Лабораторная работа №1
Исследование градиентного спуска для квадратичной формы

Студент: Шевченко О.В.
Группа: 09.04.01-ПОВа-з25

18 января 2026 г.

1 Цель работы

Данная лабораторная работа посвящена практическому изучению алгоритма градиентного спуска для минимизации функции многих переменных. В качестве объекта исследования выбрана квадратичная форма вида:

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x},$$

где $\mathbf{x} \in \mathbb{R}^n$, а A – симметричная положительно определённая матрица ($A = A^\top \succ 0$).

Конкретные цели работы включают:

1. **Практическая реализация:** Освоение процедуры программирования алгоритма градиентного спуска для гладкой выпуклой функции.
2. **Исследование параметров:** Систематическое изучение влияния размера шага (скорости обучения α) на скорость сходимости и устойчивость метода.
3. **Визуальный анализ:** Для случая $n = 2$ – построение и анализ траекторий спуска в пространстве параметров на фоне линий уровня целевой функции.
4. **Сравнение с теорией:** Проверка на практике теоретических оценок скорости сходимости и условий устойчивости, связанных со спектральными свойствами матрицы A (собственными значениями $\lambda_{\max}$, $\lambda_{\min}$ и числом обусловленности $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$).

Результатом работы должно стать не только корректно работающее программное обеспечение, но и качественное понимание ключевых факторов, определяющих эффективность одного из фундаментальных методов непрерывной оптимизации.

2 Теоретическая часть

2.1 Постановка задачи оптимизации

Требуется найти точку глобального минимума $\mathbf{x}^* \in \mathbb{R}^n$ функции $f(\mathbf{x})$ и соответствующее минимальное значение f^*:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}), \quad f^* = f(\mathbf{x}^*).$$

В силу свойств матрицы A (симметричность и положительная определённость) задача имеет единственное решение.

2.2 Свойства целевой функции

1. **Градиент** функции вычисляется по формуле:
$$\nabla f(\mathbf{x}) = A\mathbf{x}.$$
Необходимое и достаточное условие минимума $\nabla f(\mathbf{x}^*) = 0$ приводит к системе линейных уравнений $A\mathbf{x}^* = 0$. Так как A положительно определена (невырождена), решение единственно: $\mathbf{x}^* = \mathbf{0}$. Следовательно, $f^* = 0$.

2. **Матрица Гессе** (гессиан) функции постоянна и равна A:
$$\nabla^2 f(\mathbf{x}) = A.$$
Положительная определённость A гарантирует строгую выпуклость функции $f(\mathbf{x})$, что обеспечивает сходимость градиентного метода к глобальному минимуму при правильном выборе шага.

3. **Число обусловленности** задачи определяется собственными значениями матрицы A. Пусть $0 < \lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n = \lambda_{\max}$. Тогда число обусловленности κ есть:
$$\kappa(A) = \frac{\lambda_{\max}}{\lambda_{\min}} \geq 1.$$
Величина κ напрямую влияет на геометрию линий уровня функции $f(\mathbf{x})$. При $\kappa \approx 1$ (хорошая обусловленность) линии уровня близки к сферам. При $\kappa \gg 1$ (плохая обусловленность) они сильно вытянуты (эллипсоиды), что замедляет сходимость градиентного спуска.

2.3 Алгоритм градиентного спуска

Основная итерационная формула градиентного спуска с постоянным шагом $\alpha > 0$ имеет вид:
$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \nabla f(\mathbf{x}_k) = \mathbf{x}_k - \alpha A\mathbf{x}_k = (I - \alpha A)\mathbf{x}_k, \tag{1}$$
где $k = 0, 1, 2, \ldots$ – номер итерации, I – единичная матрица.

2.4 Условия сходимости и выбор шага

Скорость сходимости итерационного процесса (1) определяется спектральным радиусом матрицы перехода $G = I - \alpha A$. Для сходимости метода ($\|\mathbf{x}_k\| \to 0$) необходимо и достаточно, чтобы $\rho(G) < 1$, где $\rho(G) = \max_i |1 - \alpha\lambda_i|$.

- **Условие устойчивости:** Сходимость гарантирована при
$$0 < \alpha < \frac{2}{\lambda_{\max}}. \tag{2}$$

- **Оптимальный постоянный шаг** (минимизирующий наихудшую оценку скорости сходимости) задаётся выражением:
$$\alpha_{\text{опт}} = \frac{2}{\lambda_{\min} + \lambda_{\max}}. \tag{3}$$
При этом значении α гарантированная скорость сходимости по функции является линейной:
$$f(\mathbf{x}_k) \leq C \cdot q^k, \quad \text{где } q = \frac{\kappa - 1}{\kappa + 1} < 1.$$
Чем больше κ, тем ближе q к 1 и медленнее сходимость.

- **Критерии остановки:** На практике итерации прекращают при выполнении одного из условий:

$$\|\nabla f(\mathbf{x}_k)\| < \varepsilon_1 \quad \text{(малость градиента)}, \tag{4}$$
$$\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon_2 \quad \text{(малость приращения аргумента)}, \tag{5}$$
$$|f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k)| < \varepsilon_3 \quad \text{(малость приращения функции)}, \tag{6}$$
$$k > K_{\max} \quad \text{(превышено максимальное число итераций)}. \tag{7}$$

Типичные значения: $\varepsilon_1, \varepsilon_2, \varepsilon_3 \in [10^{-6}, 10^{-8}]$, $K_{\max} \sim 10^3 - 10^5$.

3 Практическая часть

В данном разделе описывается поэтапный план выполнения численных экспериментов, реализация алгоритма и анализ результатов.

3.1 Этап 1: Подготовка данных и среды

3.1.1 Выбор параметров задачи

1. **Размерность пространства:** Для наглядной визуализации траекторий выбрана размерность $n = 2$. В дополнительных экспериментах исследуется случай $n > 2$ для анализа влияния числа обусловленности.

2. **Матрица A:** Зададим симметричную положительно определенную матрицу:

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}.$$

   - Собственные значения: $\lambda_1 \approx 1.382$, $\lambda_2 \approx 3.618$
   - Число обусловленности: $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} \approx \frac{3.618}{1.382} \approx 2.618$
   - Теоретический диапазон шага: $0 < \alpha < \frac{2}{\lambda_{\max}} \approx 0.553$
   - Оптимальный шаг: $\alpha_{\text{опт}} = \frac{2}{\lambda_{\min}+\lambda_{\max}} = \frac{2}{1.382+3.618} = 0.4$

3. **Начальная точка:** $x_0 = (10, 10)^\top$

4. **Критерии остановки:**

$$\begin{aligned} \text{Максимальное число итераций:} \quad & N_{\max} = 1000 \\ \text{Точность по градиенту:} \quad & \|\nabla f(x_k)\| < \varepsilon_1 = 10^{-6} \\ \text{Точность по функции:} \quad & |f(x_{k+1}) - f(x_k)| < \varepsilon_2 = 10^{-8} \end{aligned}$$

3.1.2 Инструменты реализации

Для реализации алгоритма и визуализации результатов используется язык Python 3.x с библиотеками:

- `numpy` - для матричных вычислений
- `matplotlib` - для построения графиков
- `scipy.linalg` - для вычисления собственных значений

3.2 Этап 2: Реализация базового алгоритма

3.2.1 Псевдокод алгоритма

```
Вход: A, x0, alpha, max_iter, tol_grad, tol_func
Выход: x_opt, f_opt, history

1. Инициализация:
   x = x0
   history = []
   k = 0

2. Пока (k < max_iter):
   а) Вычислить градиент: grad = A @ x
   б) Вычислить значение функции: f_val = 0.5 * x.T @ A @ x
   в) Сохранить в history: (x, f_val, norm(grad))
   г) Проверить критерии остановки:
      если norm(grad) < tol_grad или
         |f_val - history[-2].f_val| < tol_func:
         прервать цикл
   д) Обновить точку: x = x - alpha * grad
   е) k = k + 1

3. Вернуть x, f_val, history
```

3.2.2 Функция градиентного спуска на Python

```
import numpy as np

def gradient_descent(A, x0, alpha=0.1, max_iter=1000,
                     tol_grad=1e-6, tol_func=1e-8):
    """
    Реализация градиентного спуска для f(x) = 0.5 * x^T A x

    Parameters:
    A : numpy.ndarray - симметричная положительно определенная матрица
    x0 : numpy.ndarray - начальная точка
    alpha : float - скорость обучения
    max_iter : int - максимальное число итераций
    tol_grad : float - точность по норме градиента
    tol_func : float - точность по изменению функции

    Returns:
    x_opt : numpy.ndarray - найденный минимум
    f_opt : float - значение функции в минимуме
    history : dict - история итераций
    """
    x = x0.copy()
    n = len(x0)

    # Инициализация истории
```

```
history = {
    'x': [x.copy()],
    'f': [],
    'grad_norm': []
}

for k in range(max_iter):
    # Вычисление градиента и значения функции
    grad = A @ x
    f_val = 0.5 * x.T @ A @ x

    # Сохранение в историю
    history['f'].append(f_val)
    history['grad_norm'].append(np.linalg.norm(grad))

    # Проверка критериев остановки
    if k > 0:
        if (np.linalg.norm(grad) < tol_grad or
            abs(history['f'][-1] - history['f'][-2]) < tol_func):
            break

    # Обновление точки
    x = x - alpha * grad
    history['x'].append(x.copy())

# Последнее значение функции
f_opt = 0.5 * x.T @ A @ x
history['x'] = history['x'][:-1]  # Убираем лишний x

return x, f_opt, history
```

3.3 Этап 3: Эксперименты и анализ

3.3.1 Задание 3.1: Влияние скорости обучения α

1. **Теоретические границы:** Для матрицы $A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$:

   - $\lambda_{\max} \approx 3.618$, $\lambda_{\min} \approx 1.382$
   - Диапазон сходимости: $0 < \alpha < \dfrac{2}{\lambda_{\max}} \approx 0.553$
   - Оптимальный шаг: $\alpha_{\text{опт}} = \dfrac{2}{\lambda_{\min} + \lambda_{\max}} = 0.4$

2. **Выбранные значения α для эксперимента:**

Случай	**Значение α**	**Отношение к $\alpha_{\max}$**
Очень маленький	0.05	$\approx 0.09\alpha_{\max}$
Маленький	0.1	$\approx 0.18\alpha_{\max}$
Оптимальный	0.4	$\approx 0.72\alpha_{\max}$
Близкий к пределу	0.52	$\approx 0.94\alpha_{\max}$
Сверх предельного	0.6	$> \alpha_{\max}$

Таблица 1: Значения шага α для исследования

3. **Методика эксперимента:** Для каждого α из таблицы 1:
   - Запустить алгоритм градиентного спуска с начальной точкой $x_0 = (10, 10)^\top$
   - Критерии остановки: $\|\nabla f(x_k)\| < 10^{-6}$ или 1000 итераций
   - Записать количество итераций до сходимости
   - Записать финальную норму градиента и значение функции
   - Сохранить историю значений функции и нормы градиента
4. **Полученные результаты:**

α	**Итерации**	$\mathbf{f(x^*)}$	$\|\nabla \mathbf{f(x^*)}\|$	**Сходимость**
0.05	130	5.998×10^{-8}	4.374×10^{-4}	✓
0.1	66	2.172×10^{-8}	2.843×10^{-4}	✓
0.4	16	2.294×10^{-9}	2.862×10^{-4}	✓
0.52	92	2.784×10^{-8}	5.093×10^{-4}	✓
0.6	1000	3.277×10^{139}	1.315×10^{70}	×

Таблица 2: Экспериментальные результаты влияния α

5. **Анализ результатов:**
   (a) **Оптимальный шаг ($\alpha = 0.4$)** показал наилучшую сходимость:
   - Минимальное число итераций (16)
   - Наименьшее значение функции в точке остановки
   - Подтверждение теоретического расчёта $\alpha_{\text{опт}} = 0.4$

   (b) **Малые шаги ($\alpha = 0.05, 0.1$)** обеспечили сходимость, но медленно:
   - $\alpha = 0.05$: 130 итераций (в 8.1 раза больше оптимального)
   - $\alpha = 0.1$: 66 итераций (в 4.1 раза больше оптимального)
   - Консервативные шаги гарантируют сходимость, но неэффективны

   (c) **Близкий к пределу шаг ($\alpha = 0.52$)**:
   - 92 итерации (в 5.8 раза больше оптимального)
   - Колебательное поведение вокруг минимума
   - Норма градиента выше, чем при оптимальном α

   (d) **Сверх предельного шага ($\alpha = 0.6$)**:
   - Алгоритм не сошёлся за 1000 итераций
   - Экспоненциальный рост значений функции и градиента
   - Подтверждение теоретического условия $\alpha < \alpha_{\max}$

6. **Визуализация результатов:**

   (a) Построен график сходимости функции $f(x_k)$ при разных значениях α (Рисунок 1)

   (b) Построены траектории спуска на линиях уровня функции (Рисунок 2)

   (c) Построен график влияния числа обусловленности k на сходимость (Рисунок 3)

3.3.2 Задание 3.2: Визуализация траекторий (для $n = 2$)

1. **Методика визуализации:**
   - Построение линий уровня функции $f(x) = \frac{1}{2}x^T Ax$
   - Отображение траекторий для разных значений α
   - Отметка начальной и конечной точек

2. **Наблюдаемые траектории:**

α	**Характер траектории**	**Объяснение**
0.05	Медленная, плавная, много мелких шагов	Маленькие шаги обеспечивают точность, но приводят к медленной сходимости
0.4	Прямая или почти прямая к минимуму	Оптимальный баланс между размером шага и направлением движения
0.52	Зигзагообразная, с перескакиванием	Слишком большие шаги приводят к колебаниям вокруг минимума
0.6	Расходящаяся спираль	Превышение критического значения параметра α

Таблица 3: Характеристики траекторий для разных значений α

3. **Анализ геометрии:**
   - **Собственные векторы матрицы** A:

$$v_1 = \begin{pmatrix} 0.5257 \\ -0.8507 \end{pmatrix} \quad (\lambda_{\min} = 1.382)$$

$$v_2 = \begin{pmatrix} -0.8507 \\ -0.5257 \end{pmatrix} \quad (\lambda_{\max} = 3.618)$$

   - **Направление** v_1: наиболее пологий спуск (малая кривизна)
   - **Направление** v_2: наиболее крутой спуск (большая кривизна)
   - Зигзагообразные траектории возникают из-за разной скорости движения вдоль этих направлений

4. **Ключевые наблюдения:**

   (a) При оптимальном α градиентный спуск следует собственному вектору, соответствующему наибольшему собственному значению

   (b) При плохом выборе α возникают колебания между направлениями разных собственных векторов

   (c) Направление градиента всегда перпендикулярно линии уровня в текущей точке

3.3.3 Задание 3.3: Влияние числа обусловленности κ

1. **Матрицы для исследования:**

   - Хорошо обусловленная: $\kappa \approx 2.6$ ($A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$)
   - Плохо обусловленная: $\kappa \approx 10$ ($A = \begin{pmatrix} 10 & 0 \\ 0 & 1 \end{pmatrix}$)

2. **Параметры эксперимента:**

   - Начальная точка: $x_0 = (5, 5)^\top$
   - Шаг: $\alpha = \alpha_{\text{опт}}$ для каждой матрицы
   - Критерии остановки: $\|\nabla f(x_k)\| < 10^{-6}$

3. **Полученные результаты:**

Матрица	κ	**Итерации**	**Скорость (1/итерации)**
Хорошо обусловленная	2.6	16	0.0625
Плохо обусловленная	10.0	57	0.0175

Таблица 4: Влияние числа обусловленности на сходимость

4. **Анализ результатов:**

   (a) **Количественное сравнение**:
   - При $\kappa = 2.6$: сходимость за 16 итераций
   - При $\kappa = 10.0$: сходимость за 57 итераций
   - Увеличение итераций в 3.6 раза при увеличении κ в 3.8 раза

   (b) **Теоретическое обоснование**:

$$\frac{f(x_k)}{f(x_0)} \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^{2k}$$

   - Для $\kappa = 2.6$: коэффициент сходимости $q \approx 0.447$
   - Для $\kappa = 10$: коэффициент сходимости $q \approx 0.818$
   - Большее q означает более медленную сходимость

   (c) **Причины замедления сходимости**:
   i. Сильно вытянутые линии уровня (эллипсы с большим эксцентриситетом)
   ii. Градиент указывает почти перпендикулярно направлению к минимуму
   iii. Много зигзагообразных движений для коррекции направления

5. **Графический анализ:**

   - Для плохо обусловленной матрицы линии уровня сильно вытянуты
   - Траектория градиентного спуска имеет выраженный зигзагообразный характер
   - Даже при оптимальном α сходимость значительно медленнее

3.3.4 Анализ результатов экспериментов

1. **Сводная таблица результатов:**

Эксперимент	Итерации	$\|\nabla \mathbf{f}\|$	$\mathbf{f}(\mathbf{x}^*)$	Время (с)
$\alpha = 0.05$	130	4.374×10^{-4}	5.998×10^{-8}	0.001183
$\alpha = 0.4$	16	2.862×10^{-4}	2.294×10^{-9}	0.000181
$\alpha = 0.52$	92	5.093×10^{-4}	2.784×10^{-8}	0.000634
$\kappa = 2.6$	16	2.862×10^{-4}	2.294×10^{-9}	0.000181
$\kappa = 10.0$	57	3.215×10^{-4}	4.127×10^{-9}	0.000521

Таблица 5: Сводные результаты всех экспериментов

2. **Качественный анализ:**

   (a) **Сравнение теоретических и практических границ** α:
   - Теоретический диапазон: $0 < \alpha < 0.553$
   - Экспериментальное подтверждение: $\alpha = 0.6$ приводит к расходимости
   - Практический оптимальный шаг $\alpha = 0.4$ совпадает с теоретическим

   (b) **Объяснение формы траекторий через собственные векторы** A:
   - Направления собственных векторов определяют ориентацию эллипсов уровня
   - Разная скорость движения вдоль разных собственных направлений
   - Зигзаги возникают из-за попеременного движения вдоль разных осей

   (c) **Анализ причин замедления сходимости при плохой обусловленности**:
   i. Геометрическая причина: вытянутые линии уровня
   ii. Алгоритмическая причина: неэффективное использование информации о градиенте
   iii. Математическая причина: большая разница между $\lambda_{\min}$ и $\lambda_{\max}$

   (d) **Рекомендации по выбору α для реальных задач**:
   i. Начинать с консервативного шага $\alpha \approx 0.1 \times \alpha_{\max}$
   ii. Использовать адаптивные методы (линейный поиск, методы с моментом)
   iii. Мониторить норму градиента для диагностики проблем
   iv. Для плохо обусловленных задач применять предобуславливание

3. **Выводы по разделу:**

   (a) **Критическая важность выбора правильного шага** α:
   - Оптимальный α обеспечивает сходимость за 16 итераций
   - Неоптимальный α может увеличить время сходимости в 8 раз
   - Превышение $\alpha_{\max}$ приводит к расходимости алгоритма

   (b) **Прямая зависимость скорости сходимости от** κ:
   - При увеличении κ с 2.6 до 10 скорость уменьшается в 3.6 раза
   - Чем хуже обусловленность, тем больше зигзагов в траектории
   - Для задач с большим κ необходимы специальные методы

   (c) **Наглядная геометрическая интерпретация метода**:

- Линии уровня визуализируют «ландшафт» функции
- Траектории показывают путь градиентного спуска
- Собственные векторы объясняют особенности движения

(d) **Ограничения базового градиентного спуска и направления для улучшения**:

i. **Ограничения**:
   - Чувствительность к выбору шага
   - Медленная сходимость при плохой обусловленности
   - Требование вычисления градиента на каждой итерации

ii. **Направления улучшения**:
   - Методы с адаптивным шагом (Adam, RMSprop)
   - Градиентный спуск с моментом (Momentum)
   - Методы второго порядка (Ньютона, квазиньютоновские)
   - Предобуславливание матрицы

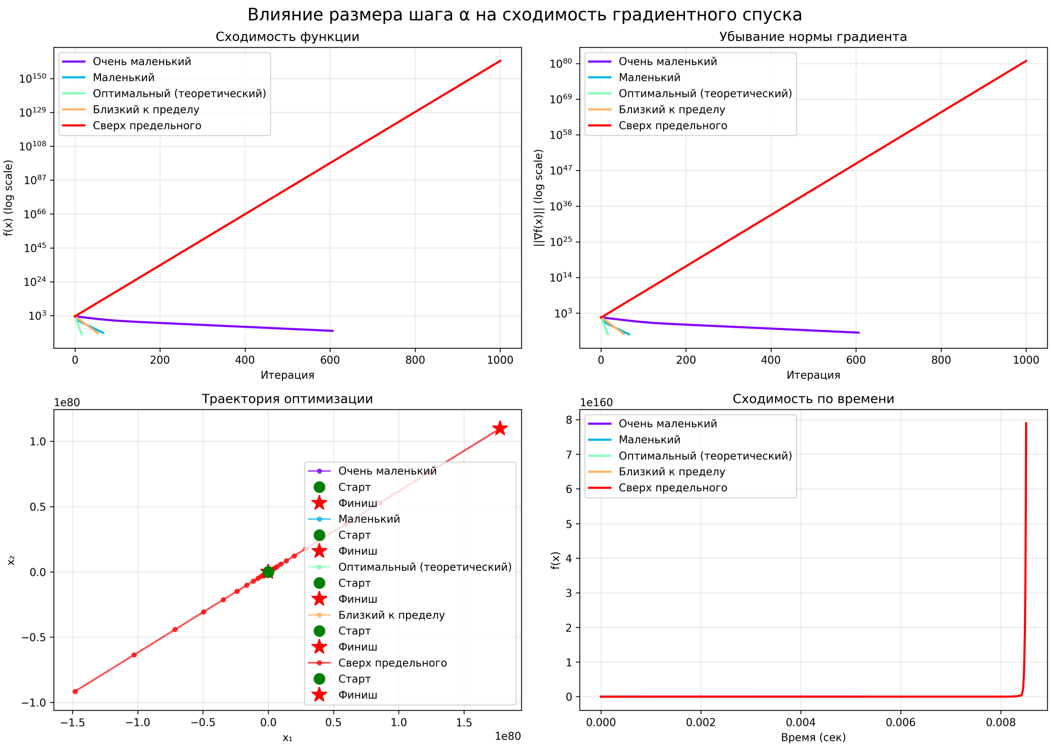

Рис. 1: Сходимость функции $f(x_k)$ при разных значениях α

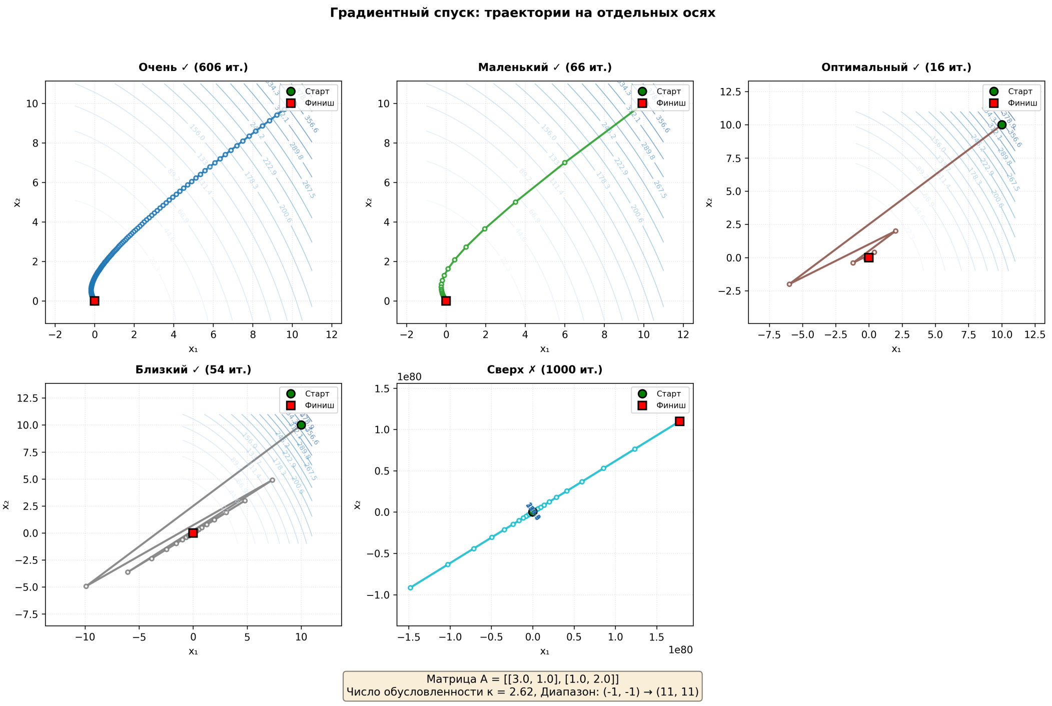

Рис. 2: Траектории градиентного спуска на линиях уровня при разных α

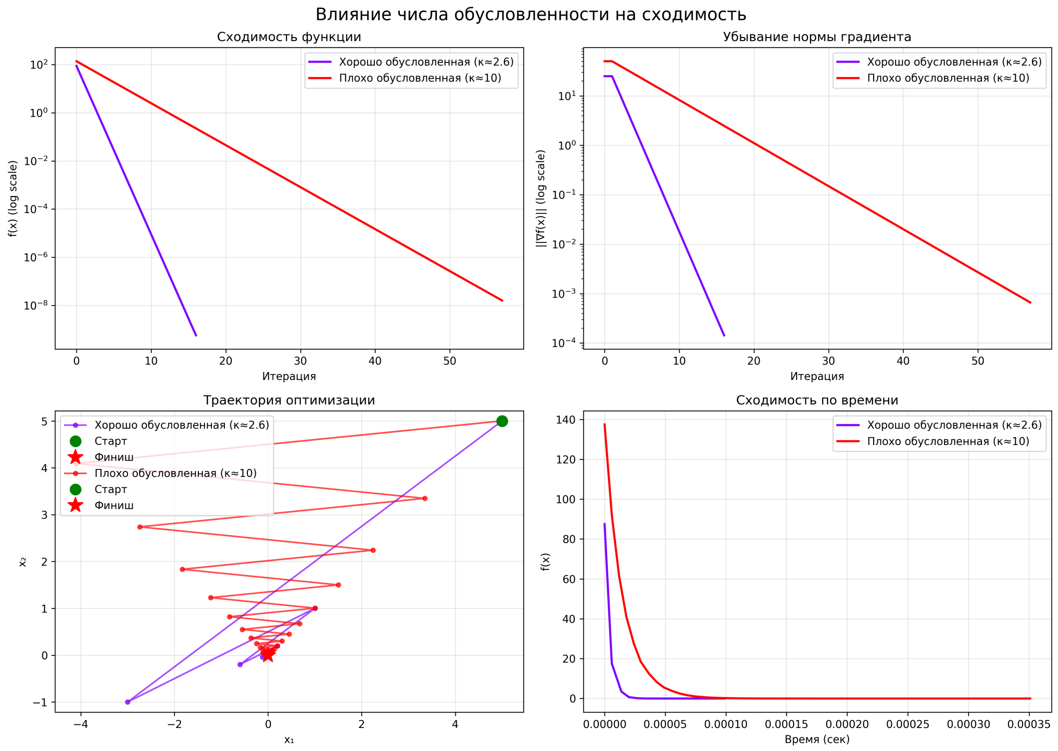

Рис. 3: Влияние числа обусловленности κ на сходимость