Нелинейное программирование

Градиентный спуск и метод Ньютона

Теория, алгоритмы и интерактивный инструмент

Русский

Вольфганг Бёме

wolfgang.boehme@economic-modeling.de

www.economic-modeling.de

25 октября 2026 г.

Полное введение с теорией, алгоритмами и разобранным примером

Документация на этом сайте разработана при поддержке ИИ Claude (Anthropic).

Содержание

1 Введение и мотивация **2**

2 Математические основы **2**
2.1 Общая задача нелинейного программирования 2
2.2 Условия оптимальности 2
2.3 Градиент и матрица Гессе 3
2.4 Численное дифференцирование (в инструменте) 3

3 Градиентный спуск **3**
3.1 Основная идея 3
3.2 Условие Армихо (Backtracking) 4
3.3 Скорость сходимости 4

4 Метод Ньютона **4**
4.1 Основная идея 4
4.2 Направление Ньютона 4
4.3 Квадратичная сходимость 5
4.4 Ограничения метода Ньютона 5

5 Полный пример: оба метода **5**
5.1 Постановка задачи 5
5.2 Метод 1: Градиентный спуск ($\alpha = 0{,}1$) 5
5.3 Метод 2: Метод Ньютона 6
5.4 Сравнение методов 6

6 Описание инструмента **6**
6.1 Ввод функции 6
6.2 Предустановленные примеры 7
6.3 Результаты 7

7 Итоги **7**

1. Введение и мотивация

Нелинейное программирование (НЛП) занимается минимизацией (или максимизацией) нелинейной целевой функции, возможно при нелинейных ограничениях.

В отличие от линейного программирования, здесь нет симплекс-подобных методов с гарантией глобального оптимума. Вместо этого используются итерационные методы, сходящиеся локально.

Типичные применения

- Машинное обучение: минимизация функции потерь
- Инженерия: оптимизация формы, механика конструкций
- Финансы: оптимизация портфеля с нелинейными рисками
- Химия: химические равновесия

2. Математические основы

2.1. Общая задача нелинейного программирования

Определение: Общая задача НЛП

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{при условии} \quad g_i(x) \leq 0,\ h_j(x) = 0$$

Для **безусловной** задачи: $\min f(x)$ без ограничений.

2.2. Условия оптимальности

Необходимое условие первого порядка (ККТ)

Локальный минимум x^* безусловной задачи удовлетворяет:

$$\nabla f(x^*) = 0 \quad \text{(стационарная точка)}$$

Достаточное условие второго порядка

Стационарная точка x^* является **локальным минимумом**, если матрица Гессе положительно определена:

$$\nabla^2 f(x^*) \succ 0$$

2.3. Градиент и матрица Гессе

Градиент и матрица Гессе

Градиент $\nabla f(x)$ указывает направление наибольшего роста:

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)^\top$$

Матрица Гессе $\nabla^2 f(x)$ описывает кривизну:

$$\left[\nabla^2 f(x) \right]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

2.4. Численное дифференцирование (в инструменте)

Инструмент вычисляет градиент и матрицу Гессе численно:

$$\frac{\partial f}{\partial x_i} \approx \frac{f(x + he_i) - f(x - he_i)}{2h}, \quad \frac{\partial^2 f}{\partial x_i^2} \approx \frac{f(x + he_i) - 2f(x) + f(x - he_i)}{h^2}$$

3. Градиентный спуск

3.1. Основная идея

Алгоритм: Градиентный спуск

1. Выбрать начальную точку x^0 и шаг $\alpha > 0$
2. Вычислить градиент $g^k = \nabla f(x^k)$
3. Если $\|g^k\| < \varepsilon$: **Стоп** (условие сходимости)
4. Установить $x^{k+1} = x^k - \alpha \cdot g^k$
5. Перейти к шагу 2

Интуиция: Всегда двигаемся в направлении наибольшего убывания.

3.2. Условие Армихо (Backtracking)

Условие Армихо

Шаг α выбирается так, чтобы значение функции убывало достаточно:

$$f(x^k - \alpha \nabla f(x^k)) \leq f(x^k) - c \cdot \alpha \cdot \|\nabla f(x^k)\|^2$$

Если условие не выполнено: $\alpha \leftarrow \beta \cdot \alpha$ (уменьшить шаг).

3.3. Скорость сходимости

Сходимость градиентного спуска

Для сильно выпуклых функций с липшицевым градиентом градиентный спуск сходится **линейно**:

$$\|x^k - x^*\| \leq \left(1 - \frac{\mu}{L}\right)^k \|x^0 - x^*\|$$

4. Метод Ньютона

4.1. Основная идея

Алгоритм: Метод Ньютона

1. Выбрать начальную точку x^0
2. Вычислить градиент g^k и матрицу Гессе H^k
3. Если $\|g^k\| < \varepsilon$: **Стоп**
4. Решить линейную систему: $H^k d^k = -g^k$
5. Установить $x^{k+1} = x^k + d^k$; перейти к шагу 2

4.2. Направление Ньютона

Направление Ньютона

Направление Ньютона минимизирует квадратичную аппроксимацию f в текущей точке:

$$d^k = -(H^k)^{-1} g^k$$

Геометрическая интерпретация: Метод Ньютона использует информацию о кривизне (матрица Гессе) для прямого перехода к минимуму локальной квадратичной аппроксимации.

4.3. Квадратичная сходимость

Квадратичная сходимость метода Ньютона

Если x^* — невырожденный локальный минимум и x^0 достаточно близко к x^*, то метод Ньютона сходится **квадратично**:

$$\|x^{k+1} - x^*\| \leq C \cdot \|x^k - x^*\|^2$$

4.4. Ограничения метода Ньютона

- Матрица Гессе должна быть **положительно определённой** (иначе нет убывания).
- Вычисление и обращение матрицы Гессе **дорого** ($O(n^3)$).
- Без хорошей начальной точки метод может **расходиться**.
- Решение: затухающий метод Ньютона (поиск вдоль направления Ньютона).

5. Полный пример: оба метода

5.1. Постановка задачи

Пример функции: квадратичный минимум

Найти глобальный минимум функции:

$$f(x_1, x_2) = (x_1 - 2)^2 + (x_2 - 3)^2$$

Аналитическое решение: $x^* = (2,\ 3)$, $f(x^*) = 0$.
Градиент:

$$\nabla f(x) = \begin{pmatrix} 2(x_1 - 2) \\ 2(x_2 - 3) \end{pmatrix}$$

5.2. Метод 1: Градиентный спуск ($\alpha = 0{,}1$)

Начальная точка: $x^0 = (5,\ 7)$
Итерация 0 → 1:

$$g^0 = \begin{pmatrix} 2(5-2) \\ 2(7-3) \end{pmatrix} = \begin{pmatrix} 6 \\ 8 \end{pmatrix}, \quad x^1 = x^0 - 0{,}1 \cdot g^0 = \begin{pmatrix} 5 \\ 7 \end{pmatrix} - \begin{pmatrix} 0{,}6 \\ 0{,}8 \end{pmatrix} = \begin{pmatrix} 4{,}4 \\ 6{,}2 \end{pmatrix}$$

Итерация 1 → 2:

$$g^1 = \begin{pmatrix} 4{,}8 \\ 6{,}4 \end{pmatrix}, \quad x^2 = \begin{pmatrix} 4{,}4 \\ 6{,}2 \end{pmatrix} - 0{,}1 \cdot \begin{pmatrix} 4{,}8 \\ 6{,}4 \end{pmatrix} = \begin{pmatrix} 3{,}92 \\ 5{,}56 \end{pmatrix}$$

Характер сходимости: Линейная — каждая итерация уменьшает ошибку на постоянный множитель $1 - 2\alpha\mu \approx 0{,}8$.

5.3. Метод 2: Метод Ньютона

Начальная точка: $x^0 = (5,\ 7)$
Матрица Гессе:

$$H = \nabla^2 f = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Уравнение Ньютона: $H\,d = -g^0$:

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} d = -\begin{pmatrix} 6 \\ 8 \end{pmatrix} \quad \Rightarrow \quad d = \begin{pmatrix} -3 \\ -4 \end{pmatrix}$$

Шаг Ньютона:

$$x^1 = x^0 + d = \begin{pmatrix} 5 \\ 7 \end{pmatrix} + \begin{pmatrix} -3 \\ -4 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} = x^*$$

Метод Ньютона достигает минимума квадратичных функций **за один шаг**!

5.4. Сравнение методов

Свойство	Градиентный спуск	Метод Ньютона
Реализация	Простая	Сложная (Гессе)
Скорость сходимости	Линейная	Квадратичная
Масштабируемость	Хорошая	Плохая
Матрица Гессе	Не нужна	Необходима

6. Описание инструмента

Нелинейное программирование — Интерактивный инструмент

Инструмент визуализирует оба метода оптимизации на произвольных функциях.

6.1. Ввод функции

Целевая функция $f(x_1, x_2)$ вводится как выражение JavaScript. Допустимые операторы: `+`, `-`, `*`, `/`, `**`, `Math.sin`, `Math.cos`, `Math.exp`, `Math.abs`.

6.2. Предустановленные примеры

Название	Формула	Минимум
Розенброк	$(1-x_1)^2+100(x_2-x_1^2)^2$	$(1,1)$
Квадратичная	$x_1^2+x_2^2$	$(0,0)$
Химмельблау	$(x_1^2+x_2-11)^2+(x_1+x_2^2-7)^2$	Несколько
Бута	$(x_1+2x_2-7)^2+(2x_1+x_2-5)^2$	$(1,3)$
Матиаса	$0{,}26(x_1^2+x_2^2)-0{,}48x_1x_2$	$(0,0)$
Растригина	$20+x_1^2+x_2^2-10(\cos 2\pi x_1+\cos 2\pi x_2)$	$(0,0)$

6.3. Результаты

- **Контурный график:** линии уровня функции и путь оптимизации.
- **Результирующий блок:** конечная точка, значение функции, итерации, статус.
- **Журнал итераций:** табличный вывод всех шагов.

7. Итоги

Ключевые положения: Нелинейное программирование

1. НЛП минимизирует нелинейные функции; гарантии глобального оптимума нет.
2. Необходимое условие: $\nabla f(x^*)=0$ (стационарная точка).
3. Достаточное условие: матрица Гессе положительно определена в x^*.
4. **Градиентный спуск:** прост, устойчив, сходится линейно.
5. **Метод Ньютона:** использует кривизну, сходится квадратично, дорог.
6. Условие Армихо гарантирует достаточное убывание функции.
7. Инструмент: интерактивная визуализация обоих методов, контурный график.