МИСиС
Национальный исследовательский технологический университет

Методы оптимизации

Лекция 6

28 апреля 2022 г.

v. 1.01

КРИТЕРИИ ОПТИМАЛЬНОСТИ В ЗАДАЧАХ С ОГРАНИЧЕНИЯМИ

Задачи безусловной оптимизации

Рассмотрим задачу

$$f(x) \to \min, \quad x \in \mathbb{R}^n \tag{1}$$

Необходимое условие локальной оптимальности первого порядка.

Теорема 1

Пусть функция f дифференцируема в точке $x^* \in \mathbb{R}^n$. Если x^* — локальное решение задачи (1), то

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

Необходимое условие локальной оптимальности второго порядка.

Теорема 2

Пусть функция f дважды дифференцируема в точке $x^* \in \mathbb{R}^n$. Если x^* — локальное решение задачи (1), то матрица Гессе $H(x^*)$ неотрицательно определена, т. е.

$$(H(x^*)h, h) \geqslant 0 \quad \text{при всех} \quad h \in \mathbb{R}^n.$$

Достаточное условие локальной оптимальности.

Теорема 3

Пусть функция f дважды дифференцируема в точке $x^* \in \mathbb{R}^n$. Предположим, что $\nabla f(x^*) = 0$, а матрица Гессе $H(x^*)$ положительно определена. Тогда x^* — строгое локальное решение задачи (1).

Определение

Симметричная квадратная матрица A размерности $n \times n$ называется положительно (отрицательно) определённой, если $x^\mathsf{T} A x > 0$ $(x^\mathsf{T} A x < 0)$ для всех ненулевых $x \in \mathbb{R}^n$.

Критерий Сильвестра

Симметричная квадратная матрица является положительно определённой тогда и только тогда, когда все её угловые миноры Δ_i положительны. Форма отрицательно определена, тогда и только тогда, когда знаки Δ_i чередуются, причём $\Delta_1 < 0$.

Задачи безусловной оптимизации

Здесь угловыми минорами матрицы A

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

называются определители вида

$$\Delta_1 = a_{11}, \quad \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \dots,$$

$$\Delta_i = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1i} \\ a_{21} & a_{22} & \dots & a_{2i} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ii} \end{vmatrix}, \dots,$$

$$\Delta_n = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Классическая задача на условный экстремум

В курсе математического анализа изучается так называемая *классическая задача на условный экстремум*.

$$f(x) \to \min, \quad x \in X$$

где X задаётся системой конечного числа уравнений

$$X = \big\{x \in \mathbb{R}^n \;\big|\; g_i(x) = 0,\, i = 1, \ldots, m\big\}.$$

Обычно такие задачи записывают в виде

$$\begin{aligned} &\min f(x) \\ &g_i(x) = 0,\, i = 1, \ldots, m \end{aligned} \tag{2}$$

т. е. явно указывается не само допустимое множество, а система, его определяющая.

Классическая задача на условный экстремум

При исследовании задачи (2) важную роль играет её функция Лагранжа, т. е. функция

$$L(x, y_0, y) = y_0 f(x) + \sum_{i=1}^{m} y_i g_i(x)$$

где $x \in \mathbb{R}^n$, $y_0 \in \mathbb{R}$, $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$.

Частные производные этой функции по координатам вектора x имеют вид

$$\frac{\partial L}{\partial x_j}(x, y_0, y) = y_0 \frac{\partial f}{\partial x_j}(x) + \sum_{i=1}^{m} y_i \frac{\partial g_i}{\partial x_j}(x), \quad j = 1, \ldots, n$$

Составленный из них вектор обозначим через $L'_x(x, y_0, y)$, т. е.

$$L'_x(x, y_0, y) = y_0 f'(x) + \sum_{i=1}^{m} y_i g'_i(x)$$

МИСиС

Необходимое условие локальной оптимальности в задаче (2).

Теорема 4 (правило множителей Лагранжа)

Пусть функции $f, g_1, \ldots, g_m$ непрерывно дифференцируемы в некоторой окрестности точки $x^* \in \mathbb{R}^n$. Если x^* — локальное решение, то существуют число y_0^* и вектор $y^* = (y_1^*, \ldots, y_m^*)$, не равные нулю одновременно и такие, что

$$L'_x(x^*, y_0^*, y^*) = 0. \tag{3}$$

Если при этом градиенты $g'_1(x^*), \ldots, g'_m(x^*)$ линейно независимы (условие регулярности), то $y_0^* \neq 0$.

Числа $y_0^*, y_1^*, \ldots, y_m^*$ называются множителями Лагранжа. Любая точка x^*, удовлетворяющая при некоторых y_0^* и y^*, $(y_0^*, y^*) \neq 0$, условию (3), а также условию допустимости

$$g_i(x^*) = 0, \quad i = 1, \ldots, m$$

называется стационарной точкой задачи (2).

Таким образом, стационарные точки определяются системой, состоящей из $n+m$ уравнений с $n+m+1$ неизвестными.

В случае $y_0^* \neq 0$ всегда можно считать $y_0^* = 1$ (для этого следует поделить все множители Лагранжа на y_0^*). Поэтому при решении указанной системы достаточно предполагать лишь два случая: $y_0^* = 0$ или $y_0^* = 1$, и, следовательно, неизвестных в системе фактически $n+m$.

Если же выполняется указанное в теореме 4 условие регулярности, то всё сводится к случаю $y_0^* = 1$ и можно ограничиться рассмотрением функции Лагранжа вида

$$L(x,y) = L(x,1,y) = f(x) + \sum_{i=1}^{m} y_i g_i(x),$$

которую называют регулярной.

Общая задача условной оптимизации

Рассмотрим общую задачу

$$f(x) \to \min, \quad x \in X \tag{4}$$

Необходимое условие оптимальности в терминах направлений.

Вектор $h \in \mathbb{R}^n$ задаёт *направление убывания* функции f в точке $x^* \in \mathbb{R}^n$, если $f(x^* + \alpha h) < f(x^*)$ при всех достаточно малых $\alpha > 0$. Множество всех таких h обозначается через $U(x^*, f)$.

Вектор $h \in \mathbb{R}^n$ задаёт *возможное направление* относительно множества X в точке $x^* \in X$, если $x^* + \alpha h \in X$ при всех достаточно малых $\alpha > 0$. Множество всех таких h обозначим через $V(x^*, X)$.

Ниже приводится необходимое условие локальной оптимальности в задаче (4), не требующее каких-либо предположений на X и f.

Теорема 5

Если x^* — локальное решение задачи (4), то

$$U(x^*, f) \cap V(x^*, X) = \varnothing$$

Дифференциальное условие оптимальности в задаче минимизации на выпуклом множестве

Теорема 6

Пусть в задаче (4) множество X выпукло, функция f дифференцируема в точке $x^* \in X$. Тогда:

1) если x^* — локальное решение задачи (4), то

$$(f'(x^*), x - x^*) \geqslant 0 \quad \text{при всех} \quad x \in X;$$

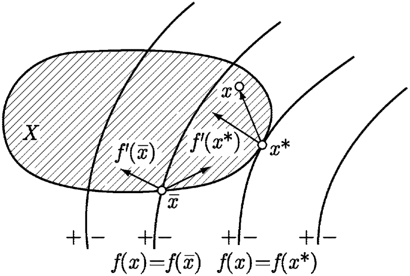
2) если функция f выпукла на X и выполняется условие 1, то x^* — (глобальное) решение задачи (4).

Таким образом, условие теоремы 6 является необходимым условием локальной оптимальности в задаче минимизации дифференцируемой функции на выпуклом множестве.

Причём, для выпуклой задачи оно оказывается и достаточным условием глобальной оптимальности.

Дифференциальное условие оптимальности в задаче минимизации на выпуклом множестве

Геометрически условие теоремы 6 означает, что градиент $f'(x^*)$ (если он отличен от нуля) составляет нетупой угол с вектором, направленным из x^* в любую точку $x \in X$.



Лемма

Если $x^* \in \operatorname{int} X$, то условие теоремы 6 эквивалентно условию:

$$f'(x^*) = 0, \quad \text{т. е.} \quad \frac{\partial f}{\partial x_j}(x^*) = 0,\ j = 1, \ldots, n.$$

Дифференциальные условия оптимальности в задаче математического программирования

Рассмотрим задачу математического программирования вида

$$
\begin{aligned}
& f(x) \to \min \\
& g_i(x) \leqslant 0, \quad i = 1, \dots, m \\
& h_i(x) = 0, \quad i = 1, \dots, l, \quad x \in P \subset \mathbb{R}^n
\end{aligned}
\tag{5}
$$

Под X будет пониматься допустимое множество задачи, т. е.

$$X = \left\{ x \in P \;\middle|\; g_i(x) \leqslant 0,\, i = 1, \dots, m;\, h_i(x) = 0,\, i = 1, \dots, l \right\}.$$

Введём функцию Лагранжа задачи (5)

$$L(x, u_0, u, v) = u_0 f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{i=1}^{l} v_i h_i(x).$$

Будем по-прежнему использовать обозначение

$$L'_x(x, u_0, u, v) = u_0 \nabla f(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) + \sum_{i=1}^{l} v_i \nabla h_i(x).$$

МИСиС

Дифференциальные условия оптимальности в задаче математического программирования

Теорема 7 (принцип Лагранжа)

Пусть в задаче (5) множество P выпукло, функции $f, g_1, \dots, g_m$ дифференцируемы в точке $x^* \in X$, функции $h_1, \dots, h_l$ непрерывно дифференцируемы в некоторой окрестности x^*. Если x^* — локальное решение задачи (5), то существуют такие числа $u_0 \geqslant 0$, $u_i \geqslant 0$, $i = 1, \dots, m$ и v_i, $i = 1, \dots, l$, не равные нулю одновременно и такие, что

$$\begin{aligned} &L'_x(x^*, u_0, u, v) = 0, \\ &u_i g_i(x) = 0,\ i = 1, \dots, m. \end{aligned} \tag{6}$$

Множители Лагранжа определены с точностью до положительной константы, т. е. можно их умножить на $1/u_0$, если $u_0 > 0$. Это позволяет без ограничения общности рассматривать в теореме 7 лишь два случая: $u_0 = 0$ или $u_0 = 1$.

Дифференциальные условия оптимальности в задаче математического программирования

В теореме 7 множитель Лагранжа, соответствующий целевой функции, не обязательно положителен. При дополнительных предположениях о множестве ограничений можно утверждать, что u_0 будет положительным (условия регулярности).

Приведённая ниже теорема Куна-Таккера является развитием необходимых условий оптимальности в теореме 7.

Теорема будет сформулирована при дополнительных предположениях относительно градиентов функций, определяющих ограничения-равенства и активные ограничения. Это гарантирует выполнение неравенства $u_0 > 0$.

Определение

Ограничения в виде неравенства $g_j(x) \geqslant 0$ называется *активным* (или *связывающим*) в точке $\bar{x}$, если $g_j(\bar{x}) = 0$, и *пассивным* (или *несвязывающим*), если $g_j(\bar{x}) > 0$.

Теорема 8 (необходимое условие Куна-Таккера)

Пусть в задаче (5) множество P выпукло, функции $f, g_1, \ldots, g_m$ дифференцируемы в точке $x^* \in X$, функции $h_1, \ldots, h_l$ непрерывно дифференцируемы в некоторой окрестности x^*. Положим $I = \left\{ j \mid g_j(x^*) = 0 \right\}$. Далее пусть $\nabla g_j(x^*)$ при $j \in I$ и $\nabla h_i(x^*)$ при $i = 1, \ldots, l$ линейно независимы. Если x^* — локальное решение задачи (5), то существуют такие числа u_i, $i = 1, \ldots, m$ и v_i, $i = 1, \ldots, l$, что

$$\nabla f(x^*) + \sum_{i=1}^{m} u_i \nabla g_i(x^*) + \sum_{i=1}^{l} v_i \nabla h_i(x^*) = 0,$$

$$u_i g_i(x^*) = 0,\ i = 1, \ldots, m,$$

$$u_i \geqslant 0,\ i = 1, \ldots, m.$$

Условие $u_i g_i(x^*) = 0$ называется *условием дополняющей нежёсткости* и означает, что множители Лагранжа, соответствующие пассивным ограничениям-неравенствам, должны обращаться в нуль.

Условия Куна-Таккера

Теорема 9 (достаточное условие Куна-Таккера)

Пусть в задаче (5) множество P выпукло, функции $f, g_1, \ldots, g_m$ выпуклы на P и дифференцируемы в точке $x^* \in X$, функции $h_1, \ldots, h_l$ линейны. Предположим, что в точке x^* выполняются условия Куна-Таккера, то есть существуют такие числа u_i, $i = 1, \ldots, m$ и v_i, $i = 1, \ldots, l$, что

$$\nabla f(x^*) + \sum_{i=1}^{m} u_i \nabla g_i(x^*) + \sum_{i=1}^{l} v_i \nabla h_i(x^*) = 0,$$

$$u_i g_i(x^*) = 0,\ i = 1, \ldots, m,$$

$$u_i \geqslant 0,\ i = 1, \ldots, m.$$

Тогда точка x^* — (глобальное) решение задачи (5).

МИСиС

Любой задаче математического программирования можно поставить в соответствие так называемую *двойственную задачу* оптимизации. Между этими задачами обнаруживаются тесные связи, изучение которых составляет предмет теории двойственности.

Данная теория тесно связана с теорией условий оптимальности и в определённом смысле может рассматриваться как её раздел.

Определение

Двойственной к задаче (5) называется задача

$$\begin{aligned} &\varphi(u, v) \to \max \\ &u \geqslant 0 \end{aligned} \tag{7}$$

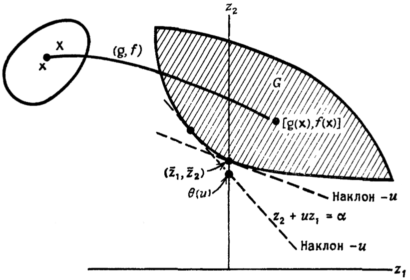
где

$$\varphi(u, v) = \inf_{x \in P} L(x, u, v) = \inf_{x \in P} \left(f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{i=1}^{l} v_i h_i(x) \right)$$

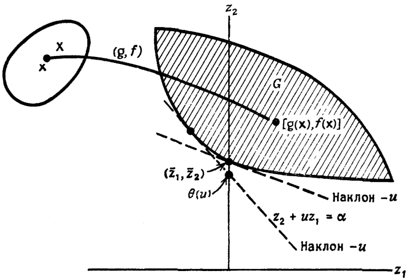
При этом задача (5) называется прямой.

Геометрическая интерпретация двойственной по Лагранжу задачи

Рассмотрим кратко геометрическую интерпретацию двойственной задачи. Для простоты возьмём задачу с единственным ограничением-неравенством. Прямая задача в этом случае имеет вид: $f(x) \to \min$ при условиях $g(x) \leqslant 0$, $x \in P$. На рисунке в плоскости (z_1, z_2) изображено множество $G = \big\{(z_1, z_2) \mid z_1 = g(x),\, z_2 = f(x),\, x \in P\big\}$.



Геометрическая интерпретация двойственной задачи



Для определения $\varphi(u)$ нужно минимизировать $z_2 + uz_1$ на множестве G, т. е. перемещать прямую $z_2 + uz_1 = \alpha$ параллельно самой себе, пока она не станет опорной к множеству G. Тогда точка пересечения этой прямой с осью z_2 укажет значение $\varphi(u)$.

Двойственная задача заключается в нахождении такого наклона опорной гиперплоскости, при котором значение координаты z_2 точки её пересечения с осью z_2 будет максимальным.

Теорема 10, называемая слабой теоремой двойственности, устанавливает, что значение целевой функции в любой допустимой точке двойственной задачи является оценкой снизу для значений целевой функции в любой допустимой точке прямой задачи.

Теорема 10 (Слабая теорема двойственности)

Пусть x — допустимое решение задачи (5), т. е. $x \in P$, $g(x) \leqslant 0$, $h(x) = 0$, а (u, v) — допустимое решение задачи (7), т. е. $u \geqslant 0$. Тогда $f(x) \geqslant \varphi(u, v)$.

Слабая теорема двойственности. Следствия

Следствие 1.

$$f^* \geqslant \varphi^*$$

$$f^* = \inf\big\{f(x) \;\big|\; x \in P,\, g(x) \leqslant 0,\, h(x) = 0\big\},\; \varphi^* = \sup\big\{\varphi(u, v) \;\big|\; u \geqslant 0\big\}$$

Следствие 2.

Если $f(\bar{x}) \leqslant \varphi(\bar{u}, \bar{v})$, где $\bar{u} \geqslant 0$ и $\bar{x} \in \big\{x \in P \;\big|\; g(x) \leqslant 0,\, h(x) = 0\big\}$ то $\bar{x}$ и $(\bar{u}, \bar{v})$ — оптимальные решения прямой и двойственной задач соответственно.

Следствие 3.

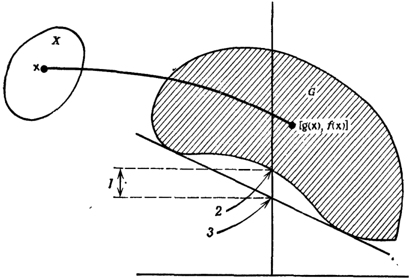
Если $\inf\big\{f(x) \;\big|\; x \in P,\, g(x) \leqslant 0,\, h(x) = 0\big\} = -\infty$, то $\varphi(u, v) = -\infty$ для всех $u \geqslant 0$.

Следствие 4.

Если $\sup\big\{\varphi(u, v) \;\big|\; u \geqslant 0\big\} = \infty$, то прямая задача не имеет допустимых решений.

Разрыв двойственности

По следствию 1 из теоремы 10 оптимальное значение целевой функции прямой задачи не меньше, чем оптимальное значение целевой функции двойственной задачи. Если выполняется строгое неравенство, то говорят, что имеет место *разрыв двойственности*.



Условия, гарантирующие отсутствие разрыва двойственности, сформулированы в теореме 11.

Сильная теорема двойственности

Теорема 11 (Сильная теорема двойственности)

Пусть в задаче (5) множество P выпукло, функции $f, g_1, \ldots, g_m$ выпуклы на P и дифференцируемы в точке $x^* \in X$, функции $h_1, \ldots, h_l$ линейны. Предположим, что выполняется следующее условие регулярности (Слейтера). Существует такой вектор $\hat{x} \in P$, что $g(\hat{x}) < 0$, $h(\hat{x}) = 0$, кроме того, $0 \in \operatorname{int} h(P)$. Здесь $h(P) = \big\{h(x) \bigm| x \in P\big\}$. Тогда

$$f^* = \varphi^*$$

Если нижняя грань конечна, то φ^* достигается в точке $(\bar{u}, \bar{v})$, для которой $\bar{u} \geqslant 0$. Если нижняя грань достигается в точке $\bar{x}$, то $\bar{u}^\mathsf{T} g(\bar{x}) = 0$.

Если оптимальное значение прямой задачи $(f^* > -\infty)$, в частности, если она имеет решение, то множество решений двойственной задачи (7) непусто, и совпадает с множеством векторов Куна-Таккера задачи (5). При этом выполняется соотношение двойственности.

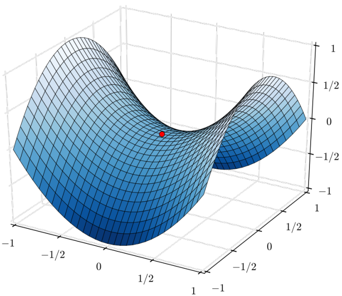
МИСиС

Критерий седловой точки

В условиях оптимальности Куна-Таккера предполагалось, что целевая функция и функции, входящие в ограничения, являются дифференцируемыми. Далее рассматриваются критерии оптимальности для недифференцируемых функций при наличии ограничений.

Определение

Говорят, что функция $f(x, y)$ имеет седловую точку (x^*, y^*), если $f(x^*, y) \leqslant f(x^*, y^*) \leqslant f(x, y^*)$ для всех x и y.



Критерий седловой точки

Теорема (о седловой точке)

Предположим, что существуют $\bar{x} \in P$ и $(\bar{u}, \bar{v})$, такие, что $\bar{u} \geqslant 0$

$$L(\bar{x}, u, v) \leqslant L(\bar{x}, \bar{u}, \bar{v}) \leqslant L(x, \bar{u}, \bar{v}) \tag{8}$$

для всех $x \in P$ и всех (u, v), для которых $u \geqslant 0$, где $L(x, u, v) = f(x) + u^{\mathsf{T}} g(x) + v^{\mathsf{T}} h(x)$. Тогда $\bar{x}$ и $(\bar{u}, \bar{v})$ являются соответственно решениями прямой задачи (5) и двойственной задачи (7). Обратно, предположим, что P — выпуклое множество, функции $f, g_1, \ldots, g_m$ выпуклы на P, функции $h_1, \ldots, h_l$ линейны. Кроме того, предположим, что $0 \in \operatorname{int} h(P)$ и существует $\hat{x} \in P$, такой, что $g(\hat{x}) < 0$ и $h(\hat{x}) = 0$. Если $\bar{x}$ — оптимальное решение задачи (5), то найдётся вектор $(\bar{u}, \bar{v})$, такой, что $\bar{u} \geqslant 0$ и справедливы неравенства (8).

Заметим, что необходимое условие оптимальности требует выпуклости и регулярности, в то время как достаточное условие не нуждается в этих предположениях.

Заключение

В данной лекции рассмотрены необходимые и достаточные условия оптимальности для задач условной оптимизации. Изложение началось с обсуждения условий оптимальности Лагранжа для задач с ограничениями в виде равенств. Затем были введены в рассмотрение ограничения в виде неравенств и построены условия оптимальности Куна-Таккера, т. е. условия первого порядка, использующие градиенты целевой функции и функций, входящих в ограничения.

Показано, что условия Куна-Таккера являются необходимыми, если соответствующие функции дифференцируемы, а ограничения удовлетворяют некоторому условию регулярности. Условия Куна-Таккера становятся достаточными условиями наличия глобального минимума, если целевая функция и ограничения-неравенства образованы выпуклыми функциями, а ограничения в виде равенств содержат линейные функции. Рассмотрены также условия оптимальности, связанные с существованием седловых точек функции Лагранжа. Эти условия применяются для исследования недифференцируемых функций.

Вопросы?