Сборник задач

Нелинейное программирование – градиентный метод и метод Ньютона

Русский

Вольфганг Бёме

wolfgang.boehme@economic-modeling.de

www.economic-modeling.de

25 октября 2026 г.

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

Содержание

1 Уровень 1: Лёгкий **2**
Задача 1.1 2
Задача 1.2 3

2 Уровень 2: Средний **4**
Задача 2.1 4
Задача 2.2 5

3 Уровень 3: Сложный **7**
Задача 3.1 7
Задача 3.2 8

1 Уровень 1: Лёгкий

Задача 1.1 – Стационарные точки и условия оптимальности

Задача 1.1

Дана функция $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(x_1, x_2) = x_1^2 + 4x_2^2 - 4x_1 - 8x_2 + 10$$

(a) Рассчитай градиент $\nabla f(x_1, x_2)$ и матрицу Гессе $\mathbf{H}(x_1, x_2)$.

(b) Определи все стационарные точки (положи $\nabla f = \mathbf{0}$).

(c) С помощью матрицы Гессе определи, является ли стационарная точка минимумом, максимумом или седловой точкой.

(d) Рассчитай значение функции в оптимуме.

Решение: Задача 1.1

(a) Градиент и матрица Гессе:

$$\nabla f = \begin{pmatrix} 2x_1 - 4 \\ 8x_2 - 8 \end{pmatrix}, \qquad \mathbf{H} = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}$$

(b) Стационарная точка: $\nabla f = \mathbf{0}$:

$$2x_1 - 4 = 0 \Rightarrow x_1^* = 2, \qquad 8x_2 - 8 = 0 \Rightarrow x_2^* = 1$$

Единственная стационарная точка: $\mathbf{x}^* = (2, 1)^\top$.
(c) Тип стационарной точки:
Собственные значения $\mathbf{H}$: $\lambda_1 = 2 > 0$, $\lambda_2 = 8 > 0 \Rightarrow \mathbf{H} \succ 0$ (положительно определена).
$\Rightarrow \mathbf{x}^* = (2, 1)^\top$ является **строгим глобальным минимумом**.
(d) Значение функции:

$$f(2, 1) = 4 + 4 - 8 - 8 + 10 = 2$$

Оптимум: $\mathbf{x}^* = (2, 1)^\top$, $f^* = 2$

Задача 1.2 – Один шаг градиентного спуска

Задача 1.2

Дана $f(x_1, x_2) = (x_1 - 3)^2 + 2(x_2 + 1)^2$ с начальной точкой $\mathbf{x}^{(0)} = (1, 1)^\top$.

(a) Рассчитай $\nabla f(\mathbf{x}^{(0)})$ и направление поиска $\mathbf{d}^{(0)} = -\nabla f(\mathbf{x}^{(0)})$.

(b) Выполни один шаг градиентного спуска с фиксированным шагом $\alpha = 0{,}25$.

(c) Сколько итераций требуется градиентному спуску с оптимальным шагом для достижения минимума? (Обоснуй.)

(d) Рассчитай число обусловленности $\kappa = \lambda_{\max}/\lambda_{\min}$ матрицы Гессе.

Решение: Задача 1.2

(a) Градиент и направление поиска:

$$\nabla f = \begin{pmatrix} 2(x_1 - 3) \\ 4(x_2 + 1) \end{pmatrix}, \quad \nabla f(\mathbf{x}^{(0)}) = \begin{pmatrix} 2(1 - 3) \\ 4(1 + 1) \end{pmatrix} = \begin{pmatrix} -4 \\ 8 \end{pmatrix}$$

$$\mathbf{d}^{(0)} = -\nabla f(\mathbf{x}^{(0)}) = \begin{pmatrix} 4 \\ -8 \end{pmatrix}$$

(b) Шаг градиентного спуска с $\alpha = 0{,}25$:

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + 0{,}25 \cdot \mathbf{d}^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0{,}25 \begin{pmatrix} 4 \\ -8 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

$$f(\mathbf{x}^{(1)}) = (2 - 3)^2 + 2(-1 + 1)^2 = 1 + 0 = 1$$

(c) Оптимальный шаг:
Матрица Гессе $\mathbf{H} = \operatorname{diag}(2, 4)$. Поскольку $\mathbf{H}$ диагональна и постоянна, можно оптимизировать каждую компоненту отдельно. С оптимальным шагом $\alpha_i^* = 1/\lambda_i$:

$$x_1^{(1)} = 1 + \frac{1}{2} \cdot 4 = 3, \quad x_2^{(1)} = 1 + \frac{1}{4} \cdot (-8) = -1$$

$\mathbf{x}^{(1)} = (3, -1)^\top = \mathbf{x}^*$. Градиентный спуск с покомпонентно оптимальным шагом достигает минимума за **одну итерацию**.
(d) Число обусловленности:

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{4}{2} = 2$$

Число обусловленности 2 очень хорошее – градиентный спуск сходится быстро.
Оптимум: $\mathbf{x}^* = (3, -1)^\top$, $f^* = 0$, $\kappa = 2$

2 Уровень 2: Средний

Задача 2.1 – Метод Ньютона: две итерации

Задача 2.1

Дана нелинейная функция:

$$f(x_1, x_2) = x_1^4 + x_2^4 - 4x_1x_2 + 1$$

с начальной точкой $\mathbf{x}^{(0)} = (2, 2)^\top$.

(a) Рассчитай ∇f и $\mathbf{H}f$ (в общем виде).

(b) Выполни **две** полные итерации метода Ньютона (с шагом $\alpha = 1$).

(c) Рассчитай $\|\nabla f(\mathbf{x}^{(k)})\|$ после каждой итерации и оцени сходимость.

(d) Обоснуй, почему метод Ньютона здесь быстро сходится.

Решение: Задача 2.1

(a) Градиент и матрица Гессе:

$$\nabla f = \begin{pmatrix} 4x_1^3 - 4x_2 \\ 4x_2^3 - 4x_1 \end{pmatrix}, \qquad \mathbf{H} = \begin{pmatrix} 12x_1^2 & -4 \\ -4 & 12x_2^2 \end{pmatrix}$$

(b) Итерация Ньютона 1 (в $\mathbf{x}^{(0)} = (2, 2)^\top$):

$$\nabla f(\mathbf{x}^{(0)}) = \begin{pmatrix} 4 \cdot 8 - 8 \\ 4 \cdot 8 - 8 \end{pmatrix} = \begin{pmatrix} 24 \\ 24 \end{pmatrix}, \qquad \mathbf{H}(\mathbf{x}^{(0)}) = \begin{pmatrix} 48 & -4 \\ -4 & 48 \end{pmatrix}$$

$\det(\mathbf{H}^{(0)}) = 48^2 - 16 = 2304 - 16 = 2288$.
Система уравнений Ньютона $\mathbf{H}^{(0)}\mathbf{d}^{(0)} = -\nabla f(\mathbf{x}^{(0)})$:

$$\begin{pmatrix} 48 & -4 \\ -4 & 48 \end{pmatrix} \mathbf{d}^{(0)} = \begin{pmatrix} -24 \\ -24 \end{pmatrix}$$

Из симметрии (∇f имеет одинаковые компоненты): $\mathbf{d}^{(0)} = (d, d)^\top$ с $(48 - 4)d = -24$, то есть $d = -24/44 = -6/11$.

$$\mathbf{d}^{(0)} = \begin{pmatrix} -6/11 \\ -6/11 \end{pmatrix} \approx \begin{pmatrix} -0{,}5455 \\ -0{,}5455 \end{pmatrix}$$

$$\mathbf{x}^{(1)} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \begin{pmatrix} -6/11 \\ -6/11 \end{pmatrix} = \begin{pmatrix} 16/11 \\ 16/11 \end{pmatrix} \approx \begin{pmatrix} 1{,}4545 \\ 1{,}4545 \end{pmatrix}$$

Итерация Ньютона 2 (в $\mathbf{x}^{(1)} \approx (1{,}4545, 1{,}4545)^\top$):

$$\nabla f(\mathbf{x}^{(1)}) \approx \begin{pmatrix} 4(1{,}4545)^3 - 4(1{,}4545) \\ 4(1{,}4545)^3 - 4(1{,}4545) \end{pmatrix} = \begin{pmatrix} 4(3{,}0787) - 5{,}818 \\ 4(3{,}0787) - 5{,}818 \end{pmatrix} \approx \begin{pmatrix} 6{,}497 \\ 6{,}497 \end{pmatrix}$$

$$\mathbf{H}(\mathbf{x}^{(1)}) \approx \begin{pmatrix} 12(2{,}1156) & -4 \\ -4 & 12(2{,}1156) \end{pmatrix} = \begin{pmatrix} 25{,}387 & -4 \\ -4 & 25{,}387 \end{pmatrix}$$

Шаг Ньютона: $d = -6{,}497/(25{,}387 - 4) = -6{,}497/21{,}387 \approx -0{,}3038$.

$$\mathbf{x}^{(2)} \approx \begin{pmatrix} 1{,}4545 - 0{,}3038 \\ 1{,}4545 - 0{,}3038 \end{pmatrix} = \begin{pmatrix} 1{,}1507 \\ 1{,}1507 \end{pmatrix}$$

(c) Норма градиента:

k	$\mathbf{x}^{(k)}$	$f(\mathbf{x}^{(k)})$	$\|\nabla f(\mathbf{x}^{(k)})\|$
0	$(2, 2)$	$2 \cdot 16 - 4 \cdot 4 + 1 = 17$	$\sqrt{2} \cdot 24 \approx 33{,}94$
1	$(1{,}455, 1{,}455)$	$\approx 3{,}82$	$\approx 9{,}19$
2	$(1{,}151, 1{,}151)$	$\approx 1{,}48$	$\approx 3{,}14$

Стационарная точка в $\mathbf{x}^* = (1, 1)^\top$: $4 - 4 = 0$ ✓, $f^* = 2 - 4 + 1 = -1$.
(d) Быстрая сходимость: Функция гладкая ($\mathcal{C}^\infty$), а начальная точка расположена относительно близко к минимуму. В окрестности минимума $\mathbf{H} \succ 0$, поэтому метод Ньютона сходится квадратично.
Оптимум: $\mathbf{x}^* = (1, 1)^\top$, $f^* = -1$

Задача 2.2 – Сравнение градиентного метода и метода Ньютона: число обусловленности

Задача 2.2

Рассмотри функцию $f(x_1, x_2) = x_1^2 + 100x_2^2$ (сильно вытянутый эллипс).

(a) Рассчитай градиент, матрицу Гессе и число обусловленности κ.

(b) Выполни **три** шага градиентного спуска с оптимальным шагом $\alpha^* = 2/(\lambda_{\max} + \lambda_{\min})$, начальная точка $\mathbf{x}^{(0)} = (1, 1)^\top$.

(c) Выполни **один** шаг Ньютона и сравни.

(d) Объясни, почему градиентный спуск медленный при большом числе обусловленности (эффект «зигзага»).

Решение: Задача 2.2

(a) Градиент, матрица Гессе, число обусловленности:

$$\nabla f = \begin{pmatrix} 2x_1 \\ 200x_2 \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} 2 & 0 \\ 0 & 200 \end{pmatrix}, \quad \kappa = \frac{200}{2} = 100$$

(b) Градиентный спуск, $\alpha^* = 2/(200+2) = 1/101 \approx 0{,}0099$**:**
$k=0$: $\nabla f(1,1) = (2, 200)^\top$, $\mathbf{d} = (-2, -200)^\top$.

$$\mathbf{x}^{(1)} = \begin{pmatrix}1\\1\end{pmatrix} + \frac{1}{101}\begin{pmatrix}-2\\-200\end{pmatrix} = \begin{pmatrix}1-2/101\\1-200/101\end{pmatrix} \approx \begin{pmatrix}0{,}9802\\-0{,}9802\end{pmatrix}$$

$f(\mathbf{x}^{(1)}) = (0{,}9802)^2 + 100(0{,}9802)^2 = 101 \cdot (0{,}9802)^2 \approx 96{,}93$. (Начальное значение функции: $1 + 100 = 101$.)
$k=1$: $\nabla f(\mathbf{x}^{(1)}) \approx (1{,}960, -196{,}04)^\top$.

$$\mathbf{x}^{(2)} \approx \begin{pmatrix}0{,}9608\\0{,}9608\end{pmatrix}, \quad f(\mathbf{x}^{(2)}) \approx 93{,}97$$

$k=2$: аналогично: $\mathbf{x}^{(3)} \approx (0{,}9417, -0{,}9417)^\top$, $f \approx 89{,}06$.
Коэффициент уменьшения на итерацию: $\left(\frac{\kappa-1}{\kappa+1}\right)^2 = \left(\frac{99}{101}\right)^2 \approx 0{,}9610$ – очень медленная сходимость!
(c) Шаг Ньютона (один шаг):

$$\mathbf{d}^{(0)} = -\mathbf{H}^{-1}\nabla f(\mathbf{x}^{(0)}) = -\begin{pmatrix}1/2 & 0\\0 & 1/200\end{pmatrix}\begin{pmatrix}2\\200\end{pmatrix} = -\begin{pmatrix}1\\1\end{pmatrix}$$

$$\mathbf{x}_N^{(1)} = \begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix}-1\\-1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} = \mathbf{x}^* \quad \checkmark$$

Метод Ньютона достигает минимума за **один шаг**, градиентному спуску требуется > 10000 итераций!
(d) Эффект зигзага:
При большом κ линии уровня представляют собой сильно вытянутые эллипсы. Градиент перпендикулярен линиям уровня и направлен вдоль короткой оси эллипса (здесь – направление x_2 с большим $\lambda = 200$). Каждый шаг градиентного спуска «перескакивает» в направлении x_2 и вынужден корректироваться на следующей итерации $\Rightarrow$ зигзаг. Метод Ньютона корректирует различия в кривизне с помощью матрицы Гессе.
Ньютон: 1 итерация. Градиентный спуск: $\approx$ тысячи. Число обусловленности $\kappa = 100$ проблематично.

3 Уровень 3: Сложный

Задача 3.1 – Седловая точка и анализ сходимости

Задача 3.1

Дана функция $f(x_1, x_2) = x_1^2 - x_2^2 + 2x_1x_2$.

(a) Рассчитай все стационарные точки.

(b) Классифицируй стационарную точку с помощью матрицы Гессе и объясни, о чём говорит знак $\det(\mathbf{H})$.

(c) Какие проблемы возникают при применении метода Ньютона к этой задаче? Рассчитай шаг Ньютона из $\mathbf{x}^{(0)} = (1, 0)^\top$ и объясни результат.

(d) Как ведёт себя градиентный спуск вблизи седловой точки? Сформулируй общее утверждение.

Решение: Задача 3.1

(a) Стационарные точки:

$$\nabla f = \begin{pmatrix} 2x_1 + 2x_2 \\ -2x_2 + 2x_1 \end{pmatrix} = \mathbf{0} \Rightarrow \begin{cases} x_1 + x_2 = 0 \\ x_1 - x_2 = 0 \end{cases} \Rightarrow x_1 = x_2 = 0$$

Единственная стационарная точка: $\mathbf{x}^* = (0, 0)^\top$.

(b) Классификация:

$$\mathbf{H} = \begin{pmatrix} 2 & 2 \\ 2 & -2 \end{pmatrix}, \quad \det(\mathbf{H}) = -4 - 4 = -8 < 0$$

Поскольку $\det(\mathbf{H}) < 0$, $\mathbf{H}$ имеет одно положительное и одно отрицательное собственное значение $\Rightarrow$ **$\mathbf{H}$ неопределена** $\Rightarrow$ $(0, 0)$ – **седловая точка** (не минимум и не максимум).

Собственные значения: $\lambda_{1,2} = \frac{(2-2) \pm \sqrt{(2+2)^2 + 4 \cdot 4}}{2} = \pm\sqrt{4+4} \cdot \frac{1}{1} = \pm 2\sqrt{2}$.

Функция f неограничена как сверху, *так* и снизу.

(c) Шаг Ньютона в седловой точке:

$$\nabla f(1, 0) = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \mathbf{H}^{-1} = \frac{1}{-8} \begin{pmatrix} -2 & -2 \\ -2 & 2 \end{pmatrix} = \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & -1/4 \end{pmatrix}$$

$$\mathbf{d} = -\mathbf{H}^{-1} \nabla f = -\begin{pmatrix} 1/4 & 1/4 \\ 1/4 & -1/4 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = -\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$\mathbf{x}^{(1)} = (1, 0)^\top + (-1, 0)^\top = (0, 0)^\top$$

Метод Ньютона сходится к седловой точке за один шаг! Это проблема: метод Ньютона находит **стационарные точки**, но не

различает минимумы, максимумы и седловые точки. Проверка $\det(\mathbf{H})$ обязательна.

(d) Градиентный спуск вблизи седловых точек:
Градиентный спуск теоретически **не** сходится к седловым точкам (поскольку $-\nabla f$ указывает направление убывания, а седловая точка не является локальным минимумом). Однако при определённых начальных точках он может медленно дрейфовать к седловой точке, прежде чем отклониться от неё. На практике: старт точно на устойчивом подпространстве $\rightarrow$ сходимость к седловой точке; в остальных случаях – расхождение от неё.

$(0,0)$ **– седловая точка. Метод Ньютона находит её за 1 шаг – но это не минимум!**

Задача 3.2 – Функция Розенброка: численный анализ

Задача 3.2

Функция Розенброка $f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ – классическая тестовая задача.

(a) Покажи, что $\mathbf{x}^* = (1,1)^\top$ – единственный глобальный минимум с $f^* = 0$.

(b) Рассчитай $\mathbf{H}$ в минимуме $\mathbf{x}^* = (1,1)^\top$ и определи число обусловленности.

(c) Объясни, почему градиентный спуск на функции Розенброка чрезвычайно медленный.

(d) Полностью рассчитай первый шаг Ньютона из $\mathbf{x}^{(0)} = (0,0)^\top$.

(e) Сравни качественно: сколько итераций требуется градиентному спуску и методу Ньютона соответственно?

Решение: Задача 3.2

(a) Глобальный минимум:
$f(\mathbf{x}) \geq 0$ для всех $\mathbf{x}$ (сумма квадратов). $f(\mathbf{x}) = 0$ тогда и только тогда, когда:

$$x_2 - x_1^2 = 0 \quad \text{и} \quad 1 - x_1 = 0 \Rightarrow x_1 = 1,\ x_2 = 1$$

Таким образом, $(1,1)^\top$ – **единственный глобальный минимум** с $f^* = 0$.

(b) Матрица Гессе в минимуме:
В общем виде (из теории):

$$\mathbf{H} = \begin{pmatrix} -400(x_2 - x_1^2) + 800x_1^2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}$$

В точке $\mathbf{x}^* = (1,1)^\top$: $x_2 - x_1^2 = 0$:

$$\mathbf{H}(\mathbf{x}^*) = \begin{pmatrix} 800+2 & -400 \\ -400 & 200 \end{pmatrix} = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix}$$

$\det(\mathbf{H}^*) = 802 \cdot 200 - 400^2 = 160400 - 160000 = 400 > 0$, след $= 1002 > 0 \Rightarrow$ $\mathbf{H}^* \succ 0$ (минимум подтверждён).
Собственные значения: характеристический многочлен $(\lambda - 802)(\lambda - 200) - 160000 = 0$:

$$\lambda^2 - 1002\lambda + 400 = 0 \Rightarrow \lambda_{1,2} = \frac{1002 \pm \sqrt{1002^2 - 1600}}{2}$$

$$\lambda_1 \approx 1001{,}6, \quad \lambda_2 \approx 0{,}3996$$

$$\kappa(\mathbf{H}^*) = \frac{\lambda_1}{\lambda_2} \approx \frac{1001{,}6}{0{,}3996} \approx \mathbf{2506}$$

Чрезвычайно высокое число обусловленности!
(c) Медленная сходимость градиентного спуска:
Функция Розенброка имеет узкую изогнутую долину вдоль параболы $x_2 = x_1^2$. Градиент направлен поперёк долины (поскольку линии уровня очень плотно расположены вдоль долины). Градиентный спуск:

- быстро приближается ко дну долины (спуск в направлении x_2);
- затем чрезвычайно медленно продвигается вдоль долины (очень малые градиенты в направлении долины);
- скорость сходимости $\approx \left(\frac{2506-1}{2506+1}\right)^2 \approx 0{,}9984$ – всего 0,16% уменьшения за шаг!

(d) Шаг Ньютона из $\mathbf{x}^{(0)} = (0,0)^\top$:

$$\nabla f(0,0) = \begin{pmatrix} -400 \cdot 0 \cdot (0-0) - 2(1-0) \\ 200(0-0) \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

$$\mathbf{H}(0,0) = \begin{pmatrix} -400(0) + 0 + 2 & 0 \\ 0 & 200 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 200 \end{pmatrix}$$

Шаг Ньютона:

$$\mathbf{d} = -\mathbf{H}^{-1}\nabla f = -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/200 \end{pmatrix}\begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\mathbf{x}^{(1)} = (0,0)^\top + 1 \cdot (1,0)^\top = (1,0)^\top$$

$f(1,0) = 100(0-1)^2 + 0 = 100$. Минимум ещё не достигнут – и результат даже хуже, чем в начале.
Действительно: $f(0,0) = 100 \cdot 0 + 1 = 1$. После шага Ньютона: $f(1,0) = 100$. Здесь метод Ньютона ухудшил результат!

Матрица Гессе в начальной точке недостаточно положительно определена для глобальной сходимости с полным шагом.
С линейным поиском: $\alpha < 1$ уменьшило бы f. При $\alpha \approx 0{,}01$: $\mathbf{x}^{(1)} \approx (0{,}01, 0)^\top$, $f \approx 0{,}98$ – более удачная начальная точка для последующих шагов Ньютона.
(e) Сравнение числа итераций (типичное):

Метод	Типичное число итераций	Достижимая точ
Градиентный спуск	10 000–100 000	$\|\nabla f\| \approx 10^{-3}$
Ньютон (с линейным поиском)	20–50	$\|\nabla f\| \approx 10^{-1}$

$\mathbf{x}^* = (1, 1)^\top$, $\kappa \approx 2506$, Ньютон: ≈ 30 итераций, градиентный спуск: $> 10\,000$ итераций.
