

# Лабораторная работа 1

на тему

## **Методы градиентного спуска и метод** **Ньютона**

Выполнила:  
студентка 1 курса магистратуры  
Шахвалиева Юлиана Сергеевна



\_\_\_\_\_  
(подпись)

## ОГЛАВЛЕНИЕ

Формулы для градиента и гессиана .....	3
Эксперимент 1: траектория градиентного спуска на квадратичной функции .....	4
Эксперимент 2: зависимость числа итераций градиентного спуска от числа обусловленности и размерности пространства .....	10
Эксперимент 3: сравнение методов градиентного спуска и Ньютона на реальной задаче логистической регрессии .....	12
Эксперимент 4: стратегия выбора длины шага в градиентном спуске .....	15
Эксперимент 5: стратегия выбора длины шага в методе Ньютона.....	21

### **Формулы для градиента и гессиана**

$$f(X) = \frac{1}{m} * \ln (I_m + e^{-b \odot A^T X}) + \frac{\lambda}{2} X^T X$$

$$\begin{aligned} \nabla f(X) &= \frac{1}{m} * \frac{I_m}{I_m + e^{-b \odot A^T X}} * e^{-b \odot A^T X} * (-b \odot A^T) + \lambda X \\ &= \frac{1}{m} * \frac{I_m}{I_m + e^{b \odot A^T X}} * (-b \odot A^T) + \lambda X \end{aligned}$$

$$\nabla^2 f(X) = \frac{1}{m} * (-b \odot A^T)^2 \frac{e^{b \odot A^T X}}{(I_m + e^{b \odot A^T X})^2} + \lambda$$

## Эксперимент 1: траектория градиентного спуска на квадратичной функции

Целью эксперимента является анализ траектории градиентного спуска на квадратичной функции. Для анализа были взяты две квадратичные функции вида  $f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle$ . Первая функция имеет параметры  $A = \begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ , вторая -  $A = \begin{bmatrix} 25 & 2 \\ 2 & 1 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ . Для этих функций был запущен алгоритм градиентного спуска с разными настройками и начальными точками. Результаты эксперимента представлены на рисунках ниже.

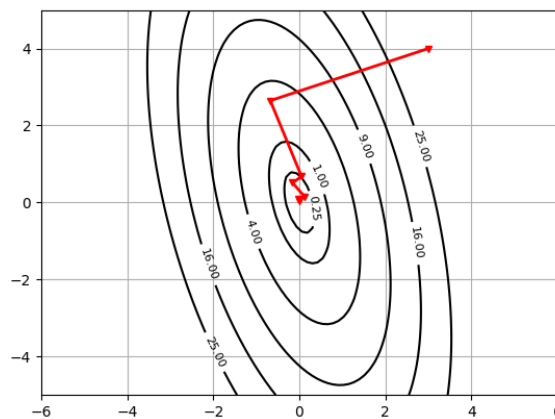


Рисунок 1.1.1 Линии уровня функции №1 с траекторией метода градиентного спуска со стратегией выбора шага Вульф и начальной точкой (3, 4).

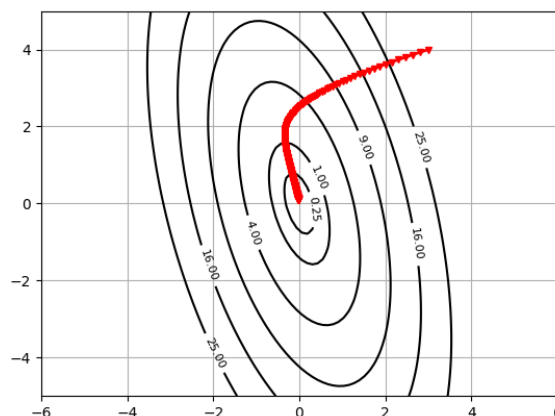


Рисунок 1.1.2 Линии уровня функции №1 с траекторией метода градиентного спуска с константной стратегией выбора шага и начальной точкой (3, 4).

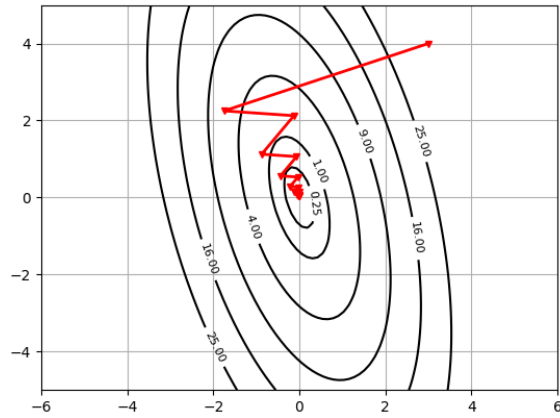


Рисунок 1.1.3 Линии уровня функции №1 с траекторией метода градиентного спуска со стратегией выбора шага Армихо и начальной точкой (3, 4).

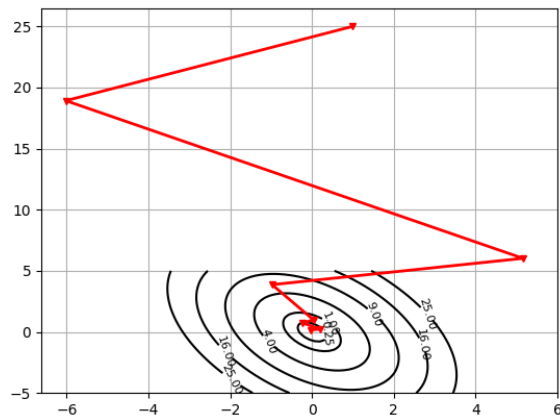


Рисунок 1.1.4 Линии уровня функции №1 с траекторией метода градиентного спуска со стратегией выбора шага Вульф и начальной точкой (1, 25).

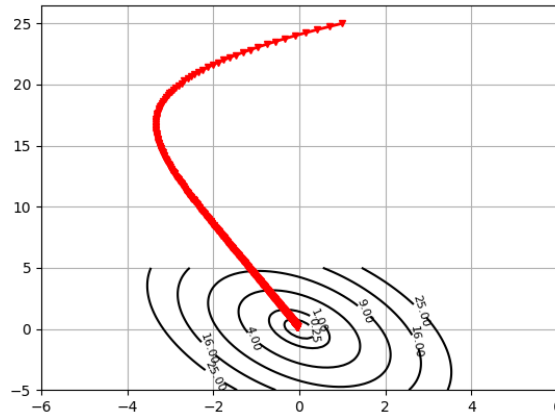


Рисунок 1.1.5 Линии уровня функции №1 с траекторией метода градиентного спуска с константной стратегией выбора шага и начальной точкой (1, 25).

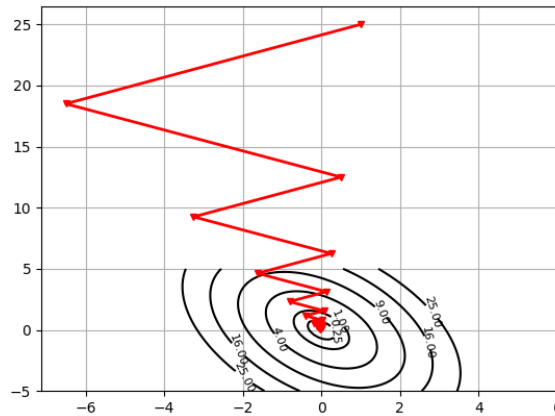


Рисунок 1.1.6 Линии уровня функции №1 с траекторией метода градиентного спуска со стратегией выбора шага Армихо и начальной точкой (1, 25).

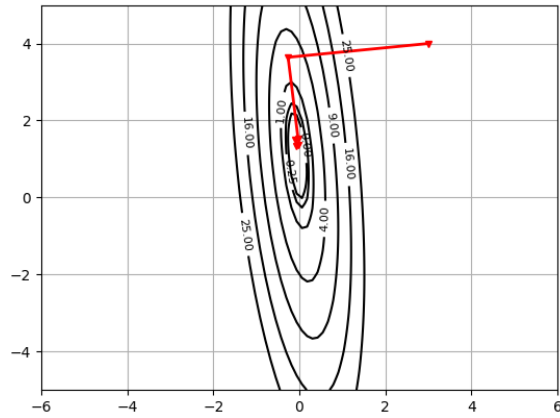


Рисунок 1.1.7 Линии уровня функции №2 с траекторией метода градиентного спуска со стратегией выбора шага Вульф и начальной точкой (3, 4).

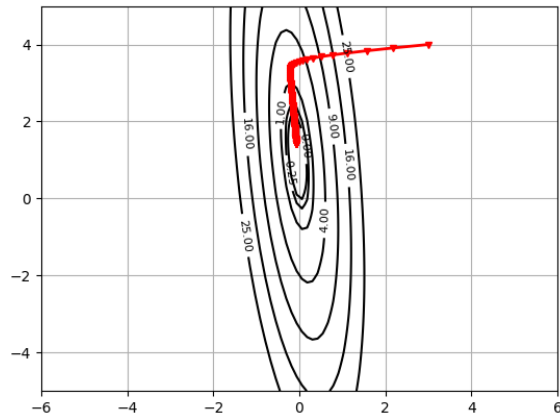


Рисунок 1.1.8 Линии уровня функции №2 с траекторией метода градиентного спуска с константной стратегией выбора шага и начальной точкой (3, 4).

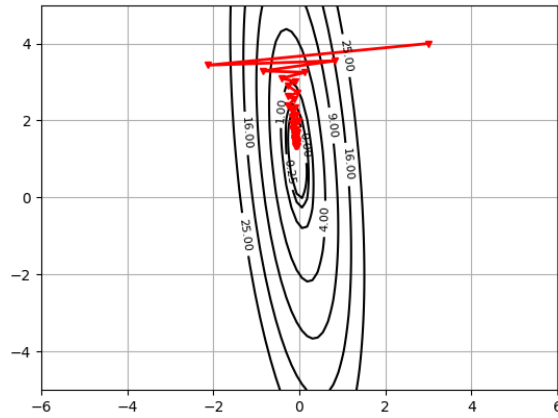


Рисунок 1.1.9 Линии уровня функции №2 с траекторией метода градиентного спуска со стратегией выбора шага Армихо и начальной точкой (3, 4).

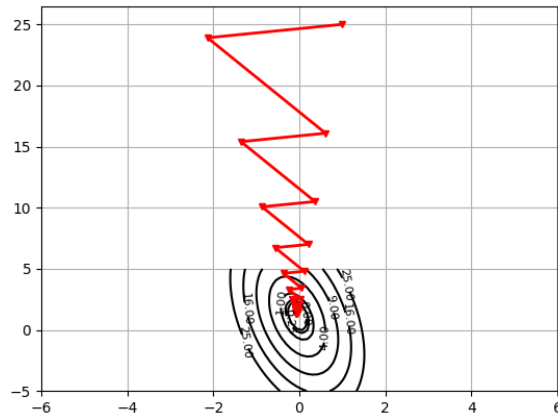


Рисунок 1.1.10 Линии уровня функции №2 с траекторией метода градиентного спуска со стратегией выбора шага Вульф и начальной точкой (1, 25).



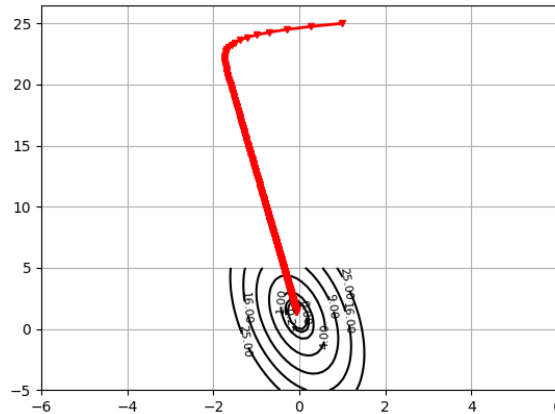


Рисунок 1.1.11 Линии уровня функции №2 с траекторией метода градиентного спуска с константной стратегией выбора шага и начальной точкой (1, 25).

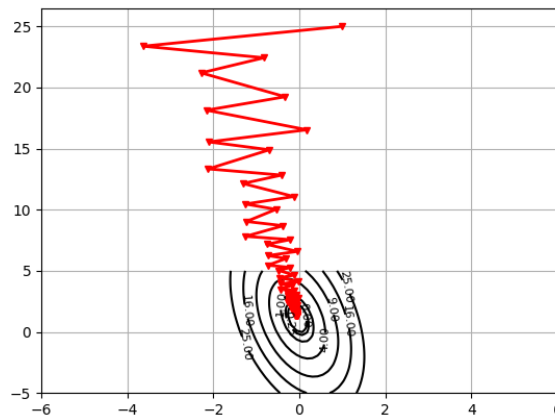


Рисунок 1.1.12 Линии уровня функции №2 с траекторией метода градиентного спуска со стратегией выбора шага Армихо и начальной точкой (1, 25).

Таким образом, поведение метода меняется в зависимости от числа обусловленности функции: чем оно больше, тем больше сходимость зависит от входных данных (начальной точки и гиперпараметров метода). Также поведение метода зависит от стратегии выбора шага: при константной стратегии траектория метода плавная, а метод Вульфа и метод Армихо имеют угловатые траектории. Выбор начальной точки влияет на скорость сходимости: чем дальше точка от искомого оптимума, тем большее количество итераций потребуется для его поиска.

Хочется отметить, что несмотря на отличия, во всех проведенных экспериментах был найден искомый оптимум.

## Эксперимент 2: зависимость числа итераций градиентного спуска от числа обусловленности и размерности пространства

Целью данного эксперимента являлось исследование зависимости числа итераций, необходимого градиентному спуску для сходимости, от числа обусловленности оптимизируемой функции и размерности пространства оптимизируемых переменных.

Для проведения анализа были сгенерированы квадратичные задачи размера  $n$  с числом обусловленности  $k$  из равномерного распределения  $U(1, 10)$ , на них был запущен метод градиентного спуска и было посчитано количество итераций, требуемое для сходимости. Далее был построен график зависимости числа итераций  $T(k, n)$  от  $k$  для фиксированного  $n$ . Поскольку каждый раз квадратичная задача генерируется случайным образом, то данный эксперимент был проведен 10 раз.

В итоге был построен график зависимости числа итераций метода градиентного спуска от числа обусловленности оптимизируемой функции. На графике цвет отвечает за размерность пространства. Так как было проведено по 10 экспериментов для каждого фиксированного  $n$ , то на графике присутствует по 10 линий каждого цвета. Также для простоты анализа был построен график среднего числа итераций.

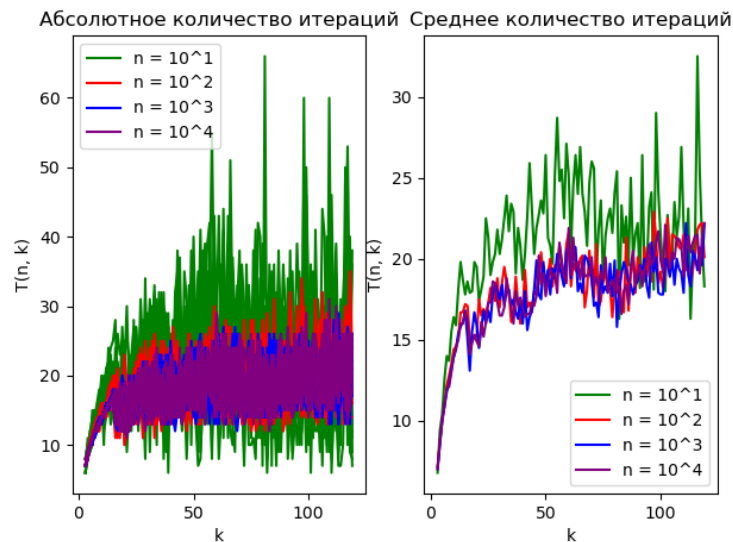


Рисунок 1.2.1 Абсолютная и средняя зависимость количества итераций метода градиентного спуска от числа обусловленности функции.

Из левого графика видно, что разброс количества итераций, необходимых для сходимости находится в обратно пропорциональной зависимости от размерности

пространства: чем больше  $n$ , тем разброс меньше. Также можно сделать вывод, что при любой размерности пространства оптимизируемых параметров удельное количество итераций снижается с ростом числа обусловленности функции: до некоторого момента  $T(k, n) \ll T(k + 1, n)$ , тем не менее через какое-то время этот перепад снижается и график становится более плавным.

Из правого графика можно сделать вывод, что в среднем при  $n = 10$  нужно большее количество итераций для сходимости при любом  $k$ , чем при большей размерности пространства оптимизируемых параметров.

### **Эксперимент 3: сравнение методов градиентного спуска и Ньютона на реальной задаче логистической регрессии**

Целью данного эксперимента являлось сравнение методов градиентного спуска и Ньютона на задаче обучения логистической регрессии на реальных данных. В качестве реальных данных были использованы следующие три набора с сайта LIBSVM3: w8a, gisette и real-sim.

Для каждого набора данных были построены 2 логистические регрессии: первая была обучена с помощью метода градиентного спуска, вторая – с помощью метода Ньютона. Параметры обоих методов были равными параметрам по умолчанию, начальная точка –  $x_0 = 0$ , коэффициент регуляризации  $\lambda = \frac{1}{m}$ .

Далее были построены график зависимости значения функции от реального времени работы метода, а также график зависимости относительного квадрата нормы градиента против реального времени работы.

Полученные результаты представлены на рисунках ниже. На каждом рисунке изображено два графика: слева - график зависимости значения функции от реального времени работы метода, справа - график зависимости относительного квадрата нормы градиента против реального времени работы.

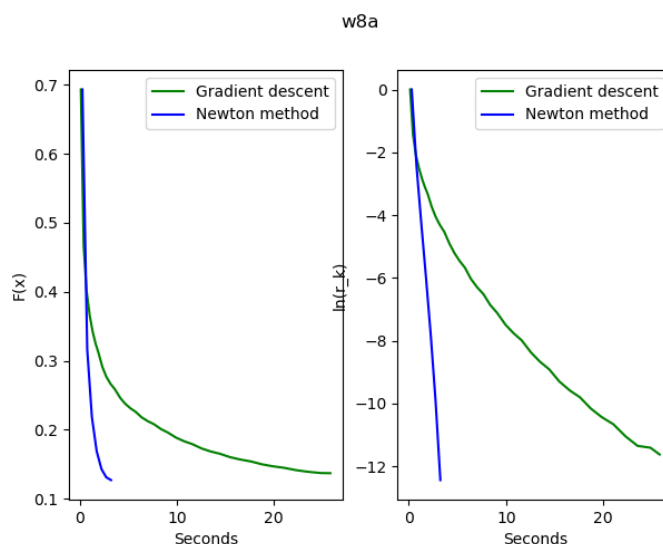


Рисунок 1.3.1 График сходимости двух видов для w8a

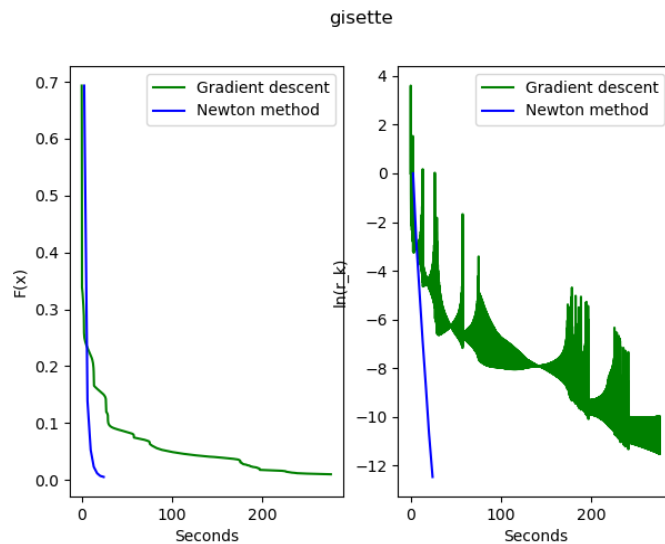


Рисунок 1.3.2 График сходимости двух видов для gisette

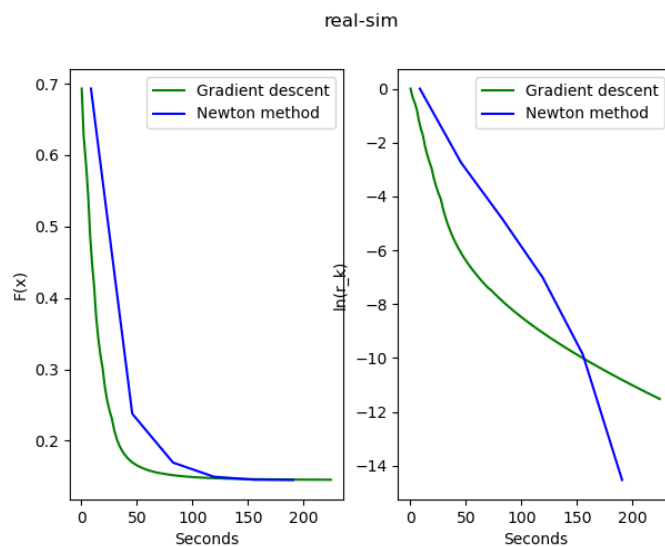


Рисунок 1.3.3 График сходимости двух видов для real-sim

Как видно из графиков для двух из трех наборов данных метод Ньютона сходил быстрее, чем метод градиентного спуска. При чем разница колоссальна: первому нужно было в несколько раз меньше времени для достижения сходимости. На данных, на которых градиентный спуск показал себя лучше, нет такого значительного перепада: градиентный спуск быстрее, но не в несколько раз.

Оценка времени работы: линейный поиск Вульфа осуществляется за линейное время, один проход цикла также осуществляется за линейное время, итоговая асимптотика:  $\mathcal{O}(max - iter * n)$ . Эта оценка одинакова для обоих методов.

Оценка памяти: память зависит от размера хранимой истории, а не от количества производимых операций. Максимально хранить мы будем  $nt$ , тогда оценка:  $\mathcal{O}(nt)$ . Эта оценка одинакова для обоих методов.

## **Эксперимент 4: стратегия выбора длины шага в градиентном спуске**

Цель эксперимента заключалась в том, чтобы исследовать, как зависит поведение градиентного спуска от стратегии подбора шага: константный шаг, бэктрекинг, условия Вульфа.

Из равномерного распределения была генерирована случайная квадратичная функция размерностью 20 на 20 с числом обусловленности равным 50. Для этой функции была построена логистическая регрессия с помощью метода градиентного спуска с использованием разных стратегий выбора длины шага, разных настроек стратегий, а также разных начальных точек.

По итогам экспериментов были нарисованы графики функции против числа итераций - для квадратичной функции, а также графики относительных квадратов нормы градиента в логарифмической шкале против числа итераций - для логистической регрессии.

Первым делом была исследована зависимость поведения метода от гиперпараметра стратегий подбора шага. На сгенерированной функции были построены несколько логистических регрессий, каждая из которых была обучена с помощью разных стратегий подбора шага с различными коэффициентами  $c$ . Начальная точка здесь нулевая.

Результаты представлены на рисунках ниже. На каждом рисунке изображено два графика: слева - график функции против числа итераций, справа - график относительных квадратов нормы градиента в логарифмической шкале против числа итераций.

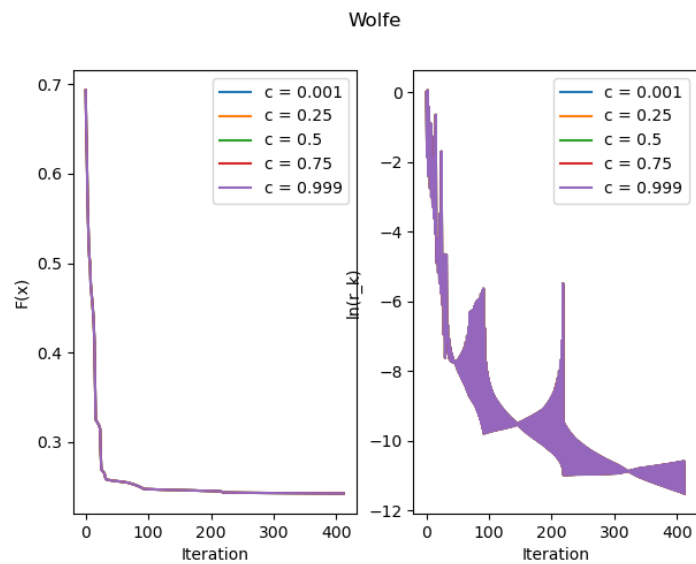


Рисунок 1.4.1 Кривые сходимости градиентного спуска со стратегией выбора шага Вульф с различными параметрами.

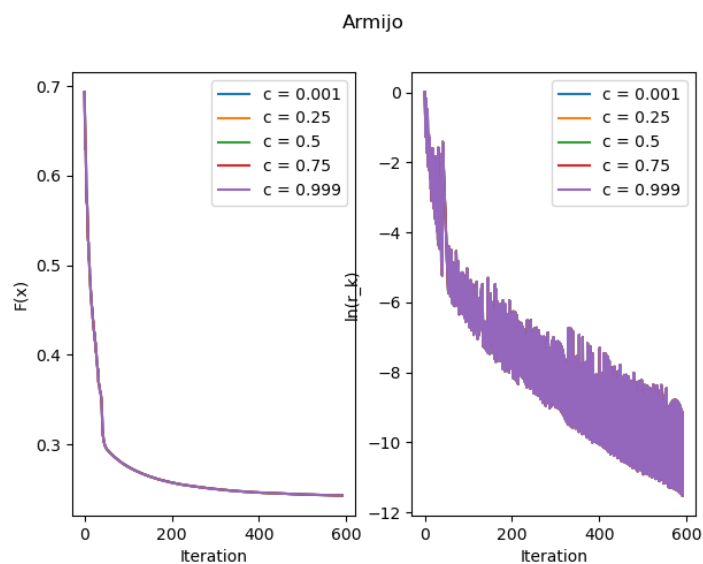


Рисунок 1.4.2 Кривые сходимости градиентного спуска со стратегией выбора шага Армихо с различными параметрами.



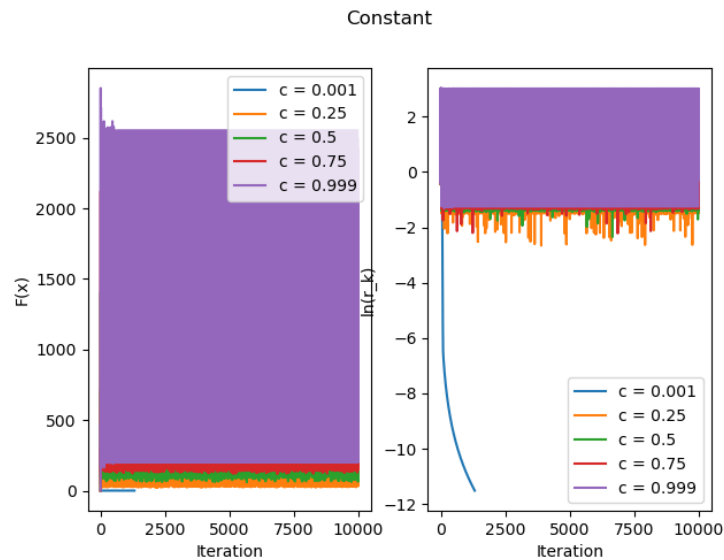


Рисунок 1.4.3 Кривые сходимости градиентного спуска с константной стратегией выбора шага с различными параметрами.

Из приведенных выше графиков можно сделать вывод, что для сгенерированной функции константа в стратегии выбора шага не оказала влияние на сходимость в методе Вульфа и Армихо. Говоря о константной стратегии выбора шага, в ней параметр  $c$  имеет существенное влияние: при всех  $c$ , кроме 0.001 градиентный спуск расходится.

Исходя из этого, при следующих экспериментах для всех стратегий выбора шага будет использована константа  $c = 0.001$ .

Далее была исследована сходимость метода градиентного спуска в зависимости от начальной точки. На сгенерированной функции были построены несколько логистических регрессий, каждая из которых была обучена с помощью разных стратегий подбора шага с различными начальными точками и коэффициентом  $c = 0.001$ . Начальные точки генерировались из семейства нормального распределения.

Результаты эксперимента представлены ниже. На рисунке цвет отвечает за различные начальные точки. Легенда отсутствует, так как  $x_0 \in \mathbb{R}^{20}$ .

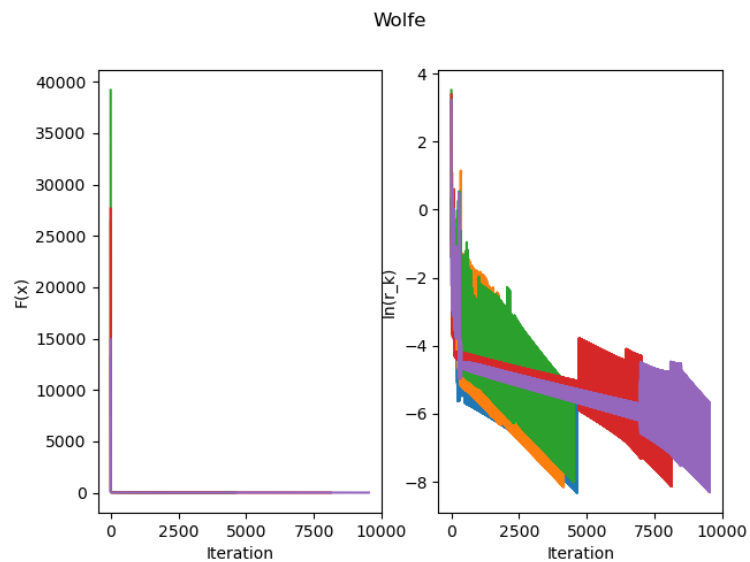


Рисунок 1.4.4 Кривые сходимости градиентного спуска со стратегией выбора шага Вульф с различными начальными точками.

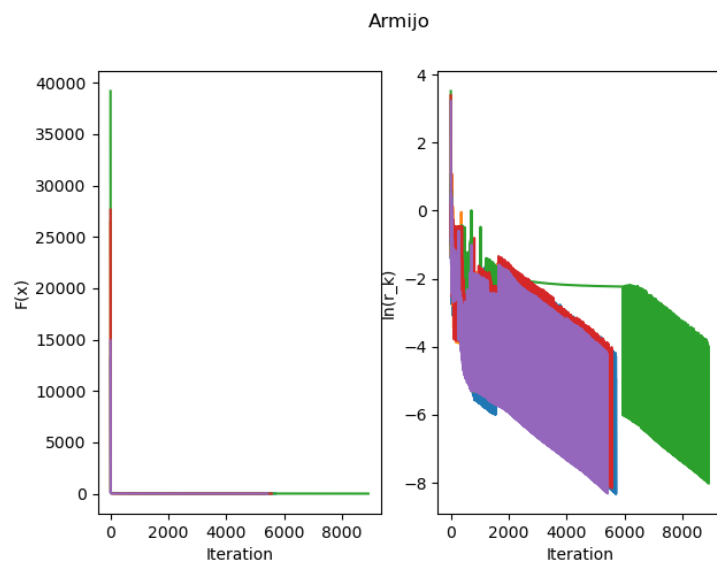


Рисунок 1.4.5 Кривые сходимости градиентного спуска со стратегией выбора шага Армихо с различными начальными точками.

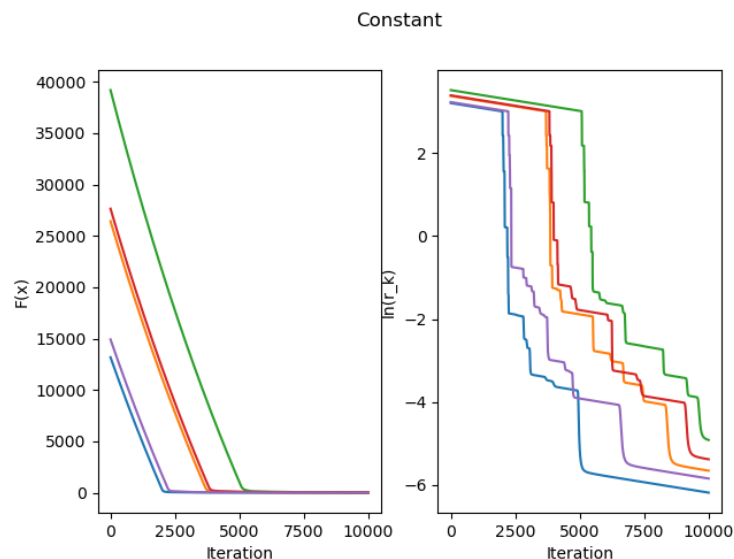


Рисунок 1.4.6 Кривые сходимости градиентного спуска с константной стратегией выбора шага с различными начальными точками.

Из приведенных выше графиков можно сделать вывод, что при различных начальных точках стратегии выбора шага Вульфа и Армихо ведут себя похожим образом: функция быстро спускается к минимуму, а потом колеблется вокруг него. Константная стратегия выбора шага более чувствительна к выбору начальной точки: чем она ближе к искомому оптимуму, тем меньше количество итераций требуется. Также видно, что для константной стратегии выбора шага относительный квадрат нормы градиента убывает достаточно плавно.

Результат сравнения трех стратегий выбора шага с константой  $c = 0.001$  и фиксированной нулевой начальной точкой представлен на рисунке ниже.

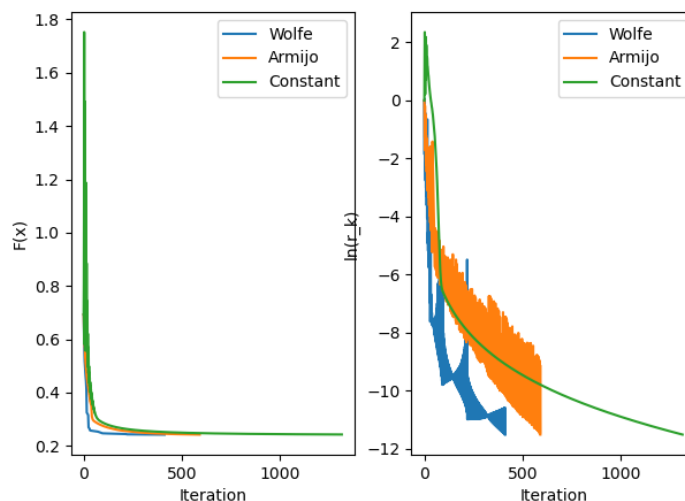


Рисунок 1.4.7 Кривые сходимости градиентного спуска с различными стратегиями выбора шага.

В данном случае меньше всего итераций для достижения оптимума понадобилось градиентному спуску, использующему стратегию Вульфа, больше всего – константной стратегии. Тем не менее, все стратегии очень быстро спустились к нулю, а потом колебались вокруг него.

Относительный квадрат нормы градиента для константного метода стабилен и плавно убывает, что нельзя сказать про остальные стратегии, для которых он ведет себя неоднозначно.

## **Эксперимент 5: стратегия выбора длины шага в методе Ньютона**

Цель эксперимента заключалась в том, чтобы исследовать, как зависит поведение метода Ньютона от стратегии подбора шага: константный шаг, бэктрекинг, условия Вульфа.

Из равномерного распределения была генерирована случайная квадратичная функция размерностью 20 на 20 с числом обусловленности равным 50. Для этой функции была построена логистическая регрессия с помощью метода градиентного спуска с использованием разных стратегий выбора длины шага, разных настроек стратегий, а также разных начальных точек.

По итогам экспериментов были нарисованы графики функции против числа итераций - для квадратичной функции, а также графики относительных квадратов нормы градиента в логарифмической шкале против числа итераций - для логистической регрессии.

Первым делом была исследована зависимость поведения метода от гиперпараметра стратегий подбора шага. На сгенерированной функции были построены несколько логистических регрессий, каждая из которых была обучена с помощью разных стратегий подбора шага с различными коэффициентами  $c$ . Начальная точка здесь нулевая.

Результаты представлены на рисунках ниже. На каждом рисунке изображено два графика: слева - график функции против числа итераций, справа - график относительных квадратов нормы градиента в логарифмической шкале против числа итераций.

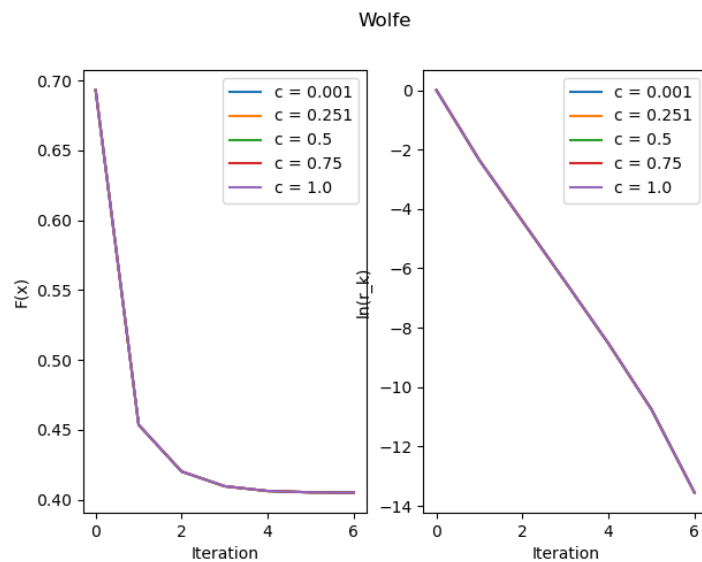


Рисунок 1.4.1 Кривые сходимости метода Ньютона со стратегией выбора шага Вульф с различными параметрами.

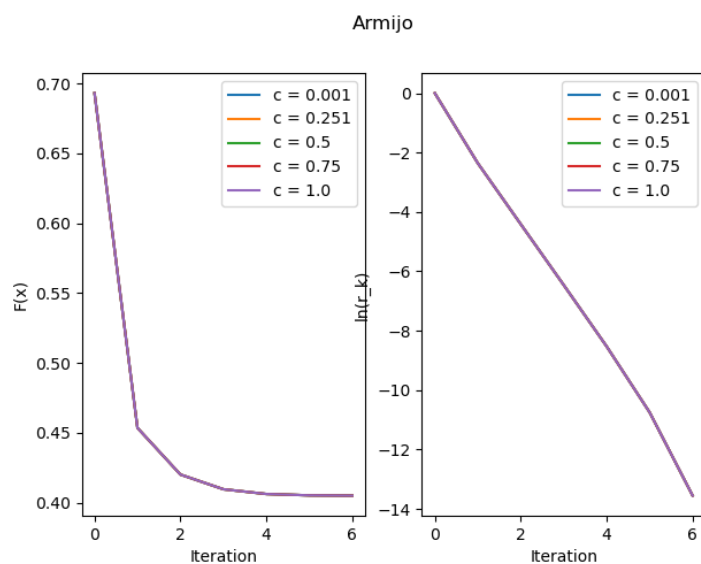


Рисунок 1.4.2 Кривые сходимости метода Ньютона со стратегией выбора шага Армихо с различными параметрами.

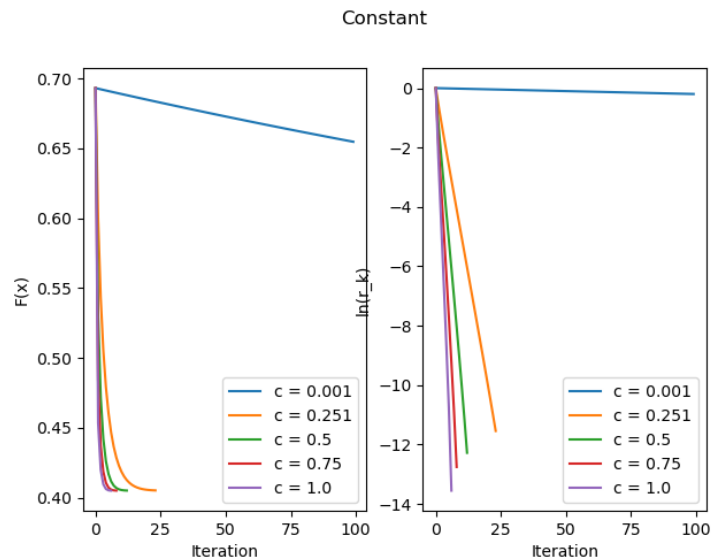


Рисунок 1.4.3 Кривые сходимости метода Ньютона с константной стратегией выбора шага с различными параметрами.

Из приведенных выше графиков можно сделать вывод, что для сгенерированной функции константа в стратегии выбора шага оказала влияние на сходимость метода Ньютона только в константной стратегии. При этом самая быстрая сходимость была достигнута при  $c = 1$ . Это подтверждает теоретические сведения о методе Ньютона. Поэтому при следующих экспериментах для всех стратегий выбора шага будет использована константа  $c = 1$ .

Далее была исследована сходимость метода Ньютона в зависимости от начальной точки. На сгенерированной функции были построены несколько логистических регрессий, каждая из которых была обучена с помощью разных стратегий подбора шага с различными начальными точками и коэффициентом  $c = 1$ . Начальные точки генерировались из семейства нормального распределения.

Результаты эксперимента представлены ниже. На рисунке цвет отвечает за различные начальные точки. Легенда отсутствует, так как  $x_0 \in \mathbb{R}^{20}$ .

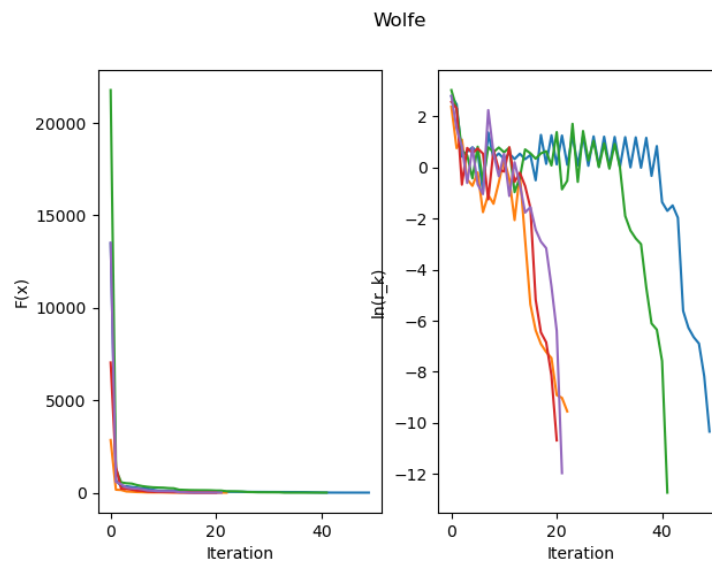


Рисунок 1.4.4 Кривые сходимости метода Ньютона со стратегией выбора шага Вульф с различными начальными точками.

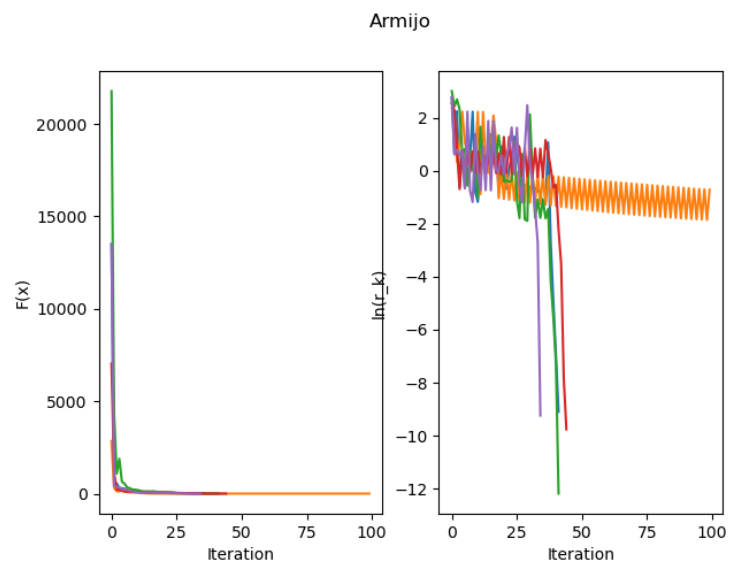


Рисунок 1.4.5 Кривые сходимости метода Ньютона со стратегией выбора шага Армихо с различными начальными точками.



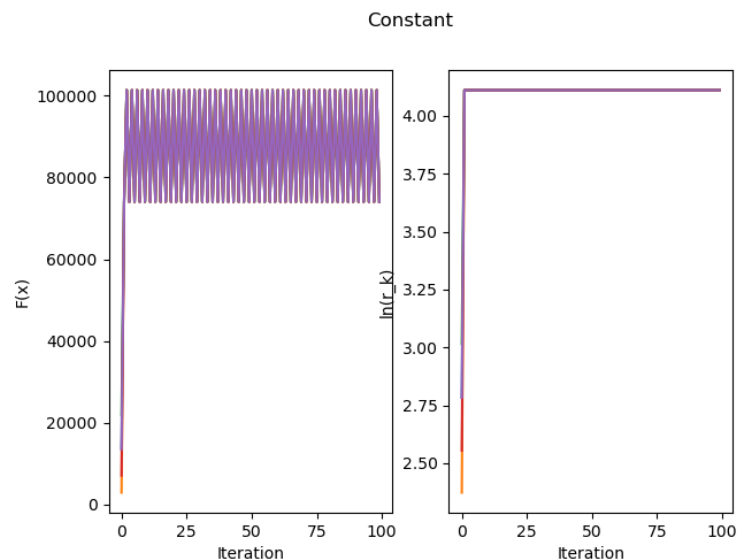


Рисунок 1.4.6 Кривые сходимости метода Ньютона с константной стратегией выбора шага с различными начальными точками.

Из приведенных выше графиков можно сделать вывод, что при различных начальных точках стратегии выбора шага Вульфа и Армихо ведут себя похожим образом: функция быстро спускается к минимуму, а потом колеблется вокруг него. При константной стратегии выбора шага метод Ньютона расходится.

Результат сравнения трех стратегий выбора шага с константой  $c = 1$  и фиксированной нулевой начальной точкой представлен на рисунке ниже.

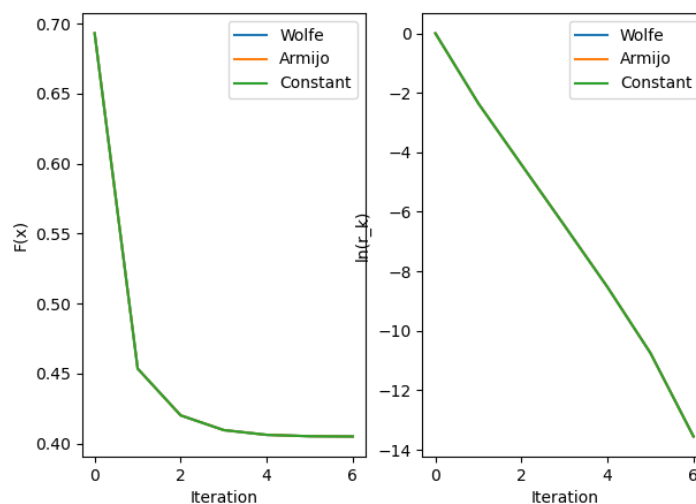


Рисунок 1.4.7 Кривые сходимости метода Ньютона с различными стратегиями выбора шага.

Несмотря на неоднозначное поведение константной стратегии выбора шага в предыдущем эксперименте, в данном случае все три стратегии показали абсолютно одинаковые результаты. Это может объясняться тем, что константная стратегия в методе Ньютона чувствительна к выбору начальной точки, и в данном случае нулевая точка – удачный выбор.

Таким образом, для метода Ньютона стратегии выбора шага Вульфа и Армихо показали очень похожие результаты, которые сильно лучше результатов константной стратегии.