# Методы оптимизации
# Оптимизация без ограничений. Выпуклость

Н. В. Артамонов

МГИМО МИД России

4 октября 2023 г.

# Содержание

1. Безусловная оптимизация

2. Численная оптимизация
   - Градиентный спуск
   - Метод Ньютона

3. Выпуклость и глобальный экстремум

# Локальный экстремум

Пусть функция $f$ определена на множестве $\mathrm{dom}(f) \subset \mathbb{R}^n$.

**Теорема (FOC)**

*Если* $\hat{x} \in \mathrm{dom}(f)$ *– локальный экстремум, то*

$$f'_{x_i}(\hat{x}) = \frac{\partial f}{\partial x_i}(\hat{x}) = 0 \qquad i = 1, \ldots, n$$

Матричная запись: на оптимальном решении

$$\mathrm{grad}\, f(\hat{x}) = 0$$

# Локальный экстремум

## Теорема (SOC)

*Пусть* $\hat{x} \in \mathrm{dom}(f)$ *удовлетворяет FOC.*

*Тогда*

1. *если* $\mathrm{Hess}(\hat{x}) > 0$, *то* $\hat{x}$ – *локальный минимум;*
2. *если* $\mathrm{Hess}(\hat{x}) < 0$, *то* $\hat{x}$ – *локальный максимум;*
3. *если* $\mathrm{Hess}(\hat{x})$ *не знакоопределена то* $\hat{x}$ – *не локальный экстремум.*

В последнем случае $\hat{x}$ – седловая точка

Знакоопределённость проверяем по критерию Сильвестра.

## Пример: оптимизация производства

Фирма производит $n$ товаров и продаёт их по ценам $P_1, \dots, P_n$.

Пусть $Q = (Q_1, \dots, Q_n)$ – объёмы производства

Издержки производства определяются функцией издержек

$$C(Q) = C(Q_1, \dots, Q_n)$$

# Пример: оптимизация производства

Фирма производит $n$ товаров и продаёт их по ценам $P_1, \dots, P_n$.

Пусть $Q = (Q_1, \dots, Q_n)$ – объёмы производства

Издержки производства определяются функцией издержек

$$C(Q) = C(Q_1, \dots, Q_n)$$

Тогда функция прибыли имеет вид

$$\pi(Q) = R - C = PQ - C(Q) = \sum_{i=1}^{n} P_i Q_i - C(Q_1, \dots, Q_n)$$

# Пример: оптимизация производства

Постановка: Найти оптимальные объёмы производства

Модель:

$$\max_{Q} \pi(Q)$$

# Пример: оптимизация производства

Постановка: Найти оптимальные объёмы производства

Модель:

$$\max_Q \pi(Q)$$

Необходимые условия:

$$\begin{cases} \pi'_{Q_1} = 0 \\ \vdots \\ \pi'_{Q_n} = 0 \end{cases} \Rightarrow \begin{cases} P_1 = C'_{Q_1} \\ \vdots \\ P_n = C'_{Q_n} \end{cases}$$

Интерпретация: На оптимальном решении предельные издержки равны ценам.

1 Безусловная оптимизация

2 Численная оптимизация
- Градиентный спуск
- Метод Ньютона

3 Выпуклость и глобальный экстремум

# Описание метода

Основной алгоритм численного решения задачи безусловной оптимизации – метод градиентного спуска.

Идея: Строим последовательные приближения

$$\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots,$$

которые в «хорошем случае» сходятся к (локальному!) экстремуму функции $f$.

Переход от $\boldsymbol{x}_i$ к $\boldsymbol{x}_{i+1}$ осуществляется по направлению

- градиента для задачи на max
- противоположного к градиенту для задачи на min

# Реализация для min

1. Выбираем «начальное приближение» $\boldsymbol{x}_0$.
2. Строим последовательность

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \lambda_i \nabla f(\boldsymbol{x}_i) \qquad\qquad i = 0, 1, \ldots$$

здесь $\lambda_i$ – параметр метода (обычно убывает).

Условие остановки:

$$\|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\| < \varepsilon$$

или

$$|f(x_{i+1}) - f(x_i)| < \varepsilon$$

где $\varepsilon$ – заданная точность

# Проблема сходимости

Пусть $\hat{x}$ – локальный минимум и $\boldsymbol{x}_0$ «достаточно близко» к $\hat{x}$.

Тогда последовательность $\{\boldsymbol{x}_i\}_{i=0}^{n}$ сходится к $\hat{x}$

Проблема практической реализации:

- выбор $\boldsymbol{x}_0$
- выбор $\lambda_i$ («скорости спуска»)

**Важно!**

*При неправильном выборе $\boldsymbol{x}_0$ и $\lambda_i$ метод может расходиться!*

# Реализация для max

1. Выбираем «начальное приближение» $\boldsymbol{x}_0$.
2. Строим последовательность

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \lambda_i \nabla f(\boldsymbol{x}_i) \qquad i = 0, 1, \dots$$

   здесь $\lambda_i$ – параметр метода (обычно убывает).

Условие остановки:

$$\|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\| < \varepsilon$$

или

$$|f(x_{i+1}) - f(x_i)| < \varepsilon$$

где $\varepsilon$ – заданная точность

1 Безусловная оптимизация

2 Численная оптимизация
- Градиентный спуск
- Метод Ньютона

3 Выпуклость и глобальный экстремум

# Реализация

Строим последовательные приближения

$$\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$$

по правилу

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \gamma \mathrm{Hess}_f^{-1}(\boldsymbol{x}_i) \nabla f(x_i)$$

где $\gamma \in (0, 1]$ – параметр

Условие остановки:

$$\|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\| < \varepsilon$$

или

$$|f(x_{i+1}) - f(x_i)| < \varepsilon$$

где $\varepsilon$ – заданная точность.

# Проблема сходимости

Пусть $\hat{x}$ – локальный экстремум и $\boldsymbol{x}_0$ «достаточно близко» к $\hat{\boldsymbol{x}}$.

Тогда последовательность $\{\boldsymbol{x}_i\}_{i=0}^{n}$ сходится к $\hat{\boldsymbol{x}}$

Проблема практической реализации:

- выбор $\boldsymbol{x}_0$
- выбор $\gamma$

**Важно!**

*При неправильном выборе $\boldsymbol{x}_0$ и $\gamma$ метод может расходиться!*

# Выпуклое множество

Одно из важных понятий в оптимизации – выпуклость. Оно используется для нахождения глобальных экстремумов.

# Выпуклое множество

Одно из важных понятий в оптимизации – выпуклость. Оно используется для нахождения глобальных экстремумов.

## Определение

*Множество $X \subset \mathbb{R}^n$ называется выпуклым, если*[a]

$$\forall M.N \in X \Rightarrow [M, N] \subset X$$

*т.е. для любых двух точек из множества отрезок, их соединяющий, также лежит в этом множестве.*

[a] Здесь $[M, N] = \{\lambda M + (1 - \lambda)N | \lambda \in [0, 1]\}$

Примеры: круг, шар, треугольник, прямоугольник, ромб, $\mathbb{R}^n_+$

# Выпуклые и вогнутые функции

Пусть числовая функция $f$ определена на $\mathrm{dom}(f) \subset \mathbb{R}^n$

График функции: поверхность в $\mathbb{R}^{n+1}$

$$\Gamma = \{(\boldsymbol{x}, f(\boldsymbol{x})) : \boldsymbol{x} \in \mathrm{dom}(f)\}$$

Далее будем предполагать, что $\mathrm{dom}(f)$ – выпуклое множество

# Выпуклые функция

### Определение

*Функция $f$ называется выпуклой, если для любых $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}(f)$ отрезок, соединяющий точки на графике $(\boldsymbol{x}, f(\boldsymbol{x}))$ и $(\boldsymbol{y}, f(\boldsymbol{y}))$, лежит выше графика функции (или на нём).*

Формально: для всех $\lambda \in [0, 1]$

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$

# Вогнутые функция

## Определение

*Функция $f$ называется вогнутой, если для любых $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom}(f)$ отрезок, соединяющий точки на графике $(\boldsymbol{x}, f(\boldsymbol{x}))$ и $(\boldsymbol{y}, f(\boldsymbol{y}))$, лежит ниже графика функции (или на нём).*

Формально: для всех $\lambda \in [0, 1]$

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \geq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$

# Строгая выпуклость и вогнутость

**Определение**

*Выпуклая функция называется* *строго выпуклой**, если для всех* $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(f)$ *и* $\lambda \in (0, 1)$ *выполнено*

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

**Определение**

*Вогнутая функция называется* *строго вогнутой**, если для всех* $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(f)$ *и* $\lambda \in (0, 1)$ *выполнено*

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) > \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

Геометрическая интерпретация: график не содержит отрезков («не участков линейности»)

# Базовые свойства

- $f$ (строго) выпукла тогда и только тогда, когда $-f$ (строго) вогнута.
- $f$ (строго) выпукла и $\alpha > 0 \Rightarrow \alpha f$ (строго) выпукла.
- $f$ (строго) вогнута и $\alpha > 0 \Rightarrow \alpha f$ (строго) вогнута.
- $f$ и $g$ выпуклы (вогнуты) $\Rightarrow f + g$ выпукла (вогнута).

# Основной результат

**Теорема**

*Дважды непрерывно дифференцируемая функция $f$ выпукла* $\iff \mathrm{Hess}_f(\boldsymbol{x}) \geq 0$ *для всех* $\boldsymbol{x} \in \mathrm{dom}(f)$.

# Основной результат

### Теорема

*Дважды непрерывно дифференцируемая функция* $f$ *выпукла* $\iff \mathrm{Hess}_f(\boldsymbol{x}) \geq 0$ *для всех* $\boldsymbol{x} \in \mathrm{dom}(f)$.

*Если* $\mathrm{Hess}_f(\boldsymbol{x}) > 0$ *для всех* $\boldsymbol{x} \in \mathrm{dom}(f)$, *то функция строго выпукла на* $\mathrm{dom}(f)$.

# Основной результат

## Теорема

*Дважды непрерывно дифференцируемая функция $f$ выпукла* $\iff \mathrm{Hess}_f(\boldsymbol{x}) \geq 0$ *для всех* $\boldsymbol{x} \in \mathrm{dom}(f)$.

*Если* $\mathrm{Hess}_f(\boldsymbol{x}) > 0$ *для всех* $\boldsymbol{x} \in \mathrm{dom}(f)$, *то функция строго выпукла на* $\mathrm{dom}(f)$.

## Следствие

*Дважды непрерывно дифференцируемая функция $f$ вогнута* $\iff \mathrm{Hess}_f(\boldsymbol{x}) \leq 0$ *для всех* $\boldsymbol{x} \in \mathrm{dom}(f)$.

# Основной результат

## Теорема

*Дважды непрерывно дифференцируемая функция $f$ выпукла* $\iff \mathrm{Hess}_f(\mathbf{x}) \geq 0$ *для всех* $\mathbf{x} \in \mathrm{dom}(f)$.

*Если* $\mathrm{Hess}_f(\mathbf{x}) > 0$ *для всех* $\mathbf{x} \in \mathrm{dom}(f)$, *то функция строго выпукла на* $\mathrm{dom}(f)$.

## Следствие

*Дважды непрерывно дифференцируемая функция $f$ вогнута* $\iff \mathrm{Hess}_f(\mathbf{x}) \leq 0$ *для всех* $\mathbf{x} \in \mathrm{dom}(f)$.

*Если* $\mathrm{Hess}_f(\mathbf{x}) < 0$ *для всех* $\mathbf{x} \in \mathrm{dom}(f)$, *то функция строго вогнута на* $\mathrm{dom}(f)$.

# Выпуклая оптимизация

**Теорема**

*Пусть $f$ – выпуклая функция. Если $\hat{x}$ – локальный минимум, то он будет глобальным минимумом.*

*Если $f$ строго выпукла, то глобальный минимум единственный.*

**Следствие**

*Пусть $f$ – вогнутая функция. Если $\hat{x}$ – локальный максимум, то он будет глобальным максимумом.*

*Если $f$ строго вогнута, то глобальный максимум единственный.*

# Случай 2D

Случай двух переменных: $f = f(x, y)$

$$\mathrm{Hess}(x,y) = \begin{pmatrix} f''_{xx} & f''_{xy} \\ f''_{xy} & f''_{yy} \end{pmatrix} \qquad \Delta_1 = f''_{xx} \qquad \Delta_2 = f''_{xx}f''_{yy} - \left(f''_{xy}\right)^2$$

Тогда

$$\mathrm{Hess} \geq 0 \iff f''_{xx}, f''_{yy} \geq 0, f''_{xx}f''_{yy} - \left(f''_{xy}\right)^2 \geq 0$$
$$\mathrm{Hess} \leq 0 \iff f''_{xx}, f''_{yy} \leq 0, f''_{xx}f''_{yy} - \left(f''_{xy}\right)^2 \geq 0$$

## Предложение

Функция двух переменных

- выпукла $\iff f''_{xx}, f''_{yy} \geq 0, f''_{xx}f''_{yy} - \left(f''_{xy}\right)^2 \geq 0$
- вогнута $\iff f''_{xx}, f''_{yy} \leq 0, f''_{xx}f''_{yy} - \left(f''_{xy}\right)^2 \geq 0$

# Выпуклость функции Кобба-Дугласа

Рассмотрим производственную функцию Кобба–Дугласа

$$f(K, L) = cK^{\alpha}L^{\beta} \qquad \operatorname{dom}(f) = \mathbb{R}^2_+ \qquad \alpha, \beta > 0$$

# Выпуклость функции Кобба-Дугласа

Рассмотрим производственную функцию Кобба–Дугласа

$$f(K, L) = cK^{\alpha}L^{\beta} \qquad \mathrm{dom}(f) = \mathbb{R}^2_+ \qquad \alpha, \beta > 0$$

Тогда

$$f'_K = \alpha c K^{\alpha-1}L^{\beta} = \frac{\alpha f}{K} \qquad f'_L = \beta c K^{\alpha}L^{\beta-1} = \frac{\beta f}{L}$$

# Выпуклость функции Кобба-Дугласа

Рассмотрим производственную функцию Кобба–Дугласа

$$f(K, L) = cK^{\alpha}L^{\beta} \qquad \mathrm{dom}(f) = \mathbb{R}^2_+ \qquad \alpha, \beta > 0$$

Тогда

$$f'_K = \alpha c K^{\alpha-1}L^{\beta} = \frac{\alpha f}{K} \qquad f'_L = \beta c K^{\alpha}L^{\beta-1} = \frac{\beta f}{L}$$

и

$$f''_{KK} = \alpha(\alpha-1)cK^{\alpha-2}L^{\beta} = \frac{\alpha(\alpha-1)f}{K^2}$$
$$f''_{LL} = \beta(\beta-1)cK^{\alpha}L^{\beta-2} = \frac{\beta(\beta-1)f}{L^2}$$
$$f''_{KL} = f''_{LK} = \alpha\beta cK^{\alpha-1}L^{\beta-1} = \frac{\alpha\beta f}{KL}$$

# Выпуклость функции Кобба-Дугласа

Гессиан функции Кобба-Дугласа

$$\mathrm{Hess}_f = \begin{pmatrix} f''_{KK} & f''_{KL} \\ f''_{LK} & f''_{LL} \end{pmatrix} = \begin{pmatrix} \frac{\alpha(\alpha-1)f}{K^2} & \frac{\alpha\beta f}{KL} \\ \frac{\alpha\beta f}{KL} & \frac{\beta(\beta-1)f}{L^2} \end{pmatrix} = \frac{f}{K^2L^2}\begin{pmatrix} \alpha(\alpha-1)L^2 & \alpha\beta KL \\ \alpha\beta KL & \beta(\beta-1)K^2 \end{pmatrix}$$

# Выпуклость функции Кобба-Дугласа

Угловые миноры функции Кобба-Дугласа

$$\Delta_1(K, L) = \frac{\alpha(\alpha - 1)f}{K^2}$$

$$\Delta_2(K, L) = \det \mathrm{Hess}_f = \frac{\alpha\beta(1 - \alpha - \beta)f^2}{K^2 L^2}$$

# Выпуклость функции Кобба-Дугласа

Угловые миноры функции Кобба-Дугласа

$$\Delta_1(K, L) = \frac{\alpha(\alpha - 1)f}{K^2}$$

$$\Delta_2(K, L) = \det \mathrm{Hess}_f = \frac{\alpha\beta(1 - \alpha - \beta)f^2}{K^2 L^2}$$

Случай $\alpha + \beta < 1$: $\Delta_1 < 0$ и $\Delta_2 > 0$ для всех $K, L > 0$.
Следовательно, функция $f$ строго вогнута.

# Выпуклость функции Кобба-Дугласа

Угловые миноры функции Кобба-Дугласа

$$\Delta_1(K, L) = \frac{\alpha(\alpha - 1)f}{K^2}$$

$$\Delta_2(K, L) = \det \mathrm{Hess}_f = \frac{\alpha\beta(1 - \alpha - \beta)f^2}{K^2 L^2}$$

Случай $\alpha + \beta < 1$: $\Delta_1 < 0$ и $\Delta_2 > 0$ для всех $K, L > 0$.
Следовательно, функция $f$ строго вогнута.

Случай $\alpha + \beta = 1$: $f''_{KK}, f''_{LL} < 0$ и $\Delta_2 = 0$ для всех $K, L > 0$.
Следовательно, функция $f$ вогнута.

# Выпуклость функции Кобба-Дугласа

Угловые миноры функции Кобба-Дугласа

$$\Delta_1(K, L) = \frac{\alpha(\alpha - 1)f}{K^2}$$

$$\Delta_2(K, L) = \det \mathrm{Hess}_f = \frac{\alpha\beta(1 - \alpha - \beta)f^2}{K^2 L^2}$$

Случай $\alpha + \beta < 1$: $\Delta_1 < 0$ и $\Delta_2 > 0$ для всех $K, L > 0$.
Следовательно, функция $f$ строго вогнута.

Случай $\alpha + \beta = 1$: $f''_{KK}, f''_{LL} < 0$ и $\Delta_2 = 0$ для всех $K, L > 0$.
Следовательно, функция $f$ вогнута.

Случай $\alpha + \beta > 1$: $\Delta_2 < 0$ и гессиан не знакоопределен.
Следовательно, функция $f$ не выпукла и не вогнута.