторой окрестности$)\big)$

3.1 Задача

$$\begin{cases} f(\mathbf{u}) \equiv f(x,y) = x \\ X_0 = \mathbb{R}^2,\ g_1(x,y) = x^2 + y^2 - 1 = 0,\ g_2(x,y) = x^3 + y^3 - 1 = 0 \end{cases}$$

Решение. Выпишем функцию Лагранжа

$$\mathscr{L}(\mathbf{u}, \boldsymbol{\lambda}) = \lambda_0 x + \lambda_1(x^2 + y^2 - 1) + \lambda_2(x^3 + y^3 - 1)$$

Найдем ее градиент и гессиан по пространственным координатам:

$$\nabla_{\mathbf{u}} \mathscr{L}(\mathbf{u}, \boldsymbol{\lambda}) = (\lambda_0 + 2\lambda_1 x + 3\lambda_2 x^2, 2\lambda_1 y + 3\lambda_2 y^2)$$

$$\nabla^2_{\mathbf{uu}}\mathscr{L}(\mathbf{u},\boldsymbol{\lambda}) = \begin{bmatrix} 2\lambda_1 + 6\lambda_2 x & 0 \\ 0 & 2\lambda_1 + 6\lambda_2 y \end{bmatrix}$$

Выпишем часть системы вида (2.2):

$$\begin{cases} \lambda_0 + 2\lambda_1 x + 3\lambda_2 x^2 = 0 \\ 2\lambda_1 y + 3\lambda_2 y^2 = 0 \\ x^2 + y^2 = 1 \\ x^3 + y^3 = 1 \end{cases}$$

Несложно убедиться в том, что последним двум уравнениям системы удовлетворяют только точки $\mathbf{u}_1 = (0, 1)$ и $\mathbf{u}_2 = (1, 0)$. Рассмотрим их по очереди.

1. $\mathbf{u}_1 = (0, 1)$:

$$\begin{cases} \lambda_0 = 0 \\ 2\lambda_1 + 3\lambda_2 = 0 \end{cases}$$

   В качестве решения данной системы рассмотрим $\boldsymbol{\lambda}_1 = (0, 3, -2)$

   Найдем $\mathbf{h} \in \mathbb{R}^2 \neq 0$, для которых будем проверять положительную определенность гессиана

$$\begin{cases} \langle \nabla g_1(\mathbf{u}_1), \mathbf{h} \rangle = 0 \\ \langle \nabla g_2(\mathbf{u}_1), \mathbf{h} \rangle = 0 \end{cases} \Leftrightarrow \begin{cases} 2h_2 = 0 \\ 3h_2 = 0 \end{cases} \Leftrightarrow h_2 = 0 \Rightarrow \mathbf{h} = (h_1, 0), h_1 \neq 0$$

   Проверим положительную определенность гессиана:

$$\langle \nabla^2_{\mathbf{uu}}\mathscr{L}(\mathbf{u}_1, \boldsymbol{\lambda}_1)\, \mathbf{h}, \mathbf{h} \rangle = \left\langle \begin{bmatrix} 6 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} h_1 \\ 0 \end{bmatrix}, \begin{bmatrix} h_1 \\ 0 \end{bmatrix} \right\rangle = 6h_1^2 > 0$$

   Так как $\lambda_0 = 0$, то $\mathbf{u}_1$ – точка локального как минимума, так и максимума. Функция в ее окрестности не константа, а значит это изолированная точка множества X.

2. $\mathbf{u}_2 = (1, 0)$ – рассматривается полностью аналогично.

3.2 Задача

$$\begin{cases} f(\mathbf{u}) \equiv f(x, y) = x^2 + y^2 \\ X_0 = \{x \geq 1, y \geq 0\},\ g(x, y) = x + y - 2 \leq 0 \end{cases}$$

Решение. Выпишем функцию Лагранжа

$$\mathscr{L}(\mathbf{u},\boldsymbol{\lambda}) = \lambda_0(x^2 + y^2) + \lambda(x + y - 2)$$

Найдем ее градиент по пространственным координатам:

$$\nabla_{\mathbf{u}}\mathscr{L}(\mathbf{u},\boldsymbol{\lambda}) = (2\lambda_0 x + \lambda, 2\lambda_0 y + \lambda)$$

Также выпишем гессиан по пространственным координатам:

$$\nabla^2_{\mathbf{uu}}\mathscr{L}(\mathbf{u},\boldsymbol{\lambda}) = \begin{bmatrix} 2\lambda_0 & 0 \\ 0 & 2\lambda_0 \end{bmatrix} = 2\lambda_0 I$$

Теперь воспользуемся необходимым условием. Если $\mathbf{u}$ – точка локального экстремума, то

$$\begin{cases} (2\lambda_0 x + \lambda)(v_1 - x) + (2\lambda_0 y + \lambda)(v_2 - y) \geq 0 \;\forall\, \mathbf{v} = (v_1, v_2) \in X_0 \\ \lambda(x + y - 2) = 0 \\ \boldsymbol{\lambda} = (\lambda_0, \lambda) \neq 0 \\ \lambda \geq 0 \end{cases} \tag{3.2.1}$$

Рассмотрим несколько случаев:

1. $x > 1, y > 0$. Имеем следующую систему:

$$\begin{cases} 2\lambda_0 x + \lambda = 2\lambda_0 y + \lambda = 0 \\ \lambda(x + y - 2) = 0 \\ \boldsymbol{\lambda} = (\lambda_0, \lambda) \neq 0 \\ \lambda \geq 0 \end{cases}$$

   (a) $\lambda_0 = 0 \Rightarrow \lambda = 0$ – противоречие

   (b) $\lambda_0 \neq 0$

      i. $\lambda = 0 \Rightarrow 2\lambda_0 x = 2\lambda_0 y = 0 \Rightarrow x = y = 0$ – противоречие

      ii. $\lambda \neq 0 \Rightarrow x + y = 2$. Из первой строки системы:

$$x = y = -\frac{\lambda}{2\lambda_0} \Rightarrow -\frac{\lambda}{\lambda_0} = 2 \Rightarrow \lambda = -2\lambda_0 \Rightarrow x = 1 \text{ – противоречие}$$

2. $x = 1, y > 0$. В первом утверждении системы (3.2.1) рассмотрим $\mathbf{v} = (1, v_2), v_2 \geq 0$, что позволяет перейти к следующей системе:

$$\begin{cases} 2\lambda_0 y + \lambda = 0 \\ \lambda(x + y - 2) = 0 \\ \boldsymbol{\lambda} = (\lambda_0, \lambda) \neq 0 \\ \lambda \geq 0 \end{cases}$$

   (a) $\lambda_0 = 0 \Rightarrow \lambda = 0$ – противоречие

   (b) $\lambda_0 \neq 0$

      i. $\lambda = 0 \Rightarrow y = 0$ – противоречие

      ii. $\lambda \neq 0 \Rightarrow y = 1 \Rightarrow \lambda = -2\lambda_0 > 0 \Rightarrow \lambda_0 < 0$ – точка $\mathbf{u}_1 = (1, 1)$ подозрительная на максимум

3. $x > 1, y = 0$. Аналогично рассуждениям предыдущего пункта, получаем:

$$\begin{cases} 2\lambda_0 x + \lambda = 0 \\ \lambda(x + y - 2) = 0 \\ \boldsymbol{\lambda} = (\lambda_0, \lambda) \neq 0 \\ \lambda \geq 0 \end{cases}$$

   (a) $\lambda_0 = 0 \Rightarrow \lambda = 0$ – противоречие

   (b) $\lambda_0 \neq 0$

   i. $\lambda = 0 \Rightarrow x = 0$ – противоречие
   ii. $\lambda \neq 0 \Rightarrow x = 2 \Rightarrow \lambda = -4\lambda_0 > 0 \Rightarrow \lambda_0 < 0$ – точка $\mathbf{u}_2 = (2, 0)$ подозрительная на максимум

4. $x = 1, y = 0$. Из второго уравнения системы (3.2.1) следует, что $\lambda = 0$, тогда первое переписывается следующим образом:

$$2\lambda_0(v_1 - 1) \geq 0 \, \forall \, v_1 \geq 1$$

Отсюда получаем, что $\lambda_0 > 0$, а значит точка $\mathbf{u}_3 = (1, 0)$ подозрительная на минимум

Подробнее рассмотрим полученные подозрительные точки:

1. $\mathbf{u}_1 = (1, 1)$. Рассмотрим две последовательности точек в X:

$$\mathbf{u}_k^1 = \left(1, 1 - \frac{1}{k}\right) \xrightarrow{k \to \infty} (1, 1)$$

$$\mathbf{u}_k^2 = \left(1 + \frac{1}{k}, 1 - \frac{1}{k}\right) \xrightarrow{k \to \infty} (1, 1)$$

При этом

$$f(\mathbf{u}_k^1) = 1 + \left(1 - \frac{1}{k}\right)^2 < 2 = f(1, 1) < f(\mathbf{u}_k^2) = \left(1 + \frac{1}{k}\right)^2 + \left(1 - \frac{1}{k}\right)^2 = 2 + \frac{2}{k^2}$$

Отсюда делаем вывод, что $\mathbf{u}_1 = (1, 1)$ не является точкой локального экстремума.

2. $\mathbf{u}_2 = (2, 0)$. Заметим, что в силу отрицательности λ_0 гессиан отрицательно определен, поэтому теорема 2.2 неприменима. Проведем дополнительный анализ:

$$y \leq 2 - x \Rightarrow y^2 \leq (x - 2)^2 \Rightarrow f(x, y) = x^2 + y^2 \leq x^2 + (x - 2)^2 = 2x^2 - 4x + 4 =$$

$$= 2(x^2 - 2x + 2) = 2((x - 1)^2 + 1) \leq \{1 \leq x \leq 2\} \leq 4 = f(\mathbf{u}_2)$$

Отсюда получаем, что $\mathbf{u}_2$ – точка глобального максимума.

3. $\mathbf{u}_3 = (1, 0)$. Заметим, что в силу положительности λ_0 гессиан положительно определен, поэтому применима теорема 2.2, а значит точка $\mathbf{u}_3$ – точка локального минимума. Уточним результат:

$$f(x, y) = x^2 + y^2 \geq \{x \geq 1, y \geq 0\} \geq 1^2 + 0^2 = 1 = f(\mathbf{u}_3)$$

Отсюда получаем, что $\mathbf{u}_3$ – точка глобального минимума.

Ответ: $\mathbf{u}_{\min} = (1, 0)$, $\mathbf{u}_{\max} = (2, 0)$

4 Условия Каруша-Куна-Таккера

Доказательства всех теорем из данного раздела можно найти в [3].

Определение. Задачу (2.1) называют *регулярной*, если $\lambda_0^* > 0$. Тут имеется в виду, что для всех пар $(\mathbf{x}_*, \boldsymbol{\lambda}^*)$, удовлетворяющих правилу множителей Лагранжа $\lambda_0^* > 0$. Тогда функция Лагранжа называется *нормальной* и имеет вид:

$$\mathscr{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(x) + \sum_{i=1}^{s} \lambda_i g_i(\mathbf{x}),$$

где $\mathbf{x} \in X_0$, а $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_s) \in \Lambda_0 = \{(\lambda_1, \dots, \lambda_s) \in \mathbb{R}^s : \lambda_i \geq 0, i = 1, \dots, m\}$

Такой вид функции Лагранжа следует из желания упростить общий вид функции Лагранжа, воспользовавшись **замечанием 3 из 2.1 параграфа.**

Определение. *Условиями Каруша-Куна-Таккера* (далее: условия ККТ) называются условия теоремы 2.1 в случае регулярной задачи.

Определение. Множество $X \subseteq \mathbb{R}^n$ называется *полиэдром*, если оно представляет собой решение системы линейных неравенств $A\mathbf{x} \leq \mathbf{b}$, где $A \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m$:

$$\mathbf{x} \in X \Leftrightarrow \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_i \; \forall\, i = 1, \dots, m$$

Замечание. В системе могут быть и равенства:

$$\langle \mathbf{a}_i, \mathbf{x} \rangle = b_i \Leftrightarrow \begin{cases} \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_i \\ -\langle \mathbf{a}_i, \mathbf{x} \rangle \leq -b_i \end{cases}$$

Примеры полиэдров: $X = \mathbb{R}^n, X = \mathbb{R}^n_+, X = \{\mathbf{x} = (x_1, \dots, x_n) : x_i \geq 0 \; \forall\, i \in I \subseteq \{1, \dots, n\}\}$
$X = \{\mathbf{x} = (x_1, \dots, x_n) : \alpha_i \leq x_i \leq \beta_i, \alpha_i, \beta_i \in [-\infty, +\infty] \; \forall\, i = 1, \dots, n\}$

Определение. *Аффинной оболочкой* множества $C \subseteq \mathbb{R}^n$ называется множество

$$\operatorname{aff} C = \left\{ \sum_{j=1}^{k} \theta_j \mathbf{x}_j : \; \forall\, k \in \mathbb{N}, \mathbf{x}_j \in C, \theta_j \in \mathbb{R} \;\; \forall\, j = 1, \dots, k, \sum_{j=1}^{k} \theta_j = 1 \right\}$$

Определение. Множество $C \subseteq \mathbb{R}^n$ называется *аффинным*, если верна следующая импликация:

$$\begin{cases} \forall\, \mathbf{x}_1, \mathbf{x}_2 \in C \\ \theta \in \mathbb{R} \end{cases} \Rightarrow \theta \mathbf{x}_1 + (1 - \theta)\mathbf{x}_2 \in C$$

Другими словами, оно должно быть замкнуто относительно операции взятия аффинной комбинации.

Замечание. Аффинная оболочка множества C – минимальное аффинное множество, содержащее множество C, или

$$\forall\, S \text{ – такое аффинное множество, что } C \subseteq S : \operatorname{aff}(C) \subseteq S$$

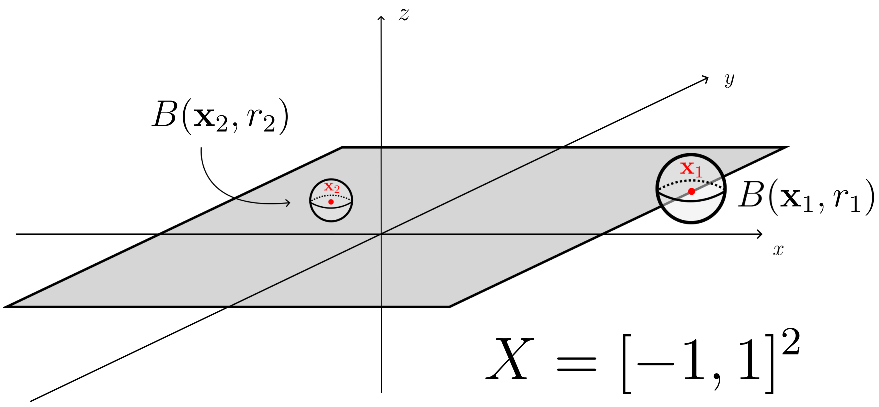

Рис. 1: Пример

Определение. *Относительной внутренностью (relative interior)* множества $C \subseteq \mathbb{R}^n$ называется следующее множество:

$$\operatorname{relint} C = \{\mathbf{x} \in C : \exists\, r > 0\; B(\mathbf{x}, r) \cap \operatorname{aff} C \subseteq C\}$$

Замечание. Если $\operatorname{int} C \neq \varnothing \Rightarrow \operatorname{int} C = \operatorname{relint} C$, в частности для открытого множества это значит, что $\operatorname{relint} C = C$, например, $\operatorname{relint} \mathbb{R}^n = \mathbb{R}^n$.
Пример. Чтобы лучше разобраться с понятием относительной внутренности множества, предлагается рассмотреть следующий пример (Рис. 1). Пусть

$$X = \{\mathbf{x} \equiv (x, y, z) \in \mathbb{R}^3 : |x| \leq 1, |y| \leq 1, z = 0\}$$

Тогда несложно убедиться в том, что $\operatorname{aff} X$ это плоскость Oxy. Поймем, что есть $\operatorname{relint} X$. Если точка множества X лежит на его границе, как $\mathbf{x}_1$ на рисунке, тогда какой бы шар мы не взяли, в его пересечении с плоскостью Oxy будут как точки из X, так и точки ему не принадлежащие, поэтому $\partial X \notin \operatorname{relint} X$. Если же рассмотреть внутреннюю точку множества X, как $\mathbf{x}_2$ на рисунке, то несложно убедиться в том, что требуемый в определении шар существует. Итак

$$\operatorname{relint} X = \{\mathbf{x} \equiv (x, y, z) \in \mathbb{R}^3 : |x| < 1, |y| < 1, z = 0\}$$

Теорема 4.1 (условия регулярности)**.** Пусть $\mathbf{x}_*$ – точка локального минимума функции f на X и

1. $f, g_1, \ldots, g_m$ – дифференцируемы в $\mathbf{x}_*$
2. $g_1, \ldots, g_m$ – выпуклы на X_0
3. $g_{m+1}, \ldots, g_s$ – линейны, то есть $g_i(\mathbf{x}) = \langle \mathbf{a}_i, \mathbf{x} \rangle - b_i\; \forall\, i = m+1, \ldots, s$

Тогда, если выполнено хотя бы одно из условий:

(a) $m = s$ (нет ограничений типа равенств) и $\exists\, \mathbf{x} \in X_0 : g_i(\mathbf{x}) < 0\; \forall\, i = 1, \ldots, m$. Иными словами, существует такая точка множества X_0, что все ограничения типа неравенств в ней пассивны.

(b) X_0 – полиэдр, $g_1, \ldots, g_m$ – линейны

(c) X_0 – полиэдр, часть ограничений типа неравенств линейны, а для оставшихся существует точка $\hat{\mathbf{x}} \in \mathbf{X}$, в которой они все пассивны, то есть: $\exists\, l : 0 \le l \le m : g_{l+1}, \ldots, g_m$ – линейны и $\exists\, \hat{\mathbf{x}} \in X : g_i(\hat{\mathbf{x}}) < 0 \,\forall\, i = 1, \ldots, l$

(d) $\exists\, l : 0 \le l \le m : g_{l+1}, \ldots, g_m$ – линейны и $\exists\, \hat{\mathbf{x}} \in (\mathrm{relint}\, X_0) \cap X : g_i(\hat{\mathbf{x}}) < 0 \,\forall\, i = 1, \ldots, l$

То существует $\boldsymbol{\lambda}^* = (1, \lambda_1, \ldots, \lambda_s)$ такой, что для $(\mathbf{x}_*, \boldsymbol{\lambda}^*)$ выполнены условия ККТ.

Определение. В теореме 4.1 условие (a) называется *условием Слейтера (SC)* , (b) – *условием линейности (LCQ)*, (c) и (d) – *модифицированными условиями Слейтера.*

Замечание 1. Ограничений типа равенств или неравенств может и не быть.
Замечание 2. Если $X_0 = \mathbb{R}^n$, то условия (c) и (d) эквивалентны.

Определение. Задача (2.1) называется *выпуклой*, если

1. $f, g_1, \ldots, g_m$ – выпуклы
2. $g_{m+1}, \ldots, g_s$ – линейны
3. X_0 – выпуклое множество

Теорема 4.2. (Теорема Куна-Таккера в дифференциальной форме) Пусть исходная задача выпукла. Тогда если для некоторого набора переменных $(\mathbf{x}_*, \boldsymbol{\lambda}^*)$ выполнены все условия ККТ, то $\mathbf{x}_*$ – точка глобального минимума.

5 Примеры решения задач

5.1 Задача

$$\begin{cases} f(\mathbf{u}) \equiv f(x,y) = (x-4)^2 + (y-4)^2 \\ X_0 = \mathbb{R}^2,\; g_1(x,y) = x + y - 4 \le 0, g_2(x,y) = x + 3y - 0 \le 0 \end{cases}$$

Решение. Заметим, что задача выпуклая, а значит условия ККТ – необходимые и достаточные. Составим функцию Лагранжа:

$$\mathscr{L}(\mathbf{u}, \boldsymbol{\lambda}) = (x-4)^2 + (y-4)^2 + \lambda_1(x + y - 4) + \lambda_2(x + 3y - 9)$$

Найдем ее градиент по пространственным координатам:

$$\nabla_{\mathbf{u}} \mathscr{L}(\mathbf{u}, \boldsymbol{\lambda}) = (2(x-4) + \lambda_1 + \lambda_2, 2(y-4) + \lambda_1 + 3\lambda_2)$$

Выпишем условия ККТ:

$$\begin{cases} 2(x-4) + \lambda_1 + \lambda_2 = 0 \\ 2(y-4) + \lambda_1 + 3\lambda_2 = 0 \\ x + y \le 4 \\ x + 3y \le 9 \\ \lambda_1 \ge 0 \\ \lambda_2 \ge 0 \\ \lambda_1(x + y - 4) = 0 \\ \lambda_2(x + 3y - 9) = 0 \end{cases}$$

Последовательно перебираем варианты равенства нулю и положительности для λ_1 и λ_2:

1. $\lambda_1 > 0, \lambda_2 > 0$. Тогда из двух последних условий:
$$\begin{cases} x + y - 4 = 0 \\ x + 3y - 9 = 0 \end{cases} \Leftrightarrow \begin{cases} x = \dfrac{3}{2} \\ y = \dfrac{5}{2} \end{cases}$$
Подставляем полученные значения в первые уравнения:
$$\begin{cases} 2\left(\dfrac{3}{2} - 4\right) + \lambda_1 + \lambda_2 = 0 \\ 2\left(\dfrac{5}{2} - 4\right) + \lambda_1 + 3\lambda_2 = 0 \end{cases} \Leftrightarrow \begin{cases} \lambda_1 = 6 \\ \lambda_2 = -1 \end{cases}$$
Получили противоречие с условием $\lambda_2 > 0$.
2. $\lambda_1 = \lambda_2 = 0 \Rightarrow x = y = 4$ – противоречие
3. $\lambda_1 > 0, \lambda_2 = 0$. Тогда первые 2 уравнения и предпоследнее преобразуются следующим образом:
$$\begin{cases} 2(x-4) + \lambda_1 = 0 \\ 2(y-4) + \lambda_1 = 0 \\ x + y - 4 = 0 \end{cases} \Leftrightarrow \begin{cases} x = y = 2 \\ \lambda_1 = 4 \end{cases}$$
Получили решение.
4. $\lambda_1 = 0, \lambda_2 > 0$. Аналогично предыдущему пункту получаем следующую систему:
$$\begin{cases} 2(x-4) + \lambda_2 = 0 \\ 2(y-4) + 3\lambda_2 = 0 \\ x + 3y - 9 = 0 \end{cases} \Leftrightarrow \begin{cases} x = \dfrac{33}{10} \\ y = \dfrac{19}{10} \\ \lambda_2 = \dfrac{14}{10} \end{cases}$$
Получили противоречие с условием $x + y \leq 4$.

Ответ: глобальный минимум достигается в точке $\mathbf{u}_* = (2, 2)$.

5.2 Задача

$$\begin{cases} f(\mathbf{x}) = \dfrac{1}{2}\|\mathbf{x}\|_2^2 \\ X_0 = \mathbb{R}^n, A\mathbf{x} - \mathbf{b} = 0 \end{cases}$$

Решение. Задача выпуклая, а значит условия ККТ – необходимые и достаточные. Составим функцию Лагранжа:

$$\mathscr{L}(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{x}\|_2^2 + \boldsymbol{\lambda}^T(A\mathbf{x} - \mathbf{b})$$

Вычислим ее градиент по пространственным координатам:

$$\nabla_{\mathbf{x}}\mathscr{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x} + A^T\boldsymbol{\lambda}$$

Выпишем условия ККТ:

$$\begin{cases} \mathbf{x} + A^T\boldsymbol{\lambda} = 0 \\ A\mathbf{x} - \mathbf{b} = 0 \end{cases} \Leftrightarrow \begin{cases} \boldsymbol{\lambda} = -(AA^T)^{-1}\mathbf{b} \\ \mathbf{x} = A^T(AA^T)^{-1}\mathbf{b} \end{cases}$$

Ответ: глобальный минимум достигается в точке $\mathbf{x}_* = A^T(AA^T)^{-1}\mathbf{b}$.

5.3 Задача

5.3.1 Интерпретация энтропии

Допустим есть конечное множество событий $\{A_i\}$, каждому из которых соответствует вероятность $\mathsf{P}(A_i)$. Мы хотим оценить «удивление» от наступления некоторого события. Для этого нужна функция, которая возрастает при убывании вероятности события. Действительно, чем меньше вероятность события, тем для нас удивительнее его пронаблюдать. Возьмем в качестве такой функции $-\log \mathsf{P}(A_i)$.

Как в таком случае посчитать среднее удивление от наблюдаемой случайной величины? Взять математическое ожидание от $-\log \mathsf{P}(A)$, где A – случайная величина, принимающая индекс события, т.е. $-\sum_i P(A_i)\log P(A_i)$ – это и есть энтропия, ее еще называют энтропией Шеннона.

Почему энтропия является мерой хаоса? Чем больше хаоса, тем больше «удивление» пронаблюдать какое-то событие в такой системе. Отсюда и связь.
Максимизируем энтропию Шеннона для дискретной случайной величины, принимающей n значений с вероятностями $p_1, \ldots, p_n$:

$$\begin{cases} \mathrm{H}(\mathbf{p}) \equiv \mathrm{H}(p_1, \ldots, p_n) = -\sum\limits_{i=1}^n p_i \log p_i \to \max \\ X_0 = \mathbb{R}^n, g_i(\mathbf{p}) = -p_i \le 0 \;\forall\, i = 1, \ldots, n \\ \sum\limits_{i=1}^n p_i = 1 \end{cases}$$

Решение. Выпишем функцию Лагранжа:

$$\mathscr{L}(\mathbf{p}, \lambda, \boldsymbol{\mu}) = \sum_{i=1}^n p_i \log p_i + \lambda\left(\sum_{i=1}^n p_i - 1\right) - \sum_{i=1}^n \mu_i p_i$$

Найдем ее градиент по пространственным (вероятностным) координатам:

$$\nabla_{\mathbf{p}}\mathscr{L}(\mathbf{p}, \lambda, \boldsymbol{\mu}) = (\log p_i + 1 + \lambda - \mu_i)_{i=1}^n$$

Запишем условия ККТ:

$$\begin{cases} \log p_i + 1 + \lambda - \mu_i = 0 \;\forall\, i = 1, \ldots, n \\ p_i \ge 0 \;\forall\, i = 1, \ldots, n \\ \mu_i \ge 0 \;\forall\, i = 1, \ldots, n \\ \mu_i p_i = 0 \;\forall\, i = 1, \ldots, n \\ \sum\limits_{i=1}^n p_i = 1 \end{cases}$$

Из первой и четвертой группы уравнений получаем:

$$p_i = \exp\{\mu_i - \lambda - 1\} > 0 \Rightarrow \mu_i = 0 \Rightarrow p_i = \exp\{-\lambda - 1\}\, \forall\, i = 1, \dots, n$$

Используем последнее уравнение системы:

$$n \exp\{-\lambda - 1\} = 1 \Rightarrow \exp\{-\lambda - 1\} = \frac{1}{n} \Rightarrow p_i = \frac{1}{n}$$

Ответ: энтропия максимальна при $p_1 = \dots = p_n = \frac{1}{n}$: действительно, наибольшее среднее удивление будет достигаться на равномерном распределении вероятностей.

5.4 Задача

Решим задачу проецирования на замкнутый единичный шар вектора, ему не принадлежащего:

$$\begin{cases} f(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_0\|_2^2 \\ X_0 = \mathbb{R}^n, X = \{\|\mathbf{x}\|_2 \le 1\} \end{cases}$$

Заметим, что выполнено условие Слейтера, а значит задача регулярная. Выпишем функцию Лагранжа и найдем ее градиент по пространственным координатам:

$$\mathscr{L}(\mathbf{x}, \lambda) = \|\mathbf{x} - \mathbf{x}_0\|_2^2 + \lambda(\|\mathbf{x}\|_2^2 - 1)$$

$$\nabla_{\mathbf{x}} \mathscr{L}(\mathbf{x}, \lambda) = 2(\mathbf{x} - \mathbf{x}_0) + 2\lambda\mathbf{x}$$

Выпишем условия ККТ:

$$\begin{cases} 2(\mathbf{x} - \mathbf{x}_0) + 2\lambda\mathbf{x} = 0 \\ \lambda \ge 0 \\ \|\mathbf{x}\|_2^2 \le 1 \\ \lambda(\|\mathbf{x}\|_2^2 - 1) = 0 \end{cases}$$

Рассмотрим несколько случаев:

1. $\lambda = 0 \Rightarrow \mathbf{x} = \mathbf{x}_0$, противоречие с непринадлежностью $\mathbf{x}_0$ шару

2. $\lambda > 0 \Rightarrow \|\mathbf{x}\|_2^2 = 1$. Умножим первое уравнение справа скалярно на $\mathbf{x}$:

   $$2\langle \mathbf{x} - \mathbf{x}_0, \mathbf{x}\rangle + 2\langle \lambda\mathbf{x}, \mathbf{x}\rangle = 0$$

   $$\lambda = \langle \mathbf{x}_0, \mathbf{x}\rangle - 1$$

   Подставим результат:

   $$\langle \mathbf{x}_0, \mathbf{x}\rangle \mathbf{x} = \mathbf{x}_0$$

   Получили, что вектора $\mathbf{x}$ и $\mathbf{x}_0$ коллинеарны, а значит пусть $\mathbf{x} = \alpha\mathbf{x}_0$, тогда:

   $$\alpha^2 \|\mathbf{x}_0\|_2^2 \mathbf{x}_0 = \mathbf{x}_0 \Rightarrow \alpha = \pm\frac{1}{\|\mathbf{x}_0\|_2}$$

   Случай отрицательного α не подходит из-за условия положительности λ.

Ответ: $\mathbf{x}_* = \frac{\mathbf{x}_0}{\|\mathbf{x}_0\|_2}$

6 Теория двойственности

Источник: [4]

Будем рассматривать регулярные задачи условной минимизации. Давайте предложим способ, как для любой задачи минимизации записать новую *(двойственную)* задачу максимизации такую, что изначальная оценивается ей снизу, то есть:

$$g(\mathbf{y}) \le f(\mathbf{x}) \,\forall\, \mathbf{y} \in \Omega, \mathbf{x} \in X \Rightarrow \max_{\Omega} g \le \min_{X} f$$

В чем плюсы такого подхода?

1. Исходная задача может быть «тяжело» решаемой. Возможно, двойственная окажется проще. Решив ее, мы получим оценку снизу на решение исходной.

2. Возможно, будет выполнена *сильная двойственность*, то есть $g(\mathbf{y}_*) = f(\mathbf{x}_*)$, тогда от перехода мы вовсе ничего не теряем.

3. Доказательство существования решения для исходной задачи

$$\min_X f = -\infty \Leftrightarrow \Omega = \varnothing$$

4. Получение оценки на невязку: пусть $f_* = f(\mathbf{x}_*)$, тогда

$$f_* \ge g(\mathbf{y}) \Leftrightarrow -f_* \le -g(\mathbf{y}) \Leftrightarrow f(\mathbf{x}) - f_* \le f(\mathbf{x}) - g(\mathbf{y})$$

Опишем вариант построения двойственной функции, когда

$$X = \{\mathbf{x} \in \mathbb{R}^n, g_1(\mathbf{x}) \le 0, \ldots, g_m(\mathbf{x}) \le 0, g_{m+1}(\mathbf{x}) = 0, \ldots, g_s(\mathbf{x}) = 0\}$$

Определение. *Двойственной функцией* к $f(\mathbf{x})$ назовем функцию

$$g(\lambda_1, \ldots, \lambda_s) \equiv g(\boldsymbol{\lambda}) := \inf_{\mathbb{R}^n} \mathscr{L}(\mathbf{x}, \boldsymbol{\lambda})$$

Утверждение 6.1. $f(\mathbf{x}) \ge g(\boldsymbol{\lambda}) \,\forall\, \mathbf{x} \in X, \boldsymbol{\lambda} \in \Lambda_0$

Доказательство.

$$\forall\, \mathbf{x} \in X : f(\mathbf{x}) \ge f(\mathbf{x}) + \sum_{i=1}^{m} \underbrace{\lambda_i g_i(\mathbf{x})}_{\le 0} + \sum_{j=m+1}^{s} \underbrace{\lambda_j g_j(\mathbf{x})}_{=0} = \mathscr{L}(\mathbf{x}, \boldsymbol{\lambda}) \ge \inf_{\mathbb{R}^n} \mathscr{L}(\cdot, \boldsymbol{\lambda}) = g(\boldsymbol{\lambda})$$

□

Утверждение 6.2. $g(\boldsymbol{\lambda})$ – вогнутая функция

Доказательство. Достаточно показать, что $-g(\boldsymbol{\lambda})$ – выпуклая. Это следует из линейности $\mathscr{L}$ по всем λ_i и сохранения выпуклости при переходе к sup. □

Определение. *Двойственной* задачей оптимизации для регулярной задачи (2.1) в случае $X_0 \equiv \mathbb{R}^n$ будем называть следующую задачу:

$$g(\boldsymbol{\lambda}) \to \max_{\Lambda_0},\ \Lambda_0 = \mathbb{R}^m_+ \times \mathbb{R}^{s-m}$$

Выпишем некоторые достаточные условия сильной двойственности:

1. задача выпуклая и выполнено условие Слейтера (см. (a) теоремы 4.1)
2. задача выпуклая и выполнено модифицированное условие Слейтера (см. (c) теоремы 4.1)

Замечание. Если выполнена сильная двойственность, то решение «прямой» задачи $\mathbf{x}_*$ и двойственной $\boldsymbol{\lambda}^*$ связаны:

$$\mathbf{x}_* \in \operatorname*{Argmin}_{\mathbb{R}^n} \mathscr{L}(\mathbf{x}, \boldsymbol{\lambda}^*)$$

7 Примеры решения задач

Введем обозначения: $\boldsymbol{\mu} := (\lambda_1, \dots, \lambda_m)$ и $\boldsymbol{\nu} := (\lambda_{m+1}, \dots, \lambda_s)$.

7.1 Задача

$$\begin{cases} \|\mathbf{x}\|_2^2 \to \min \\ A\mathbf{x} = \mathbf{b} \end{cases}$$

Составим функцию Лагранжа:

$$\mathscr{L}(\mathbf{x}, \boldsymbol{\nu}) = \|\mathbf{x}\|_2^2 + \boldsymbol{\nu}^T(A\mathbf{x} - \mathbf{b})$$

Найдем ее градиент по пространственным координатам:

$$\nabla_{\mathbf{x}}\mathscr{L}(\mathbf{x}, \boldsymbol{\nu}) = 2\mathbf{x} + A^T\boldsymbol{\nu}$$

Приравнивая полученное выражение к нулю, находим постановку двойственной задачи

$$\mathbf{x}_* = -\frac{1}{2}A^T\boldsymbol{\nu}$$

$$g(\boldsymbol{\nu}) = \inf_{\mathbb{R}^n} \mathscr{L}(\mathbf{x}, \boldsymbol{\nu}) = \mathscr{L}(\mathbf{x}_*, \boldsymbol{\nu}) = -\frac{1}{2}\boldsymbol{\nu}^T AA^T\boldsymbol{\nu} - \boldsymbol{\nu}^T\mathbf{b} \to \max_{\boldsymbol{\nu}\in\mathbb{R}^{\boldsymbol{s-m}}}$$

Найдем градиент функции g:

$$\nabla g = -AA^T\boldsymbol{\nu} - \mathbf{b}$$

Воспользуемся тем, что задача теперь безусловной оптимизации и приравняем нулю градиент:

$$\nabla g = -AA^T\boldsymbol{\nu} - \mathbf{b} = 0$$

Откуда в силу выполнения условий сильной двойственности, получаем

$$\mathbf{x}_* = \frac{1}{2}A^T(AA^T)^{-1}\mathbf{b}$$

7.2 Задача

Для следующей задачи найти к ней двойственную:

$$\begin{cases} \mathbf{c}^T\mathbf{x} \to \min \\ A\mathbf{x} = \mathbf{b} \\ E\mathbf{x} \leq \mathbf{d} \end{cases}$$

Для задачи выполнены модифицированное условие Слейтера. Выпишем функцию Лагранжа:

$$\mathscr{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\nu}) = \mathbf{c}^T\mathbf{x} + \boldsymbol{\mu}^T(E\mathbf{x} - \mathbf{d}) + \boldsymbol{\nu}^T(A\mathbf{x} - \mathbf{b})$$

Найдем ее градиент по пространственным координатам:

$$\nabla_{\mathbf{x}}\mathscr{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\nu}) = \mathbf{c} + E^T\boldsymbol{\mu} + A^T\boldsymbol{\nu}$$

Откуда в случае равенства нулю получаем следующую двойственную задачу:

$$\begin{cases} \boldsymbol{\mu}^T\mathbf{d} + \boldsymbol{\nu}^T\mathbf{b} \to \min \\ \boldsymbol{\mu} \geq 0 \\ \mathbf{c} + E^T\boldsymbol{\mu} + A^T\boldsymbol{\nu} = 0 \end{cases}$$

Список литературы

[1] Васильев Ф.П. *Методы оптимизации.* Факториал Пресс, 2002. ISBN: 5-88688-056-9.

[2] Будак Б.А. Артемьева Л.А. Потапов М.М. *Методы оптимизации под ред. Васильева Ф.П.* Издательство Юрайт, 2016. ISBN: 978-5-9916-6157-7.

[3] Федоров В.В. Сухарев А.Г. Тимохов А.В. *Курс методов оптимизации.* ФИЗМАТЛИТ, 2005. ISBN: 978-5-9221-0559-0.

[4] MachineLearning.ru. *Стандартные классы выпуклых задач и двойственность.* 2017. URL: https://shorturl.at/eT056.