

# Методы оптимизации

## Лабораторная работа №1

Лаврентьев В. Ю., ИТМО, М-4151

### 1. Формулировка задания

#### 1.1. Реализовать методы:

- метод градиентного спуска (`optimization.gradient_descent()`);
- процедура линейного спуска (`optimization.LineSearchTool.line_search()`)
- метод Ньютона (`optimization.newton()`)

#### 1.2. Реализовать оракул логистической регрессии

Функция задана в виде:

$$\frac{1}{m} \sum \ln(1 + \exp(-bAx)) + \frac{\lambda}{2} \|x\|_2^2 \quad (1)$$

Градиент:

$$\frac{\lambda}{2} x - A^T \left( \frac{b}{1 + \exp(-bAx)} \right) \frac{1}{m} \quad (2)$$

Гессиан:

$$\frac{A^T \text{diag}(bAx(1 - bAx))}{m + \frac{\lambda}{2} I} \quad (3)$$

- 1.3. Реализовать подсчет разностных производных (`oracles.grad_finite_diff()`, `oracles.hess_finite_diff()`). Выполнить проверку.

### 2. Эксперименты

#### 2.1. Траектория градиентного спуска на квадратичной функции

Проанализируйте траекторию градиентного спуска для нескольких квадратичных функций: придумайте две-три квадратичные двумерные функции, на которых работа метода будет отличаться, нарисуйте графики с линиями уровня функций и траекториями методов. Попробуйте ответить на следующий вопрос: как отличается поведение метода в зависимости от числа обусловленности функции, выбора начальной точки и стратегии выбора шага (константная стратегия, Армихо, Вульф)?

Анализ выполнен на трех вариантах квадратичной функции, заданной различными матрицами  $A$  (с разным числом обусловленности) и векторами  $b$ :

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, b = [1 \quad 1], \text{cond}(A) = 3 \quad (4.1)$$

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, b = [1 \quad 1], \text{cond}(A) = 1 \quad (4.2)$$

$$A = \begin{bmatrix} 11 & 9 \\ 9 & 11 \end{bmatrix}, b = [2 \quad 2], \text{cond}(A) = 10 \quad (4.3)$$

В качестве начальных точек взяты  $x_0 = (4, 4)$ ,  $x_0 = (2, -1)$ ,  $x_0 = (0, 0)$

Построены графики траекторий градиентного спуска (рис 1, 2, 3).

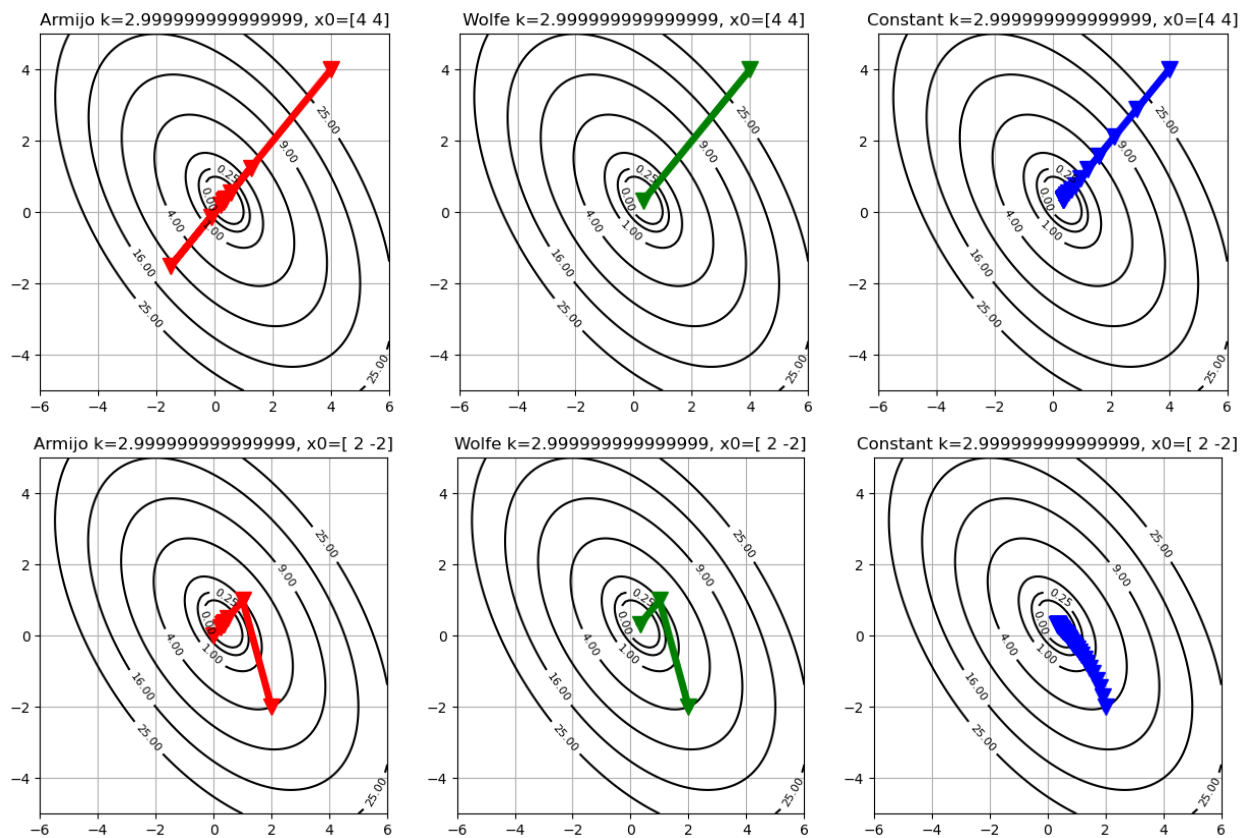


Рисунок 1. Траектория градиентного спуска для функции, заданной (4.1)

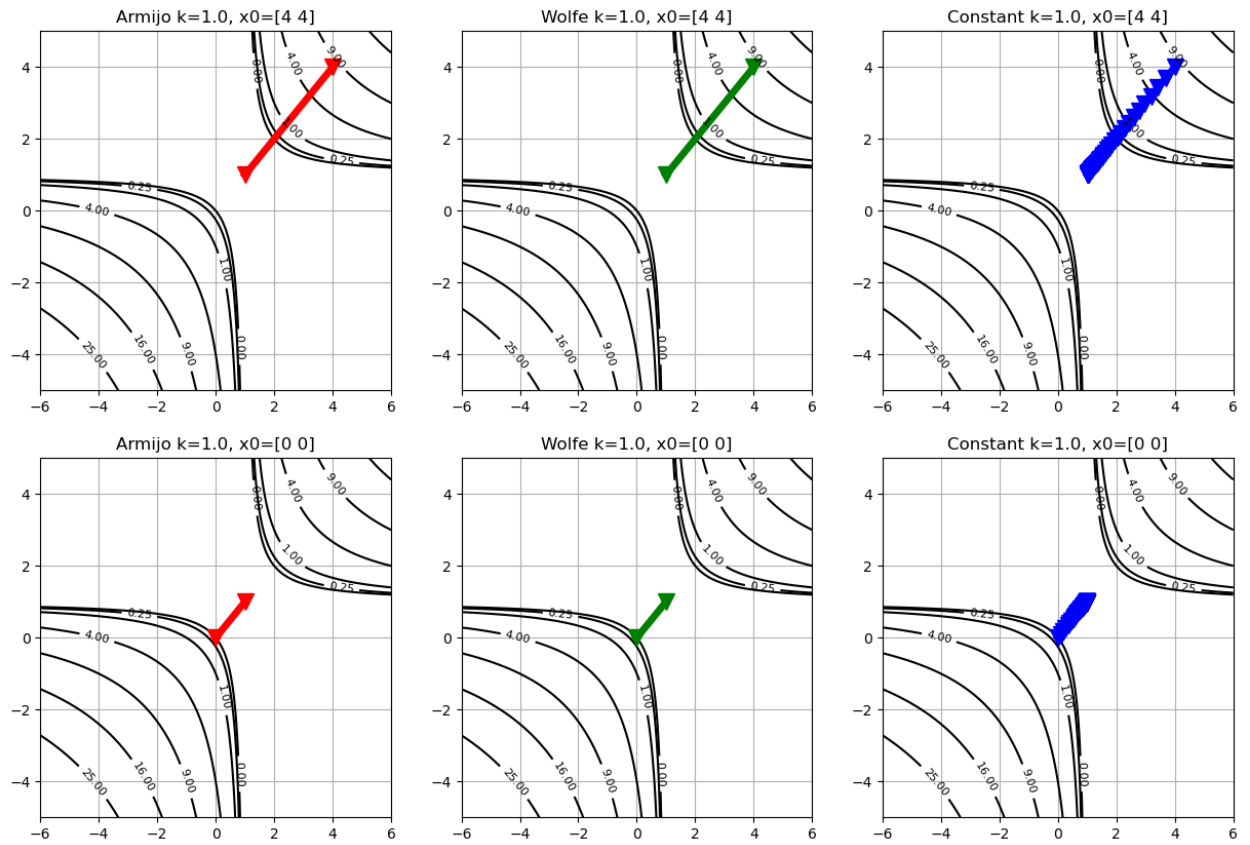


Рисунок 2. Траектория градиентного спуска для функции, заданной (4.2)

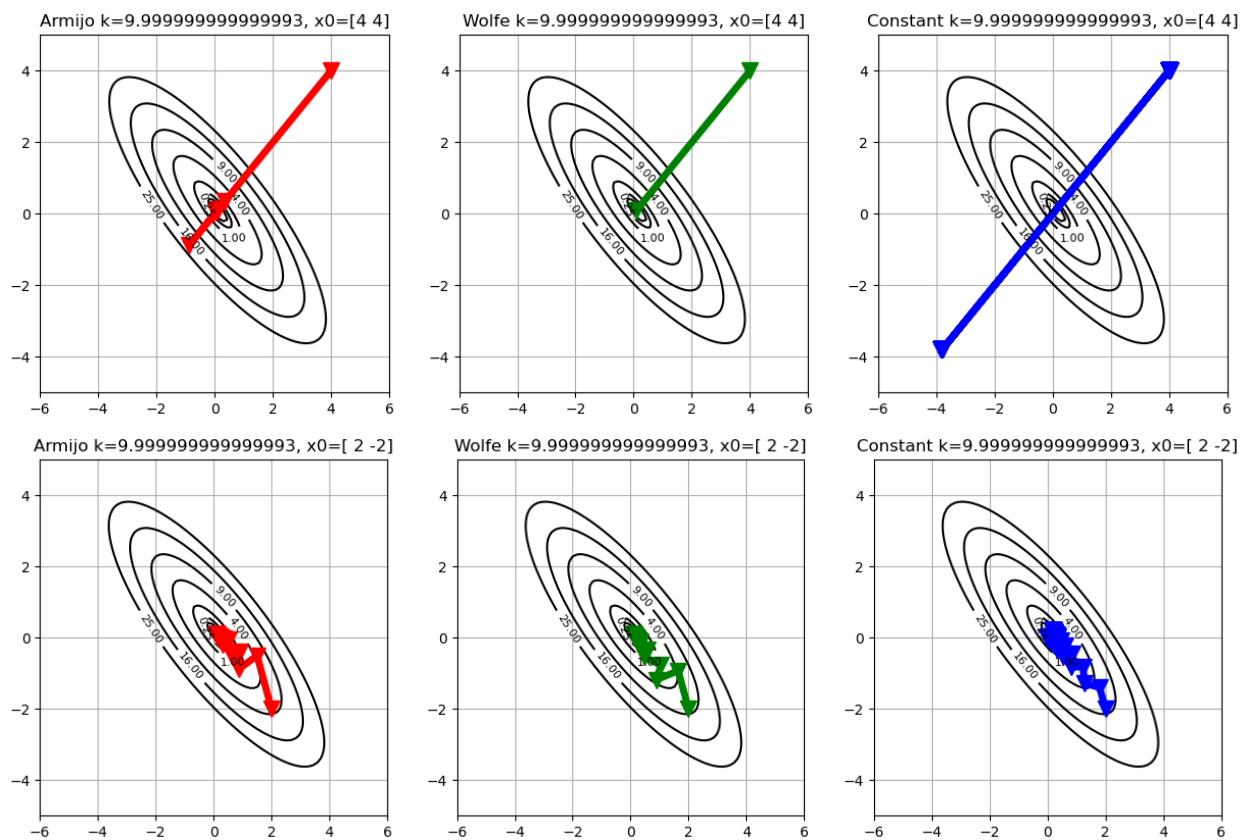


Рисунок 3. Траектория градиентного спуска для функции, заданной (4.3)

Проведенный эксперимент показывает, что все стратегии метода выбора шага сходятся к оптимальной точке. Для стратегии константного шага наблюдалось заметно большее число итераций, чем для стратегии Армихо. Спуск по стратегии Вульфа завершался за самое меньшее число итераций. Для всех трех стратегий прослеживается некоторый рост числа итераций по мере возрастания числа обусловленности матрицы  $A$ , задающей функцию.

## 2.2. Зависимость числа итераций градиентного спуска от числа обусловленности и размерности пространства

Исследуйте, как зависит число итераций, необходимое градиентному спуску для сходимости, от следующих двух параметров: 1) числа обусловленности  $\kappa \geq 1$  оптимизируемой функции и 2) размерности пространства  $n$  оптимизируемых переменных.

Для этого для заданных параметров  $n$  и  $\kappa$  сгенерируйте случайным образом квадратичную задачу размера  $n$  с числом обусловленности  $\kappa$  и запустите на ней градиентный спуск с некоторой фиксированной требуемой точностью. Замерьте число итераций  $T(n, \kappa)$ , которое потребовалось сделать методу до сходимости (успешному выходу по критерию остановки).

Для выполнения эксперимента была выполнена генерация с параметрами по сетке в диапазонах:

- размерность пространства: [1, 10, 100, 1000]
- число обусловленности: от 1 до 100 с шагом 20.

Собственно, функция задавалась случайным образом исходя из конкретного сочетания названных выше параметров. В целях минимизации влияния случайного результата, для каждого сочетания параметров (размерность пространства/число обусловленности) процедура градиентного спуска запускалась 100 раз и полученные значения числа итераций усреднялись. Итоговые данные представлены на рис. 4.

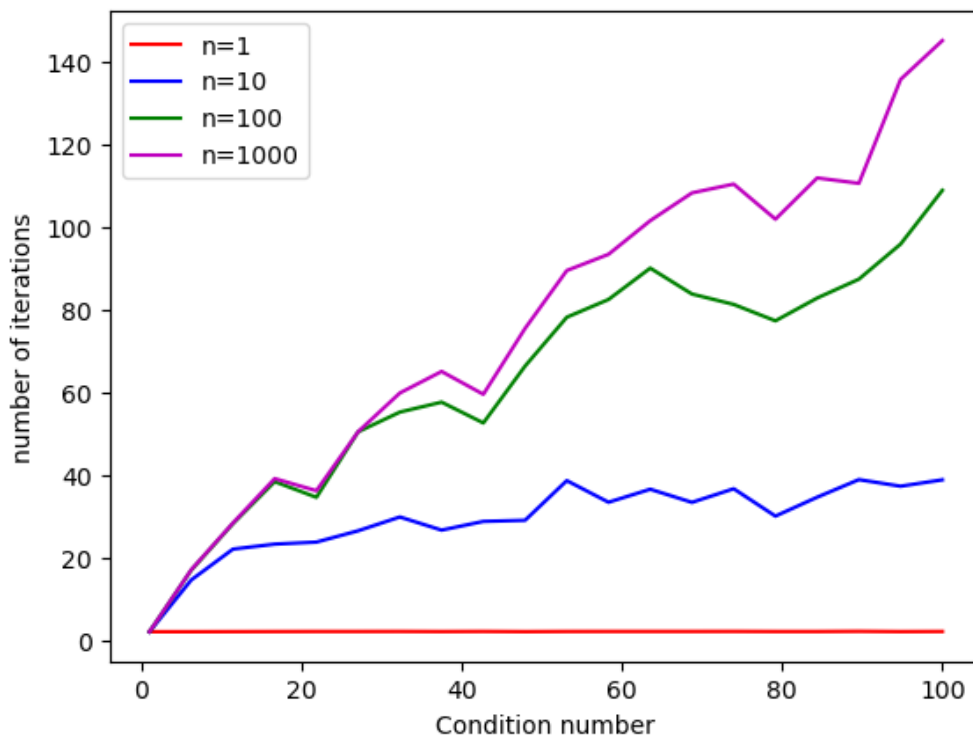


Рисунок 4. Зависимость числа итераций от размерности пространства и числа обусловленности

Результаты эксперимента показывают линейный рост числа итераций, потребовавшихся до сходимости градиентного спуска от числа обусловленности матрицы, задающей квадратичную функцию. При этом при одном и том же значении числа обусловленности изменения размерности пространства влияния не оказывает.

### **2.3. Сравнение методов градиентного спуска и Ньютона на реальной задаче логистической регрессии**

Сравнить методы градиентного спуска и Ньютона на задаче обучения логистической регрессии на реальных данных. В качестве реальных данных используйте следующие три набора с сайта LIBSVM: w8a, gisette и real-sim. Коэффициент регуляризации взять стандартным образом:  $\lambda = \frac{1}{m}$ .

Параметры обоих методов взять равными параметрам по умолчанию. Начальную точку выбрать  $x_0 = 0$ .

Построить графики сходимости следующих двух видов:

- (a) Зависимость значения функции от реального времени работы метода.
- (b) Зависимость относительного квадрата нормы градиента  $\|\nabla f(x_k)\|_2^2 / \|\nabla f(x_0)\|_2^2$  (в логарифмической шкале) против реального времени работы.

Построенные графики сходимости показаны на рис. 5, 6, 7

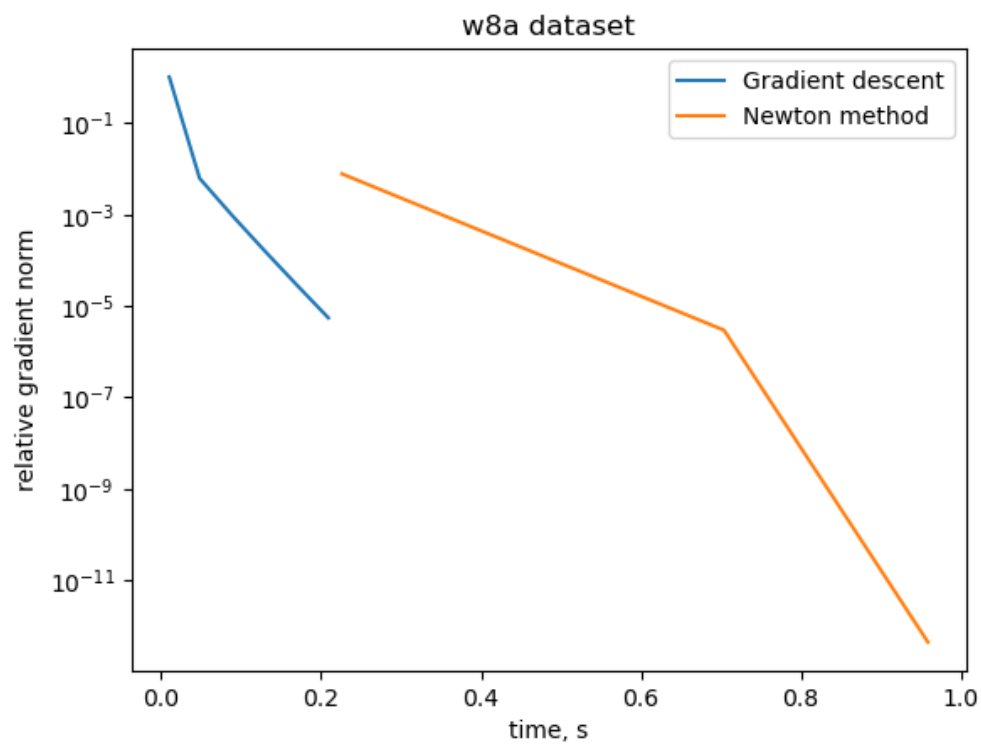
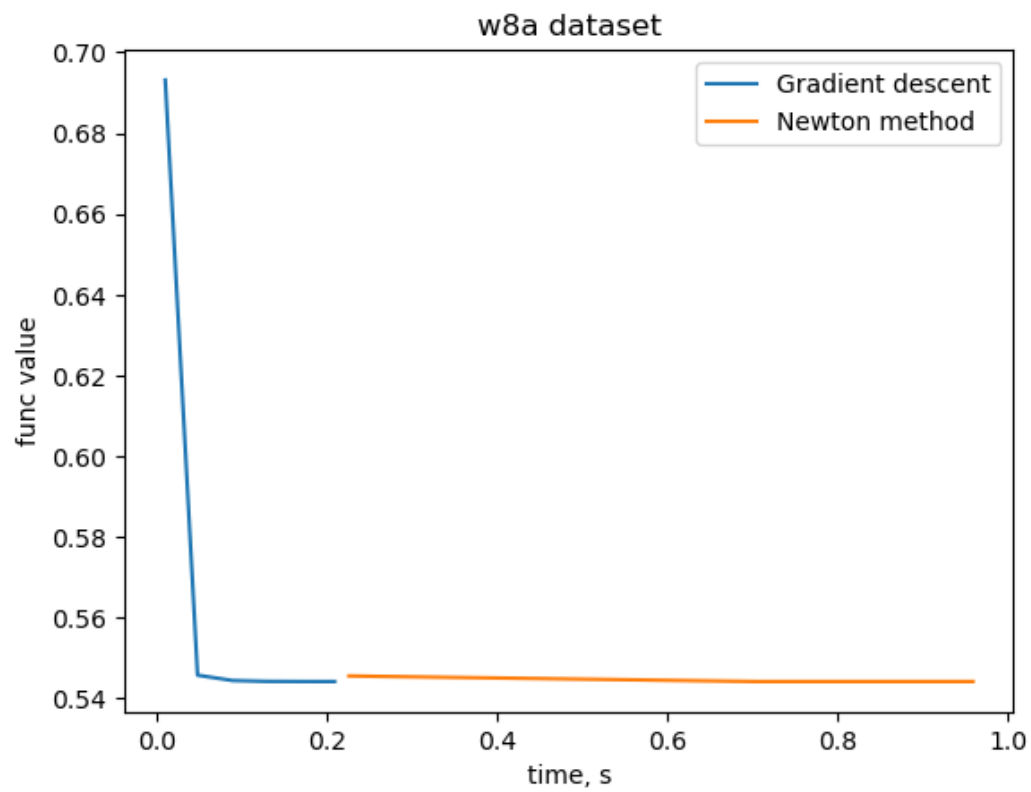


Рисунок 5. Графики сходимости (значение функции / время работы сверху, относительная норма градиента / время работы внизу) для набора данных w8a

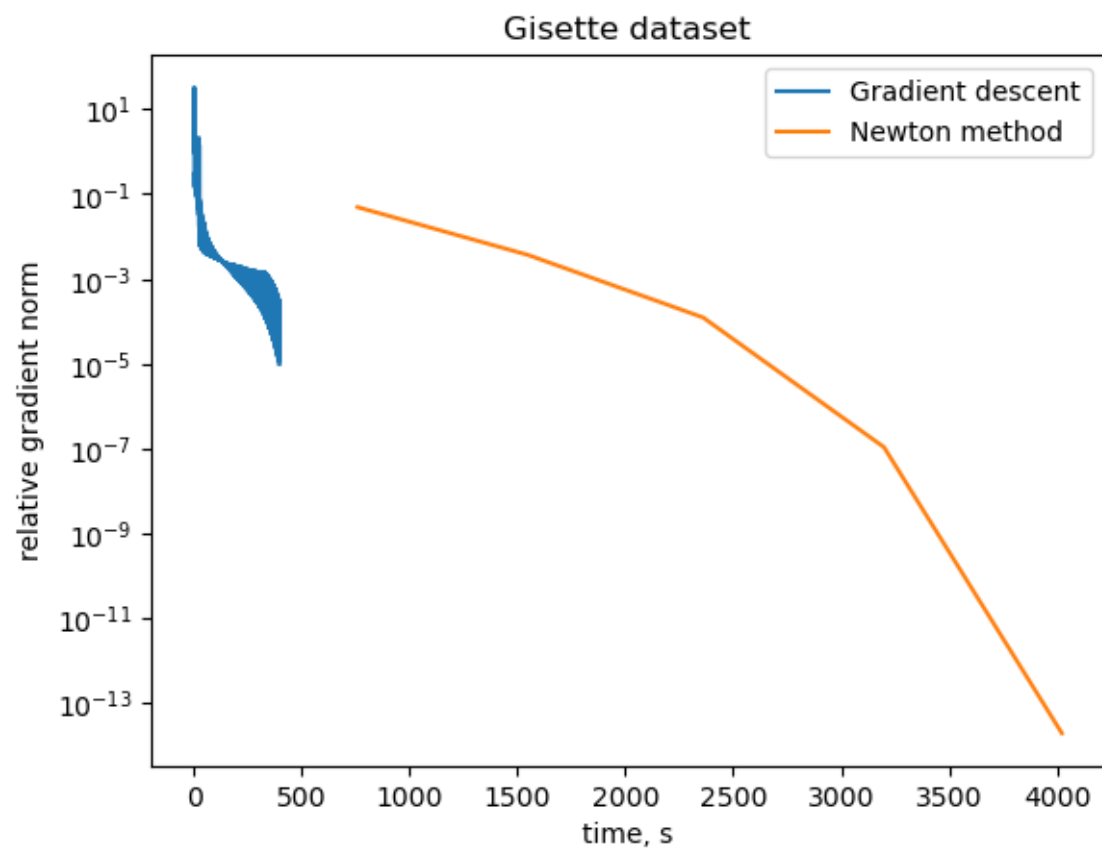
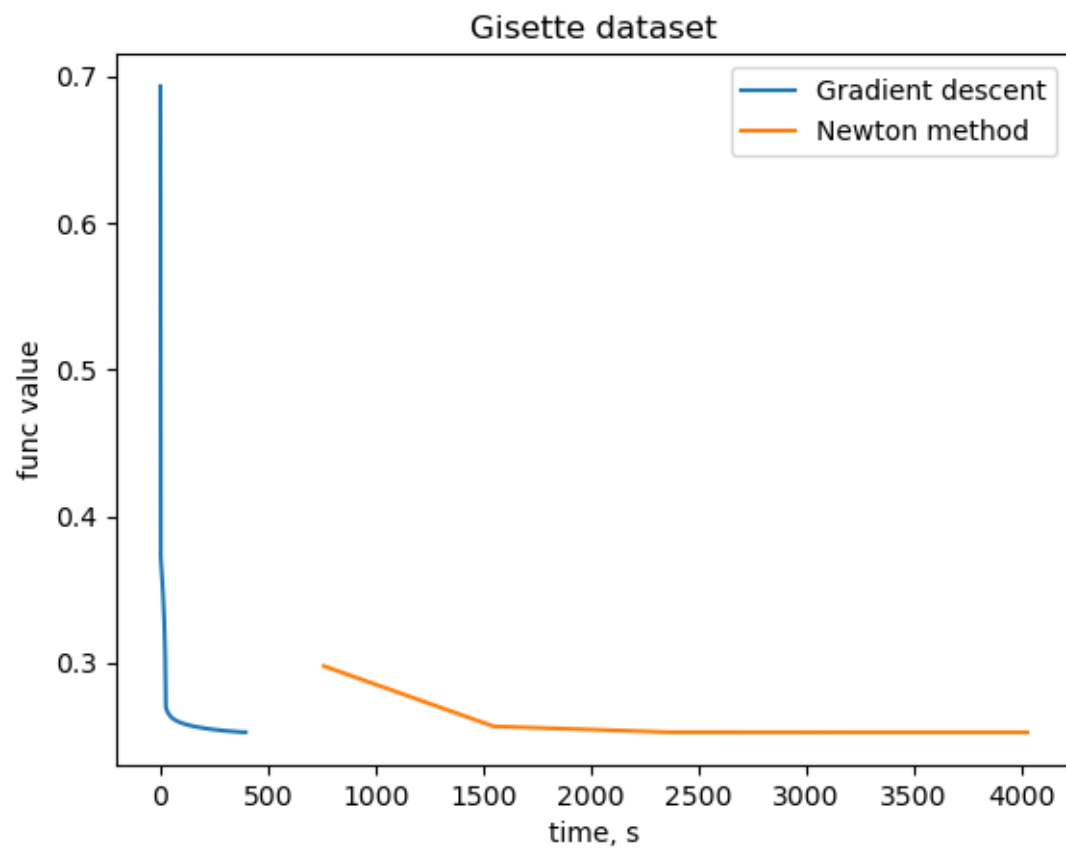


Рисунок 6. Графики сходимости (значение функции / время работы вверх, относительная норма градиента / время работы вниз) для набора данных Gisette

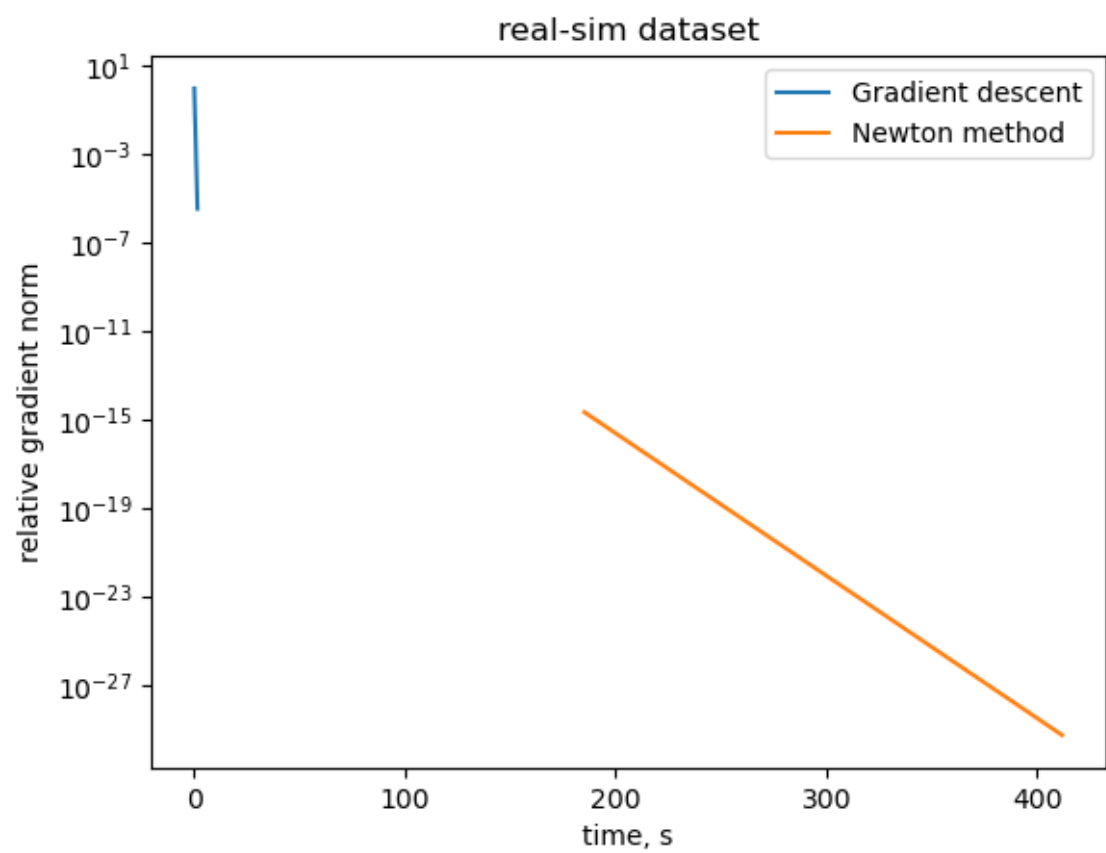
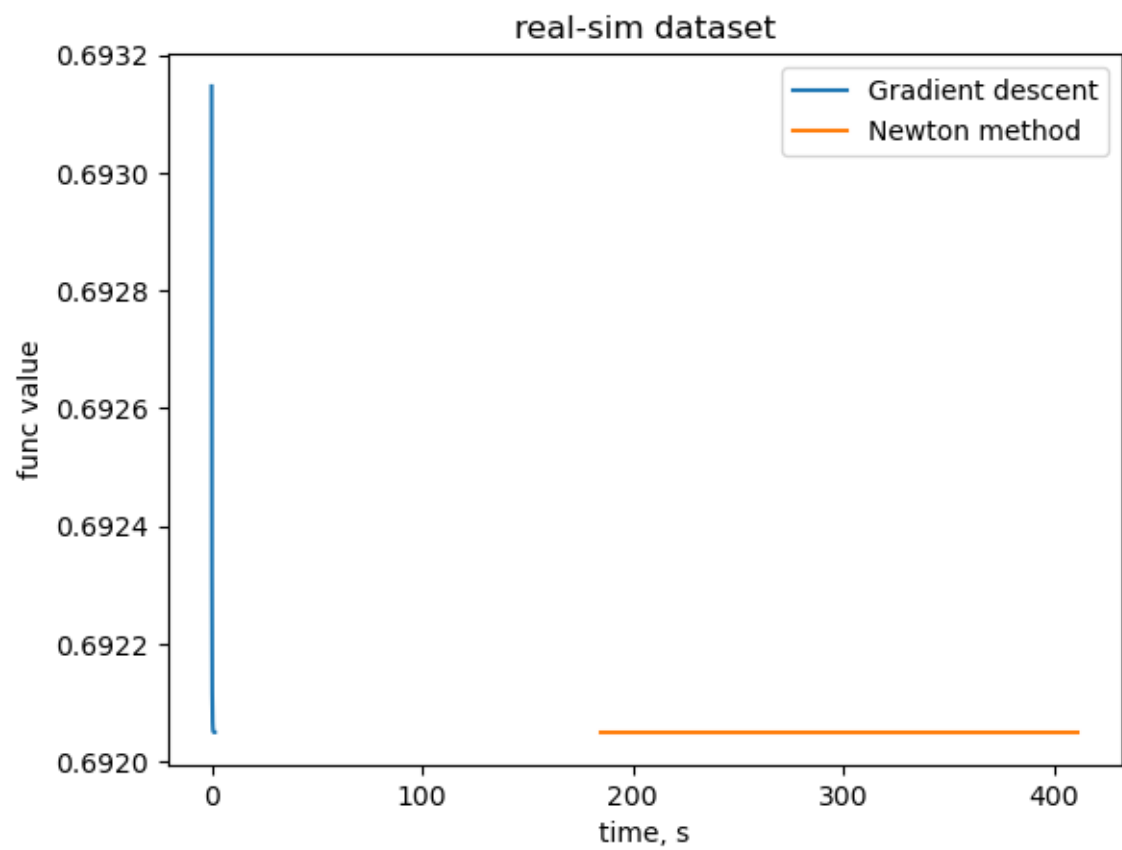


Рисунок 7. Графики сходимости (значение функции / время работы вверху, относительная норма градиента / время работы внизу) для набора данных real-sim



Оценка стоимости представлена в таблице 1.

Таблица 1. Оценка стоимостей метода градиентного спуска и метода Ньютона на трех рассмотренных наборах данных

	Градиентный спуск		Метод Ньютона	
	Память	Итерации	Память	Итерации
W8a размерность задачи $q=49749$ , размерность пространства $n=300$	$O(300)$	$O(49749) + O(300)$	$O(300^2)$	$O(300^3) + O(49749 \cdot 300^2)$
Gisette Размерность задачи $q=6000$ Размерность пространства $n=5000$	$O(5000)$	$O(6000) + O(5000)$	$O(5000^2)$	$O(5000^3) + O(6000 \cdot 5000^2)$
Real-sim Размерность задачи $q=72309$ Размерность пространства $n=20958$	$O(20958)$	$O(72309) + O(20958)$	$O(20958^2)$	$O(20958^3) + O(72309 \cdot 20958^2)$

По результатам эксперимента можно сделать вывод о том, что более простой метод (градиентный спуск) работает быстрее в начале, но достигает меньше точности. В противоположность этому, метод Ньютона показывает более медленный «разгон» в начале работы, но увеличивает скорость сходимости по мере приближения к оптимальной точке, при этом показывая более высокую точность.

#### 2.4. Стратегия выбора длины шага в градиентном спуске

Исследовать, как зависит поведение метода от стратегии подбора шага: константный шаг (попробовать различные значения), бэктрейнинг (попробовать различные константы  $c_1$ ), условия Вульфа (попробовать различные параметры  $c_2$ ).

Рассмотрите квадратичную функцию и логистическую регрессию с модельными данными (сгенерированными случайно).

Запустите для этих функций градиентный спуск с разными стратегиями выбора шага из одной и той же начальной точки. Нарисуйте кривые сходимости (относительная невязка по функции в логарифмической шкале против числа итераций – для квадратичной функции, относительный квадрат нормы градиента в логарифмической шкале против числа итераций – для логистической регрессии) для разных стратегий на одном графике.

Результаты эксперимента показаны на рис. 8,

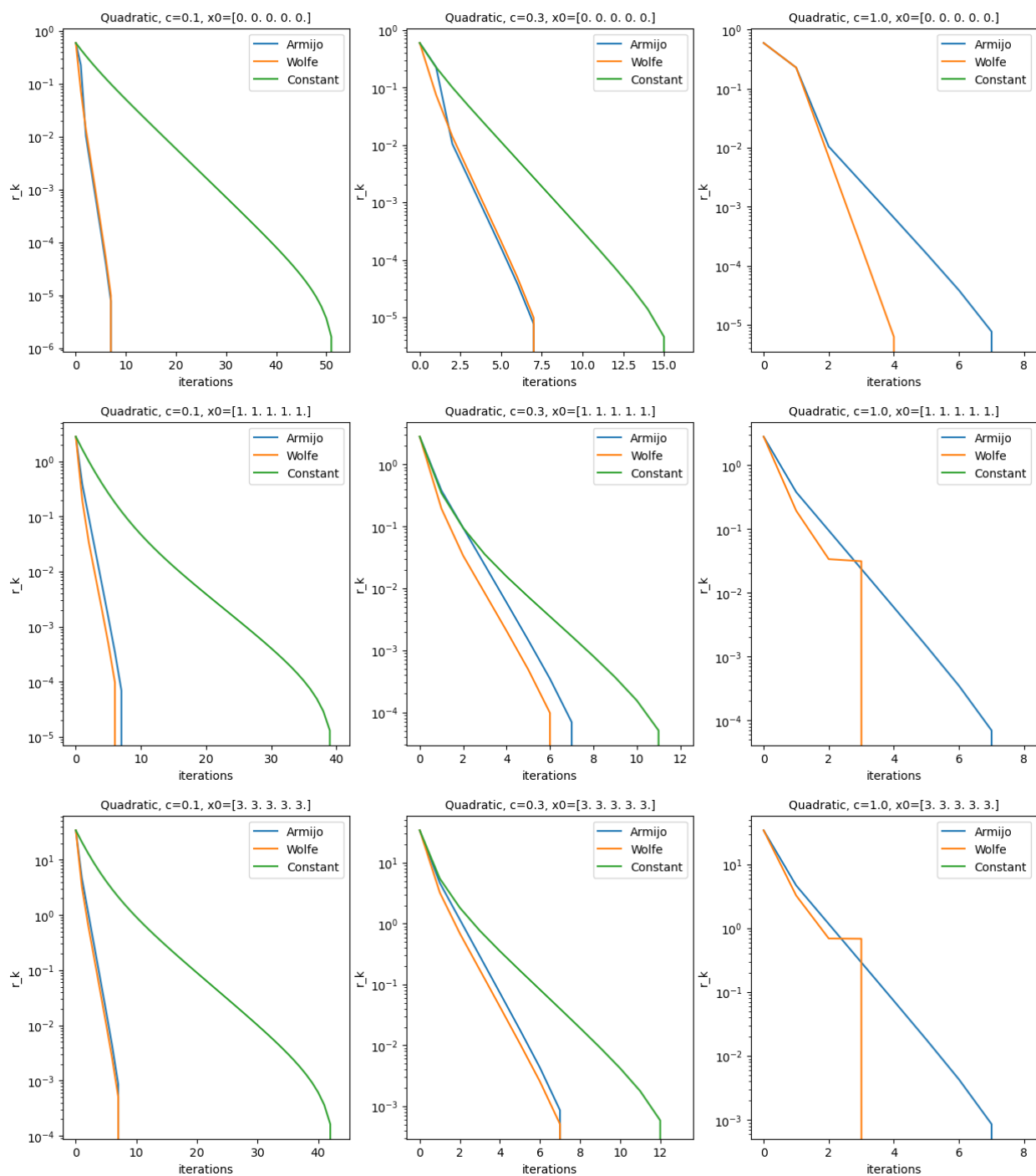


Рисунок 8. Сходимость градиентного спуска на квадратичной функции при различных стратегиях выбора шага и начальных точках

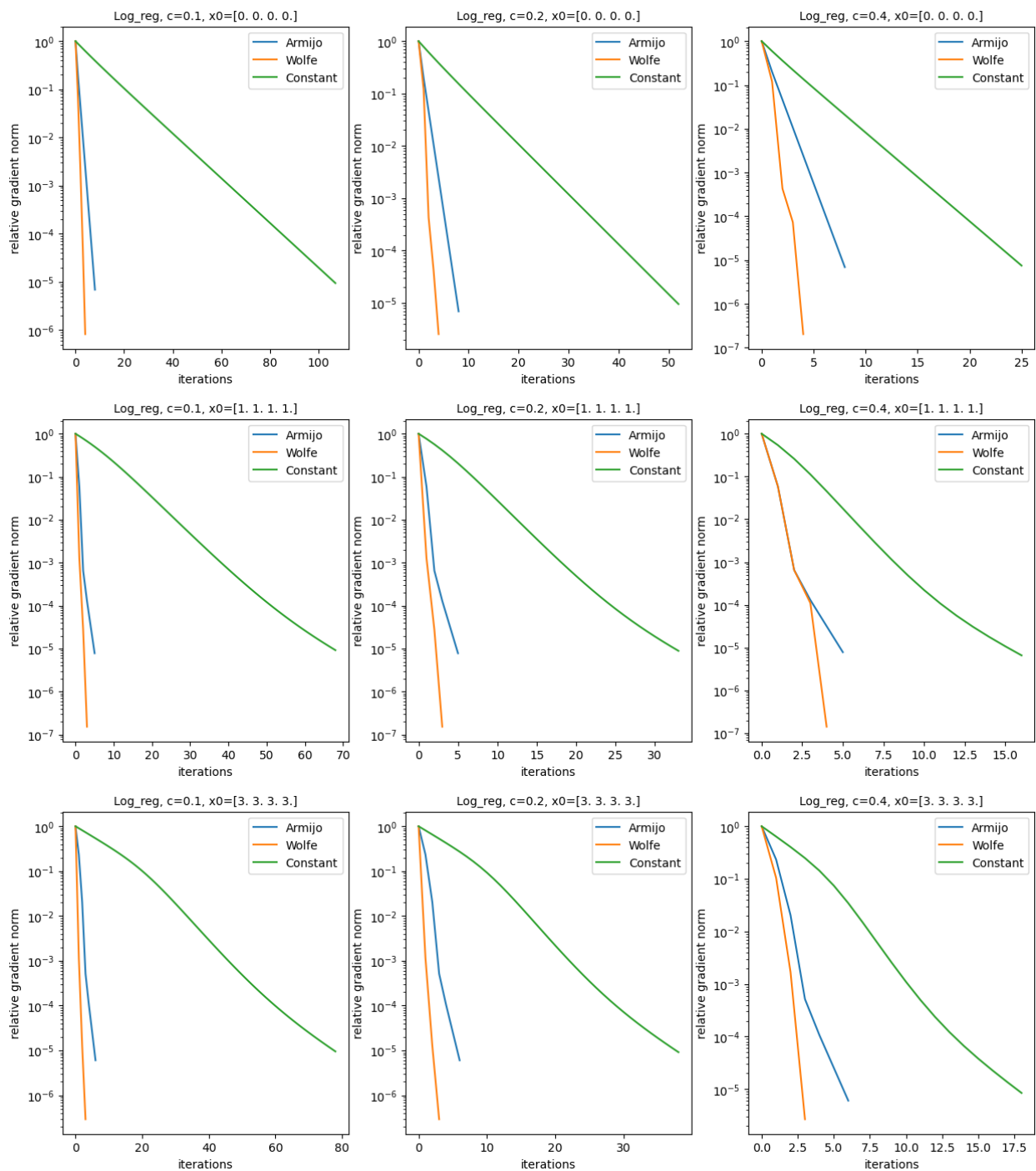


Рисунок 9. Сходимость градиентного спуска на логистической регрессии при различных стратегиях выбора шага и начальных точках

Результаты эксперимента во всех рассмотренных случаях демонстрируют преимущество стратегии Вульфа. Далее с незначительным (в ряде случаев практически «один в один» отрывом идет стратегия Армихо. В обоих названных случаях выбор той или иной начальной точки принципиально ситуацию не меняет. Длина шага также не оказывает заметного влияния, что вполне объяснимо, т.к. обе эти стратегии оптимальный размер шага определяют внутри своего алгоритма.

Значительное отставание по скорости сходимости показывает стратегия постоянного шага. При выборе излишне большой длины шага (1.0 и более) метод начинает расходиться.

## 2.5. Стратегия выбора шага в методе Ньютона

Повторите эксперимент 2.4 со сравнением стратегий выбора шага на логистической регрессии с модельной выборкой, но для метода Ньютона.

Графики, полученные в ходе эксперимента приведены на рис. 10.

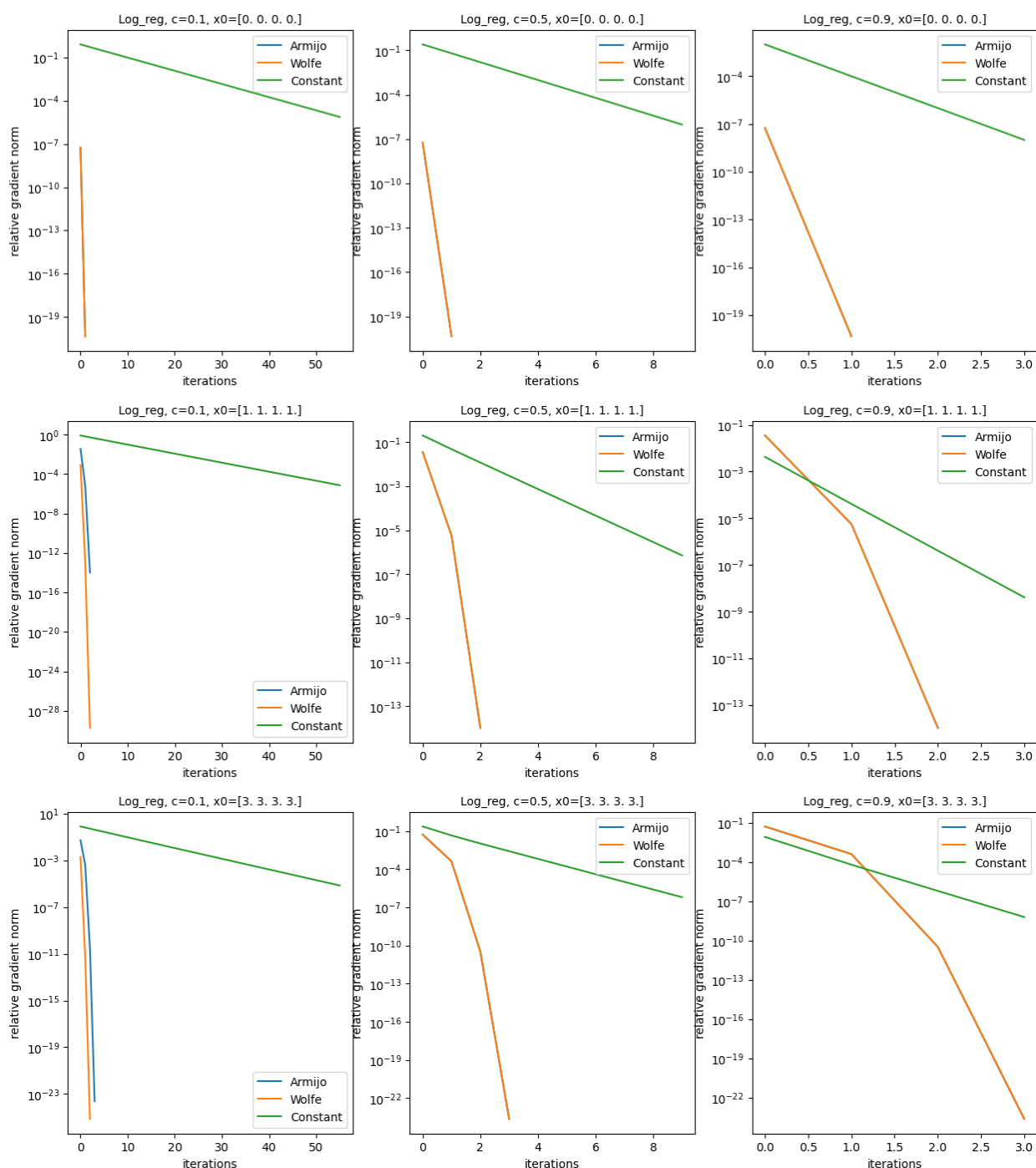


Рисунок 10. Сходимость метода Ньютона при различных стратегиях выбора шага и начальной точки.

Как и в предыдущем эксперименте, принципиальных изменений в ситуации не произошло – наилучшие результаты показывает стратегия Вульфа и, чуть хуже, стратегия Армихо. Изменение величины шага (шаг в любом случае определяется внутри алгоритма) / выбор начальной точки заметного влияния не оказывает.

Стратегия выбора постоянного шага значительно отстает по скорости сходимости.