Федеральное государственное бюджетное образовательное
учреждение высшего профессионального образования
«Санкт-Петербургский национальный исследовательский университет
информационных технологий, механики и оптики»

Отчёт по лабораторной работе №1
Методы градиентного спуска и метод Ньютона.
По дисциплине: Методы оптимизации

Выполнил:
Зозуля А.В.
Проверила:
Головкина А.Г.

Москва 2024 г.

Формулы для градиента и гессиана функции логистической регрессии

$$f(x) = \frac{1}{m}\sum_{i=1}^{m} ln(1 + exp(- b_i\langle a_i, x\rangle)) + \frac{\lambda}{2}||x||_2^2$$

$$\nabla f(x) = - \frac{1}{m}\sum_{i=1}^{m} \frac{exp(-b_i\langle a_i, x\rangle)}{1 + exp(-b_i\langle a_i, x\rangle)} b_i a_i + \lambda x =>$$

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

$$\nabla^2 f(x) = \frac{1}{m}A^T \frac{1}{1+exp(-bAx)} (1 - \frac{1}{1+exp(-bAx)})A + \lambda$$

Эксперименты

1. Траектория градиентного спуска на квадратичной функции.

Для анализа траектории градиентного спуска использовались две функции, одна с числом обусловленности равным 1.824 другая – 16.224, для каждой функции и каждого способа поиска шага использовался градиентный спуск из 3 разных точек.

Обусловленность функций 1.824:

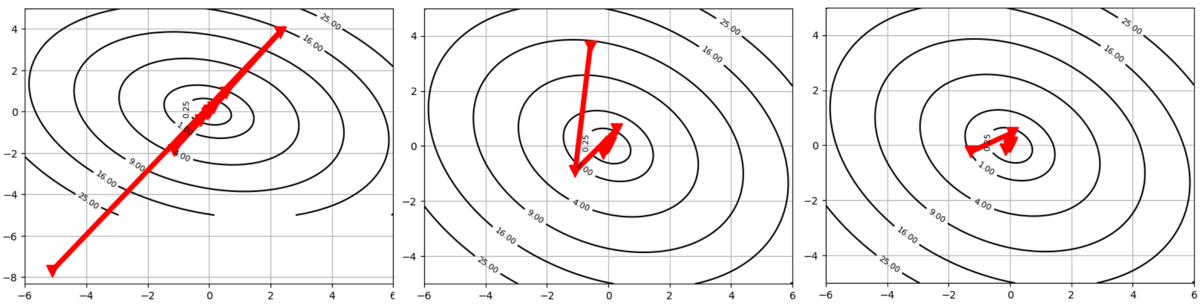

Рис. 1. Метод Wolfe шагов 10, 12, 10 соответственно

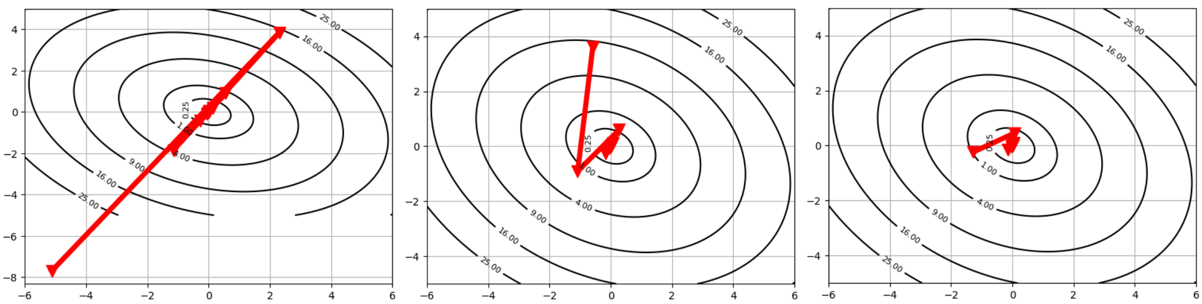

Рис. 2. Метод Armijo шагов 10, 12, 10 соответственно

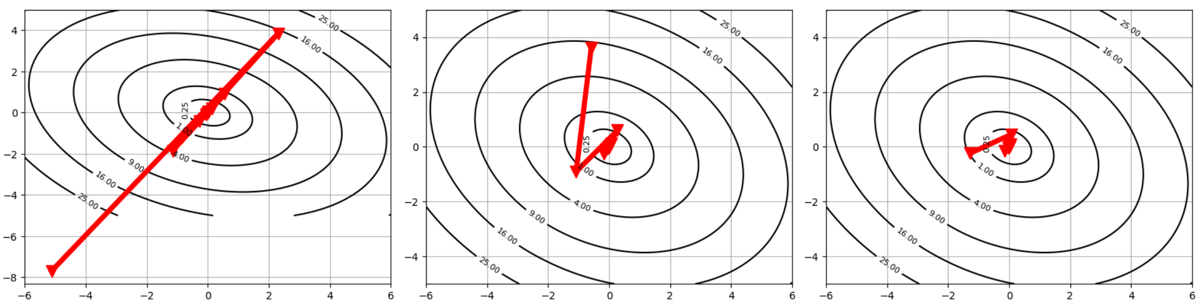

Рис. 3. Метод Constant шагов 10, 12, 10 соответственно

Обусловленность функций 16.224:

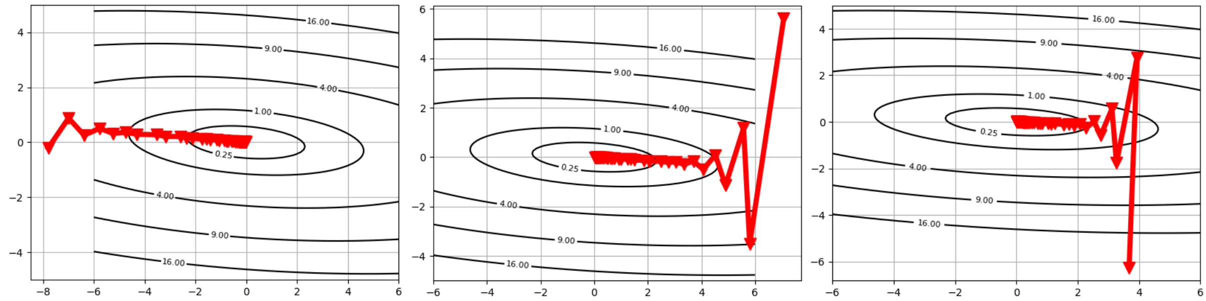

Рис. 4. Метод Wolfe шагов 44, 40, 48 соответственно

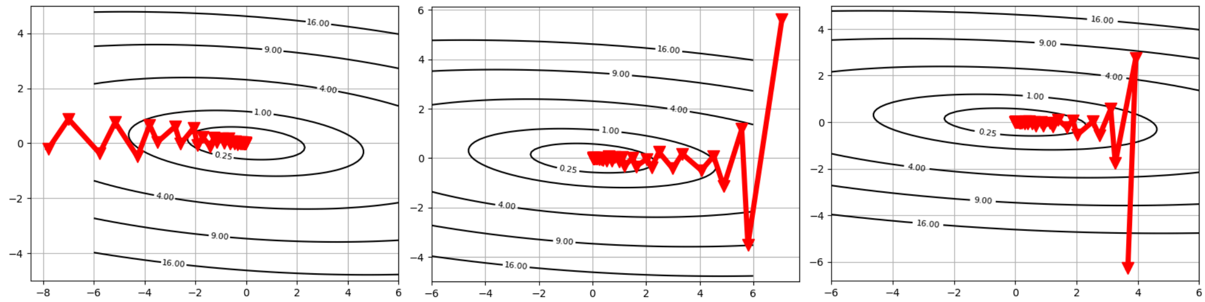

Рис. 5. Метод Armijo шагов 45, 44, 50 соответственно

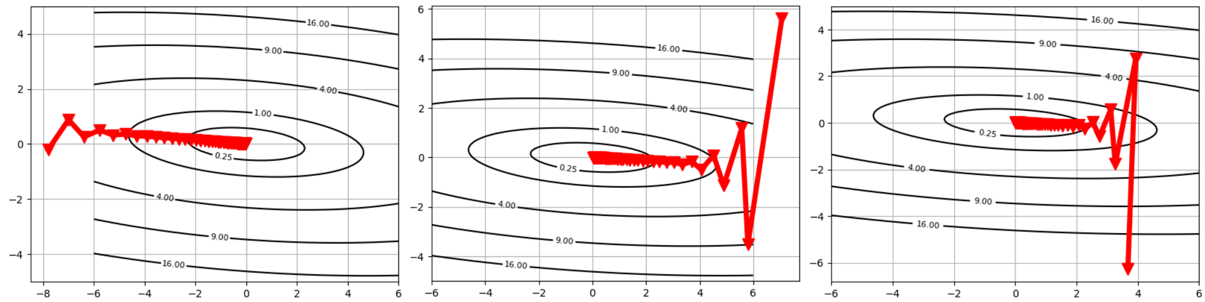

Рис. 6. Метод Constant шагов 59, 53, 65 соответственно

При росте обусловленности функции увеличивается количество итераций градиентного метода – усиливается блуждание вдоль большей оси.

На небольшой обусловленности все способы поиска шага показали одинаковые результаты, однако для большей обусловленности метод поиска шага по сильным условиям Вольфе показал себя лучше всего, затем Армихо, затем константный.

2. Зависимость числа итераций градиентного спуска от числа обусловленности и размерности пространства

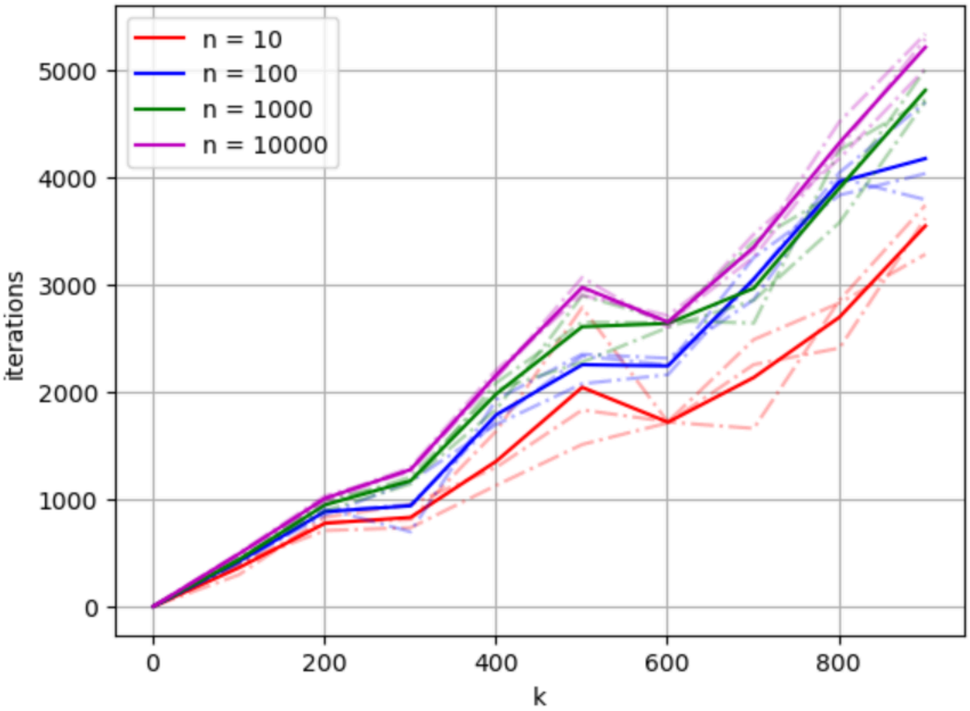

Рис. 7. Число итераций, необходимое градиентному спуску для сходимости, от числа обусловленности и размерности пространства n. Жирными линиями среднее по замерам для данной размерности.

Почти линейная зависимость числа итераций от обусловленности функции, и слабое влияние размерности задачи на количество итераций, следовательно желательно преобразовать матрицу к меньшей обусловленности.

3. Сравнение методов градиентного спуска и Ньютона на реальной задаче логистической регрессии

Градиентный метод обладает линейной скоростью сходимости, а метод Ньютона – квадратичной в области оптимального решения, а для глобального линейной

Градиентный спуск требует $O(n)$ стоимость по памяти и $O(q) + O(n)$ по времени на одну итерацию, где $O(q)$ – стоимость вычисления функции, градиента и гессиана, а $O(n)$ стоимость от размерности задачи. Метод Ньютона требует $O(n^2)$ памяти и $O(qn^2) + O(n^3)$ времени на одну итерацию где первое слагаемое отвечает за решение $\nabla^2 f(xk)dk =- \nabla f(xk)$, а второе за разложение Холецкого: A = L*L^T.

Для поиска шага – α в обоих методах использовались условия Вольфе, так как показали наилучший результат в экспериментах.

Для метода Ньютона бралась ε = 10^-9 а для градиентного – 10^-5

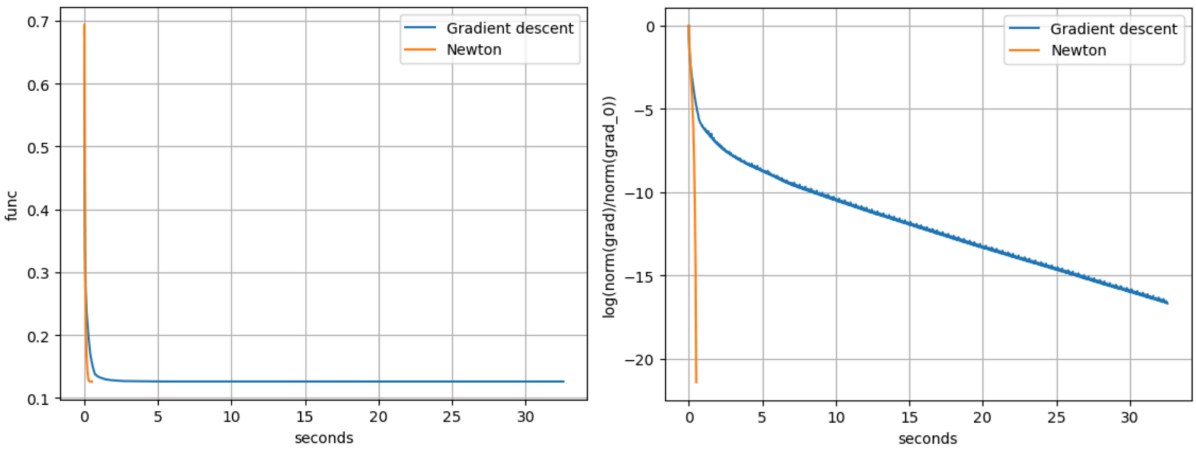

Рис. 8. Датасет w8a, сильное условие Вульфе, затрачено времени на вычисления: 33.5 сек.
Слева представлено значение целевой функции, а справа относительный квадрат нормы градиента в лог шкале.

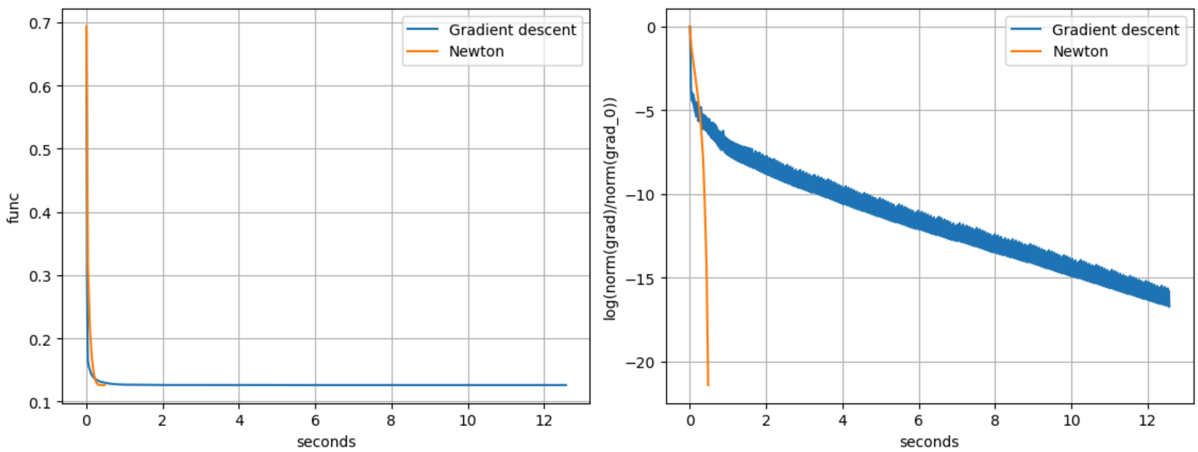

Рис. 9. Датасет w8a, условие Армихо, затрачено времени на вычисления: 13.5 сек.
Слева представлено значение целевой функции, а справа относительный квадрат нормы градиента в лог шкале.

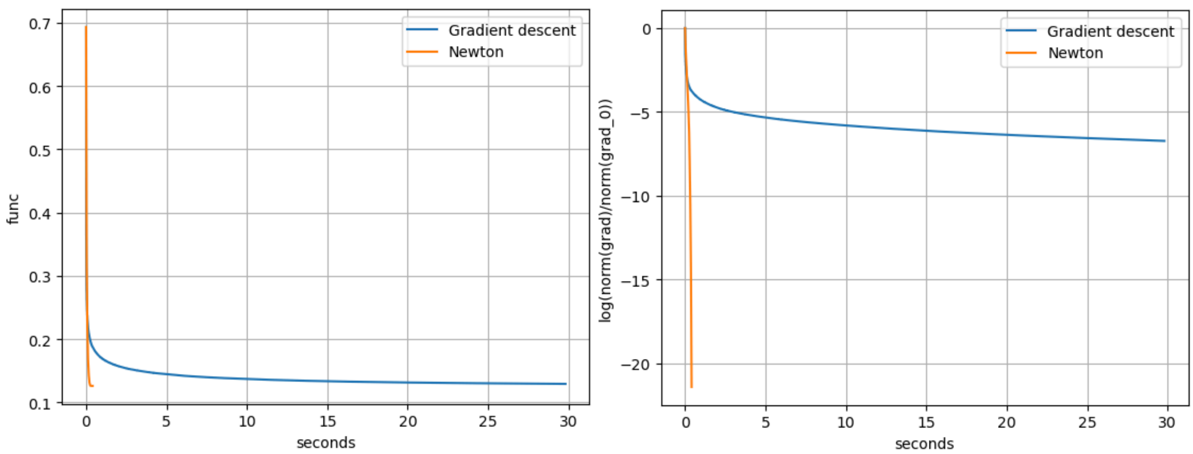

Рис. 10. Датасет w8a, константа, затрачено времени на вычисления: 30.6 сек.
Слева представлено значение целевой функции, а справа относительный квадрат нормы градиента в лог шкале.

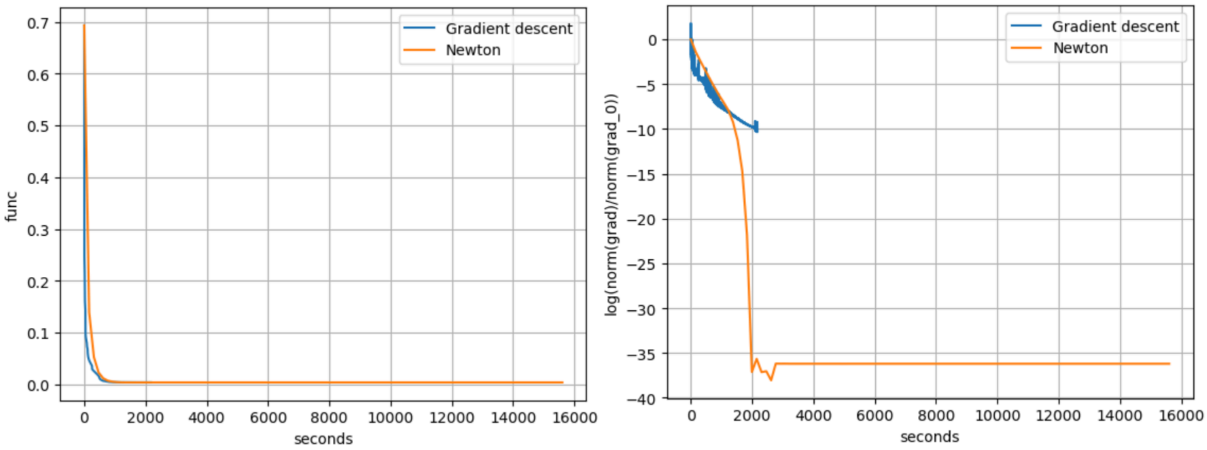

Рис. 11. Датасет gisette_scale, затрачено времени на вычисления: 4 ч 59 мин 5 сек.
Слева представлено значение целевой функции, а справа относительный квадрат нормы градиента в лог шкале

Комментарий

К сожалению слишком много времени и компьютерных ресурсов занимает вычисление оптимума.

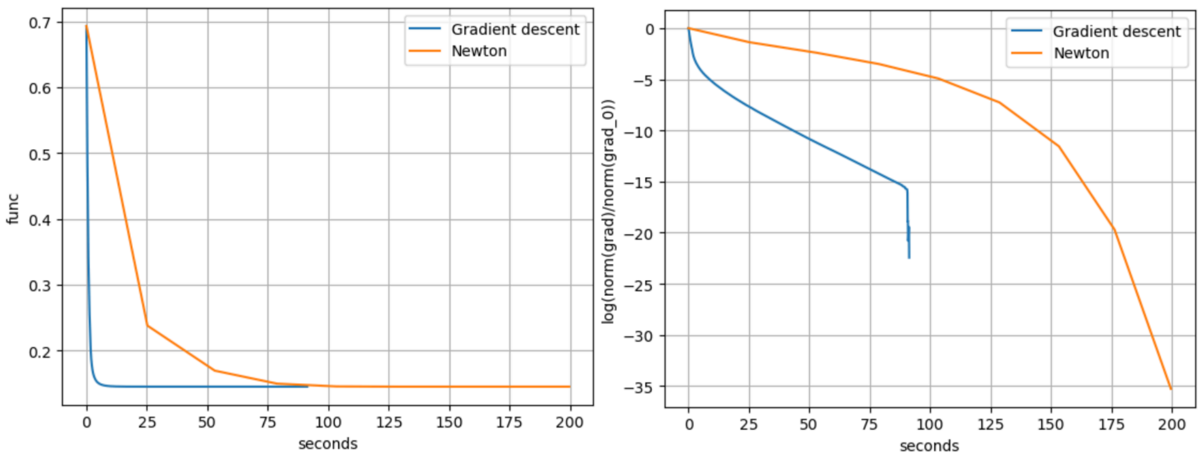

Рис. 12. Датасет real-sim, сильное условие Вульфе, затрачено времени на вычисления: 4 мин 54 сек.
Слева представлено значение целевой функции, а справа относительный квадрат нормы градиента в лог шкале

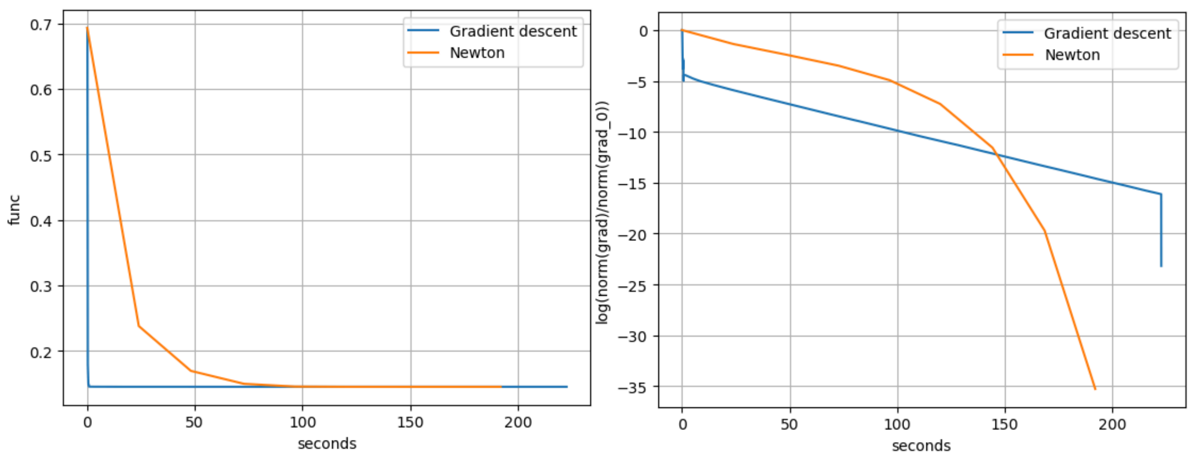

Рис. 13. Датасет real-sim, условие Армихо, затрачено времени на вычисления: 6 мин 57 сек. Слева представлено значение целевой функции, а справа относительный квадрат нормы градиента в лог шкале

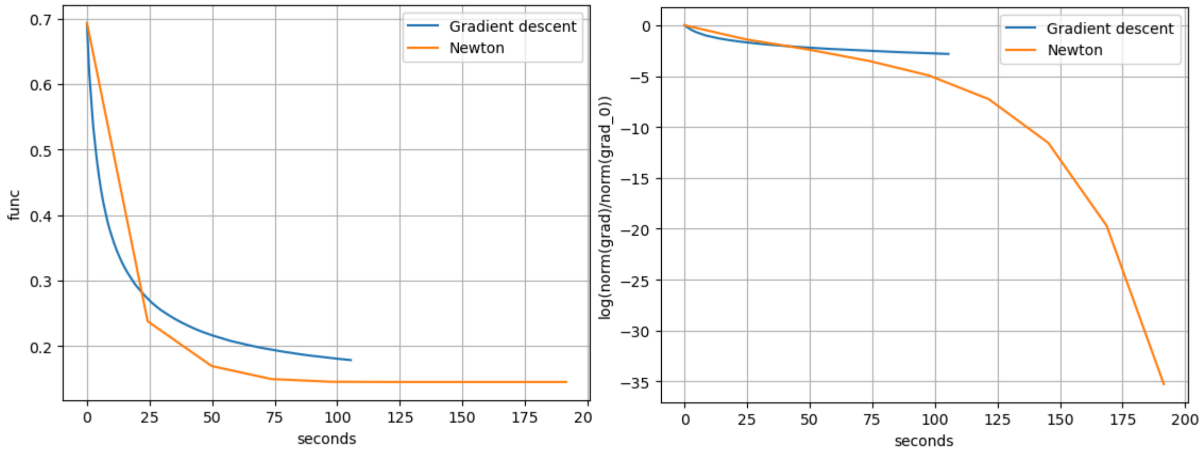

Рис. 14. Датасет real-sim, константа, затрачено времени на вычисления: 4 мин 59 сек. Слева представлено значение целевой функции, а справа относительный квадрат нормы градиента в лог шкале

Для датасета gisette_scale потребовалось в среднем 1.3 гигабайта оперативной памяти, а для real-sim 8.6 гигабайта.

На примере первого датасета видно что при небольшой размерности задачи метод Ньютона намного быстрее справляется с задачей, однако при большой размерности задачи метод ньютона начинает сильно проигрывать по скорости вычислений.

Также можно предположить что начальная точка поиска для метода Ньютона находится не в области оптимального решения, а значит линейная скорость сходимости, а учитывая что стоимость одной итерации по времени у метода Ньютона дороже чем у градиентного спуска, то можно объяснить такой результат по сравнению с градиентным.

Поэтому в задачах больших размерностей лучше использовать градиентный спуск.

В задачи малой размерности используя условие армихо алгоритм быстрее справляется с задачей, а в большой наоборот медленнее, константа примерно занимает столько же времени сколько и Вульфе.

4. Стратегия выбора длины шага в градиентном спуске.

Квадратичная задача размерностью 20x20 и числом обусловленности 500

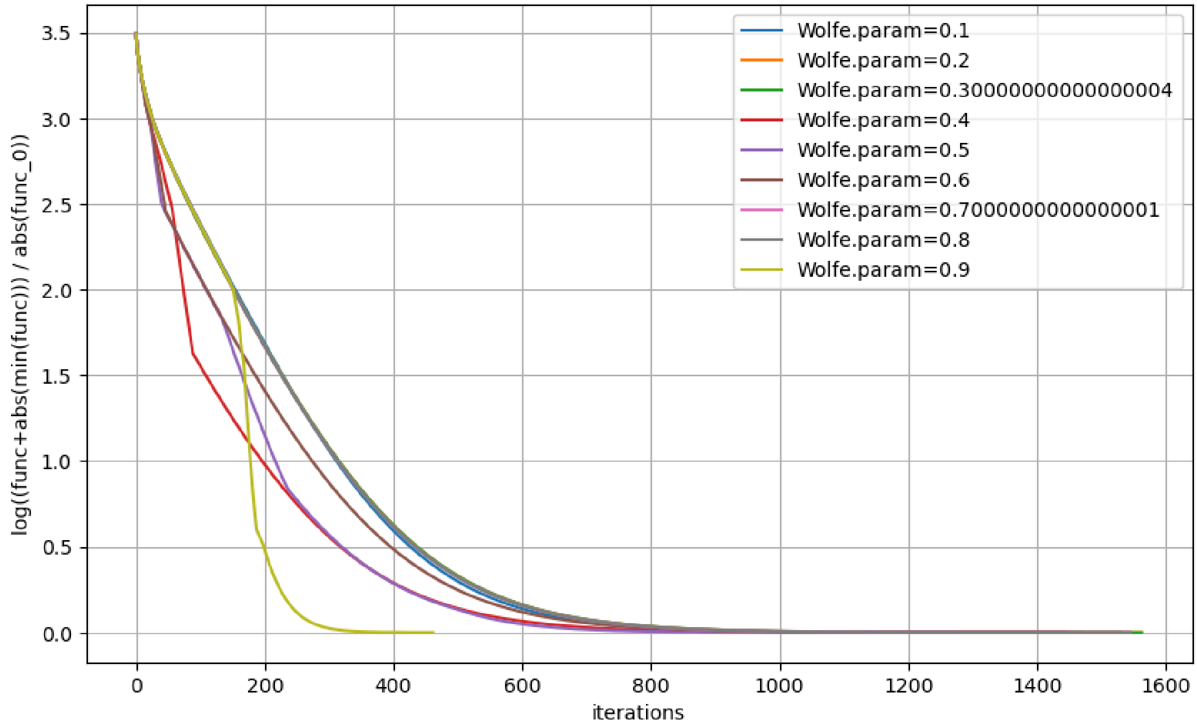

Рис. 15. Зависимость логарифма относительного значения функции от итераций, вариация параметра c2

Так как для логарифма требуется число большее нуля, то к функции прибавляется модуль минимального значения функции.

Эксперимент показал, что значение *c2* для сильного условия Вульфе для задачи с данными условиями лучше подбирать *c2* равную 0.9, обеспечивающее наибыстрейшую скорость сходимости.

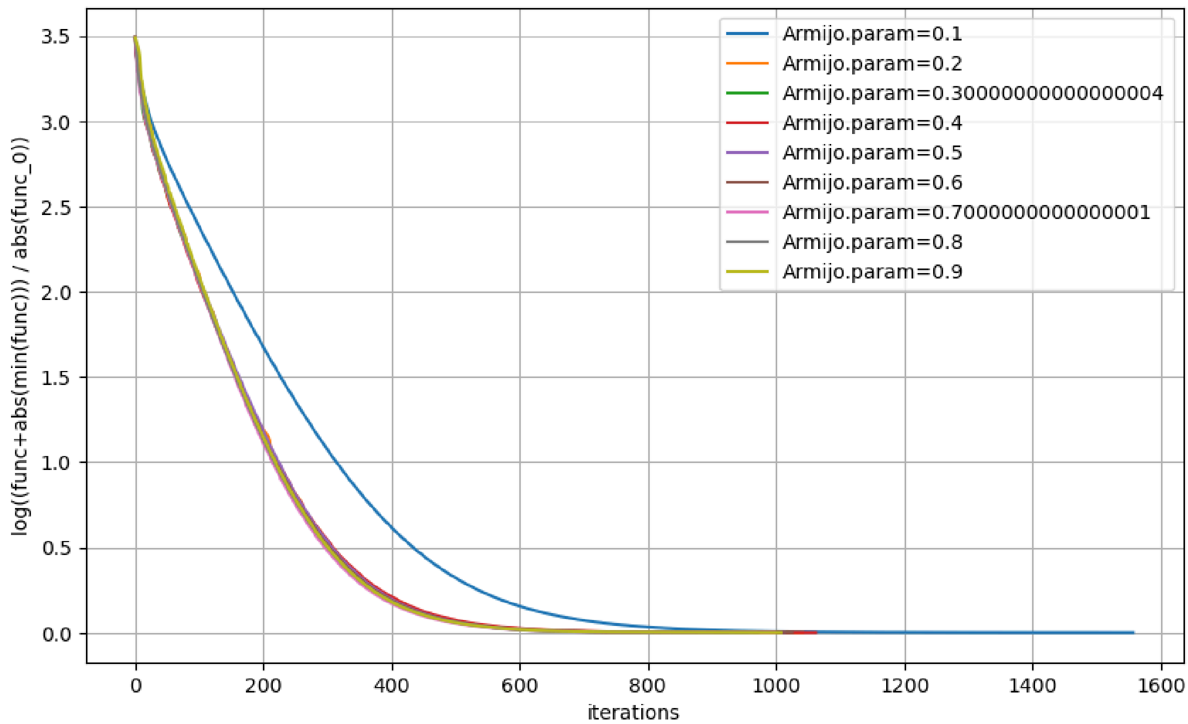

Рис. 16. Зависимость логарифма относительного значения функции от итераций, вариация параметра c1

Для условия Армихо с учетом задачи с данными условиями лучше подбирать c1 равную 0.9, обеспечивающее наибыстрейшую скорость сходимости.

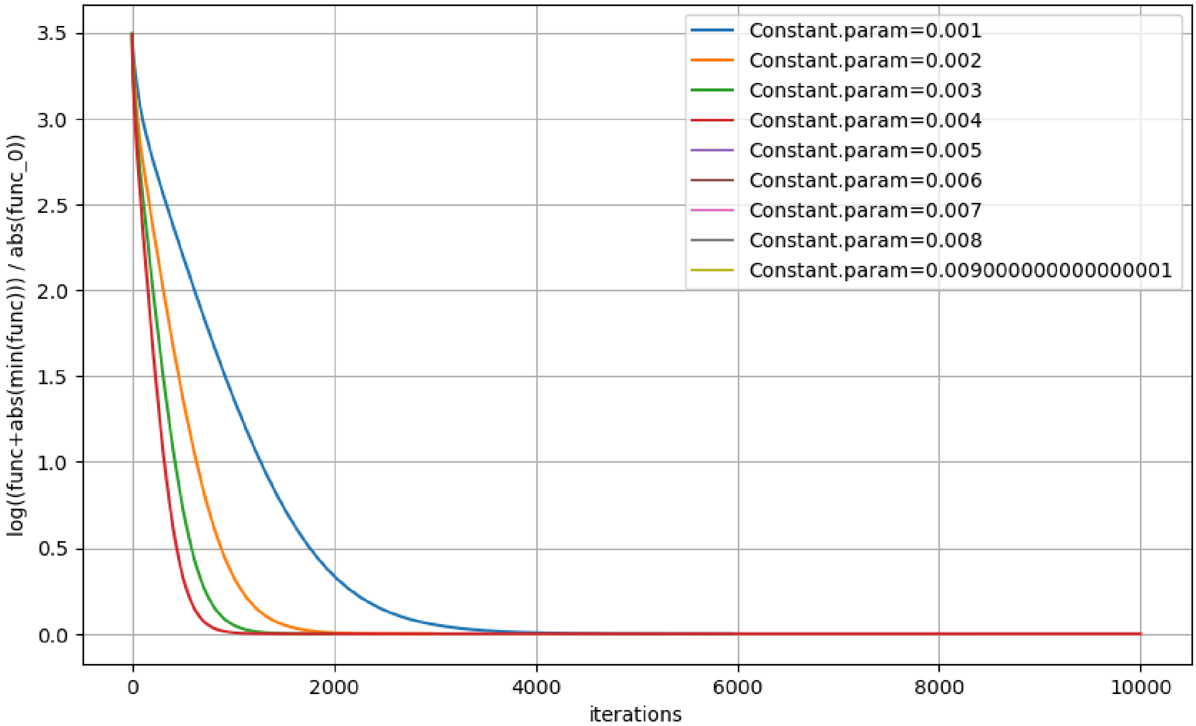

Рис. 17. Зависимость логарифма относительного значения функции от итераций, вариация параметра *c*

Для большего значения константы метод расходится:

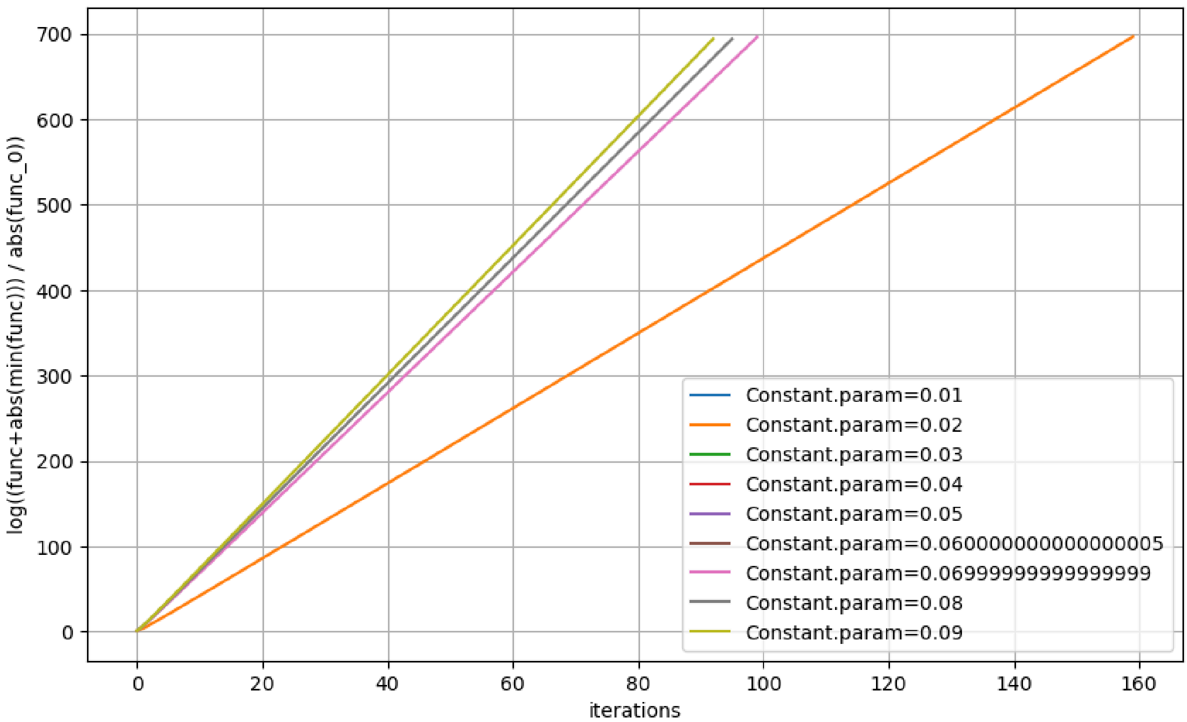

Рис. 18. Зависимость логарифма относительного значения функции от итераций, вариация параметра *c*. Расхождение метода

Для константы с учетом задачи с данными условиями лучше подбирать значение равное 9*10^-3, обеспечивающее наибыстрейшую скорость сходимости, для больших значений *c* алгоритм сходится хуже, а с некоторого значения начинается расхождение.

Задача логистической регрессии размерностью 1000x20:

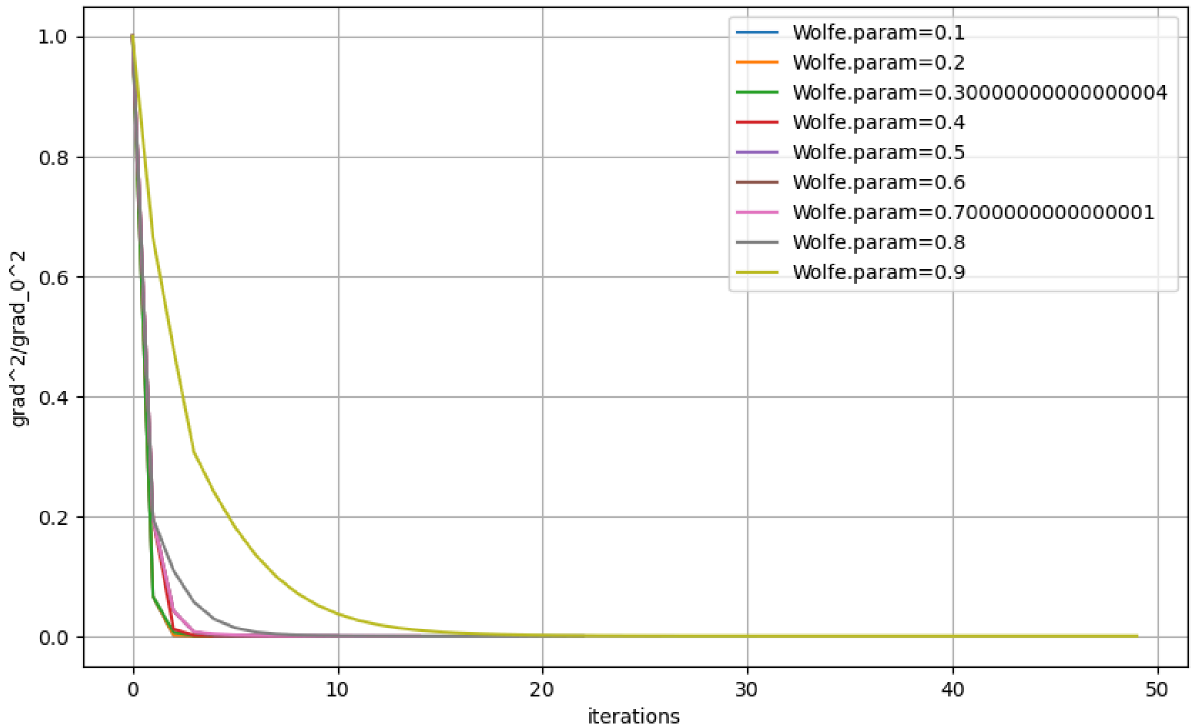

Рис. 19. Зависимость относительного градиента от итераций, вариация параметра c2

Для *c2* с учетом задачи с данными условиями лучше подбирать *c2* равную 0.1, обеспечивающее наибыстрейшую скорость сходимости.

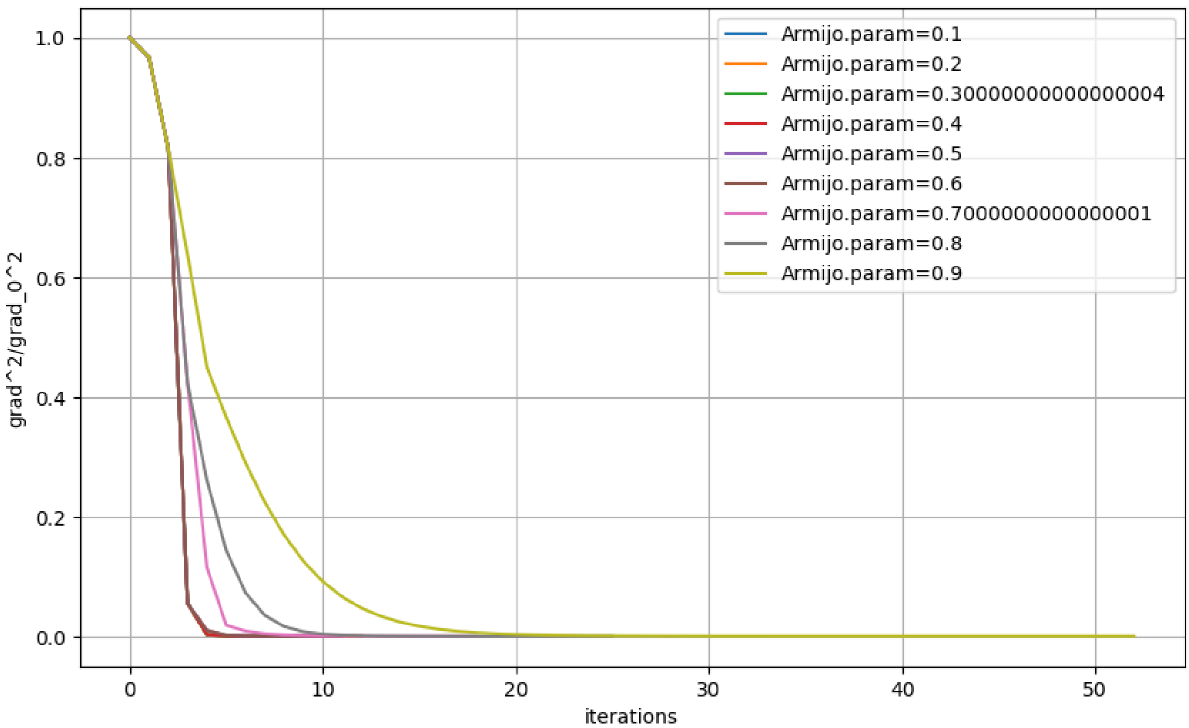

Рис. 20. Зависимость относительного градиента от итераций, вариация параметра c1

Для *c1* с учетом задачи с данными условиями лучше подбирать *c1* равную 0.1, обеспечивающее наибыстрейшую скорость сходимости.

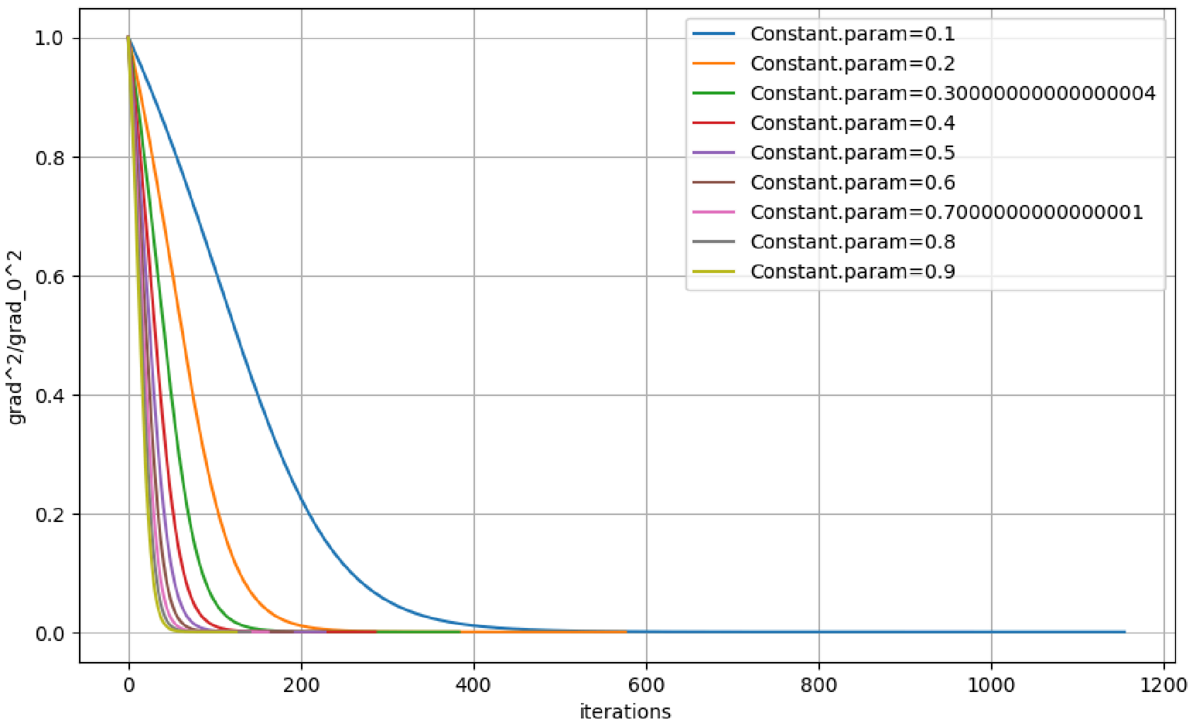

Рис. 21. Зависимость относительного градиента от итераций, вариация параметра *c*

Для константы с учетом задачи с данными условиями лучше подбирать значение равное 0.9, обеспечивающее наибыстрейшую скорость сходимости.

5. Стратегия выбора длины шага в методе Ньютона

Квадратичная задача размерностью 20x20 и числом обусловленности 500

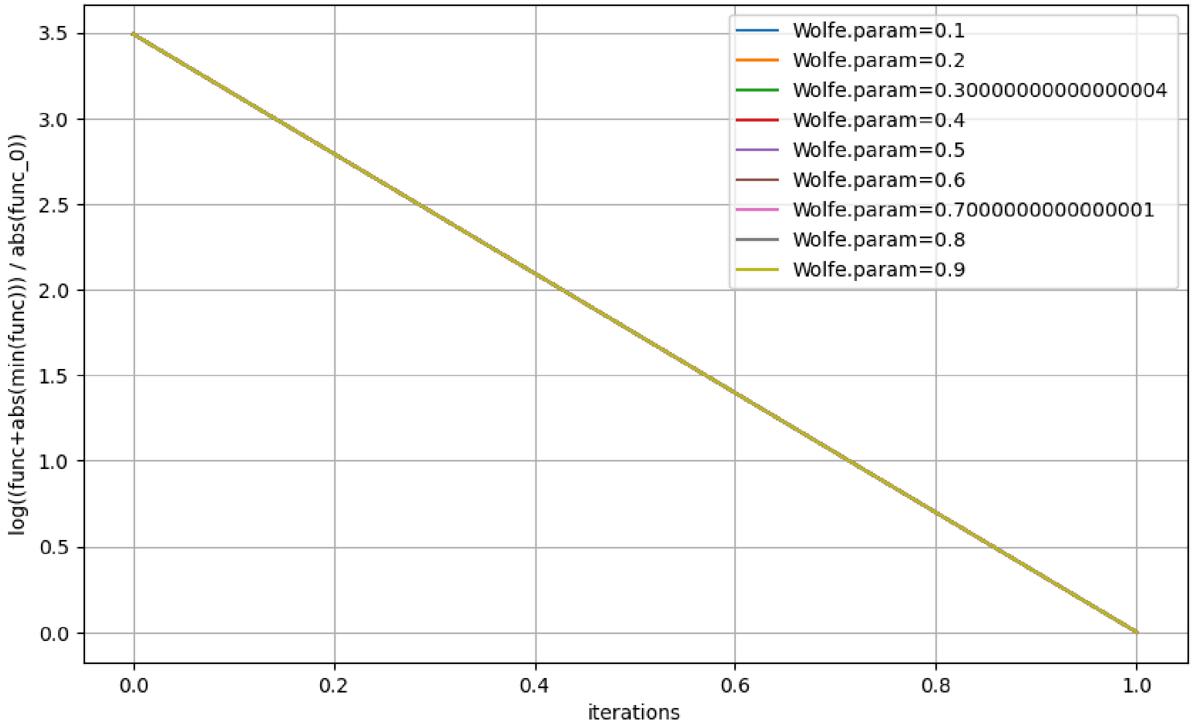

Рис. 22. Зависимость логарифма относительного значения функции от итераций, вариация параметра c2

Комментарий

Слишком быстро находится оптимум функции, если усложнить задачу, то в остальных условиях кривые накладываются друг на друга, поэтому решил оставить так.

Для *c2* с учетом задачи с данными условиями лучше подбирать *c2* равную 0.9, как наиболее частую рекомендацию.

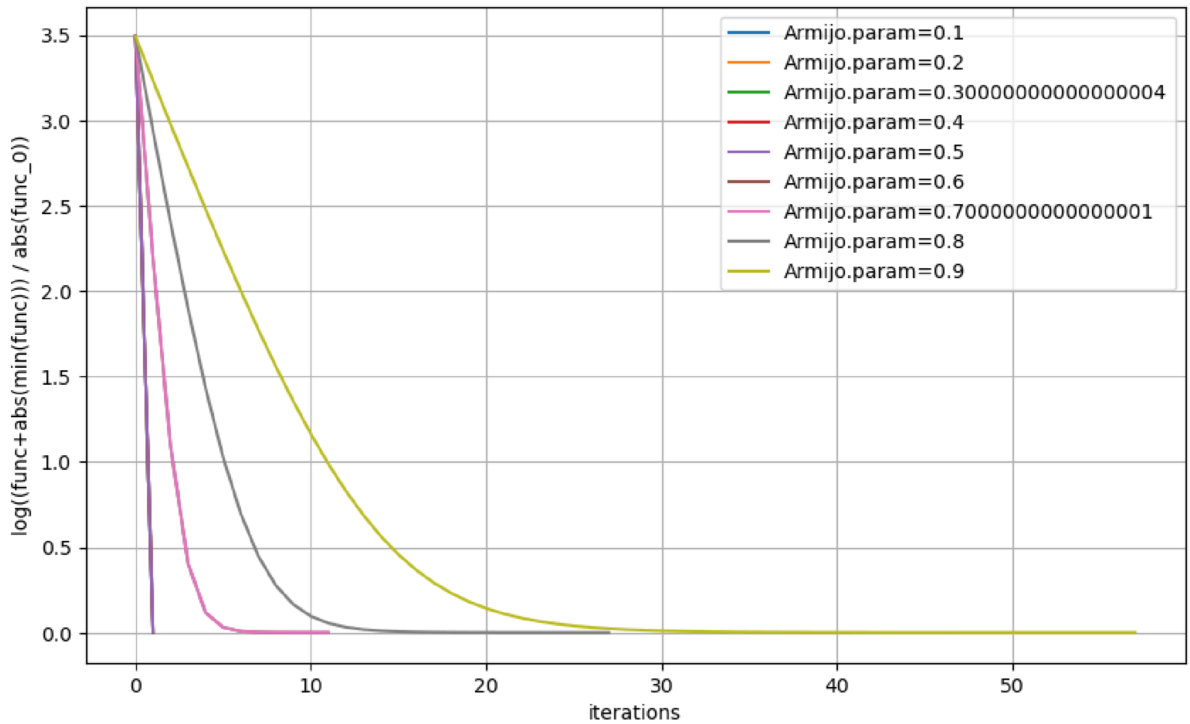

Рис. 23. Зависимость логарифма относительного значения функции от итераций, вариация параметра c1

Для *c1* с учетом задачи с данными условиями лучше подбирать *c1* равную 0.1, обеспечивающее наибыстрейшую скорость сходимости.

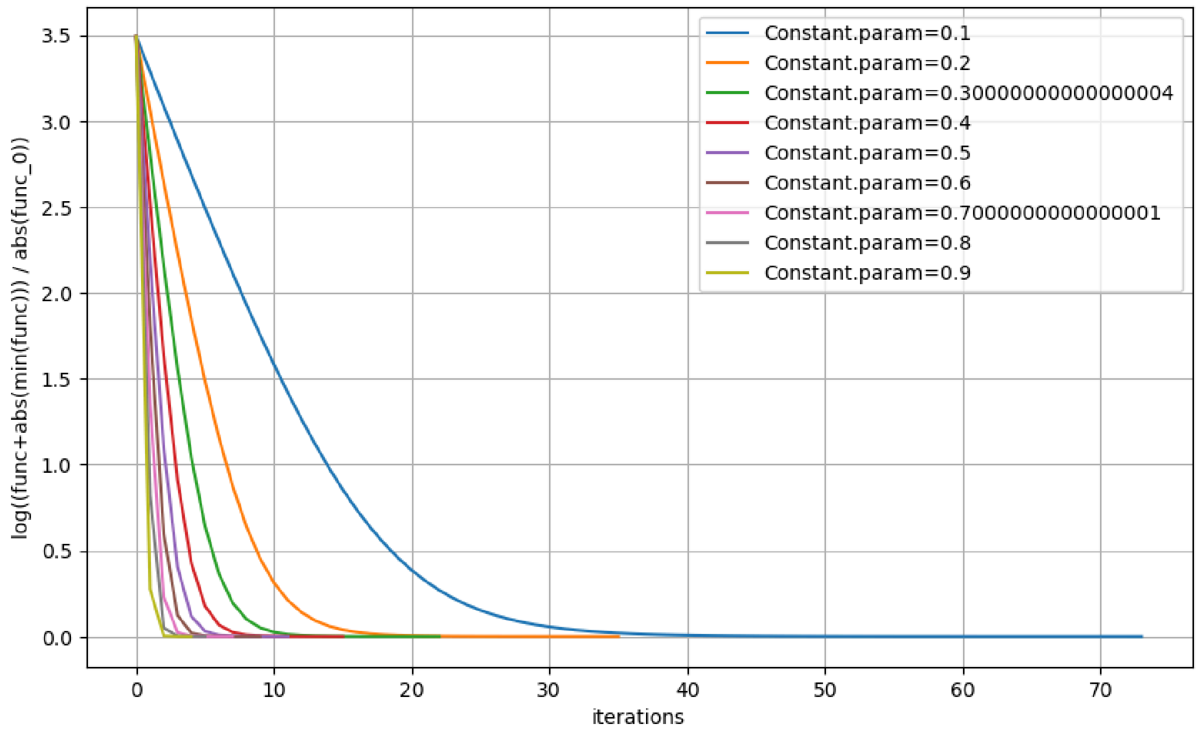

Рис. 24. Зависимость логарифма относительного значения функции от итераций, вариация параметра *c*

Для константы с учетом задачи с данными условиями лучше подбирать значение равное 0.9, обеспечивающее наибыстрейшую скорость сходимости.

Задача логистической регрессии размерностью 1000x20:

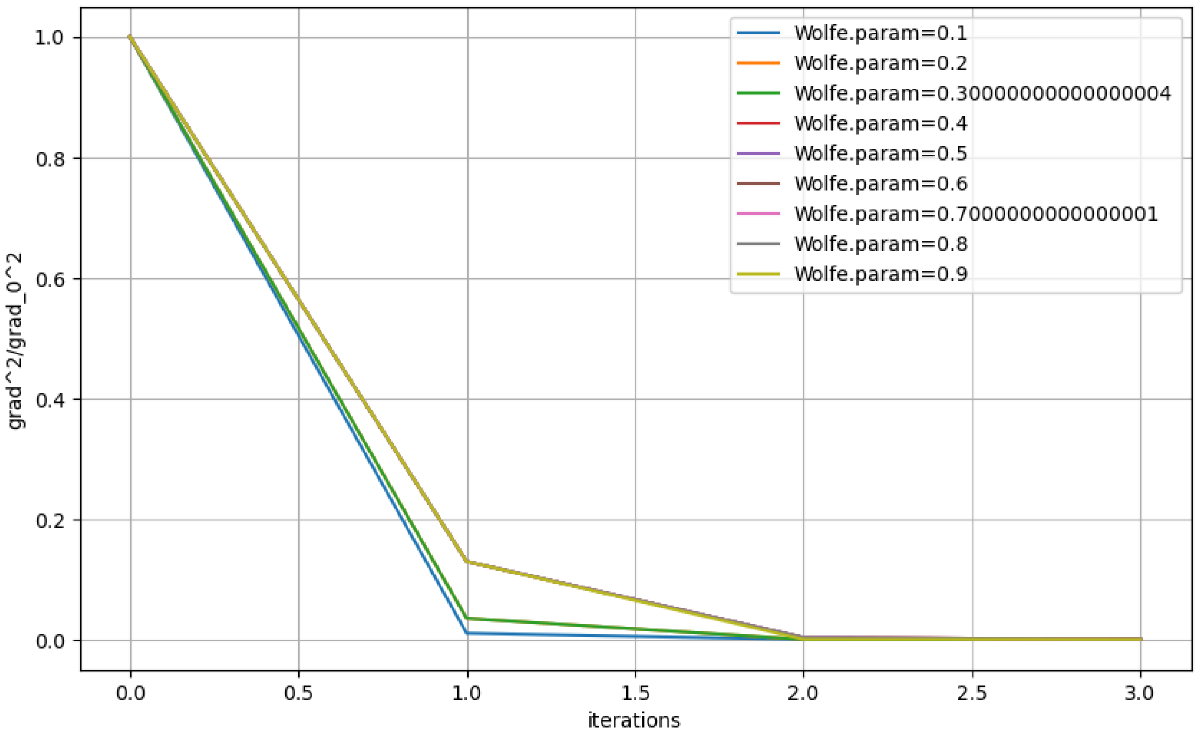

Рис. 25. Зависимость относительного градиента от итераций, вариация параметра c2

Для *c2* с учетом задачи с данными условиями лучше подбирать *c2* равную 0.1, обеспечивающее наибыстрейшую скорость сходимости.

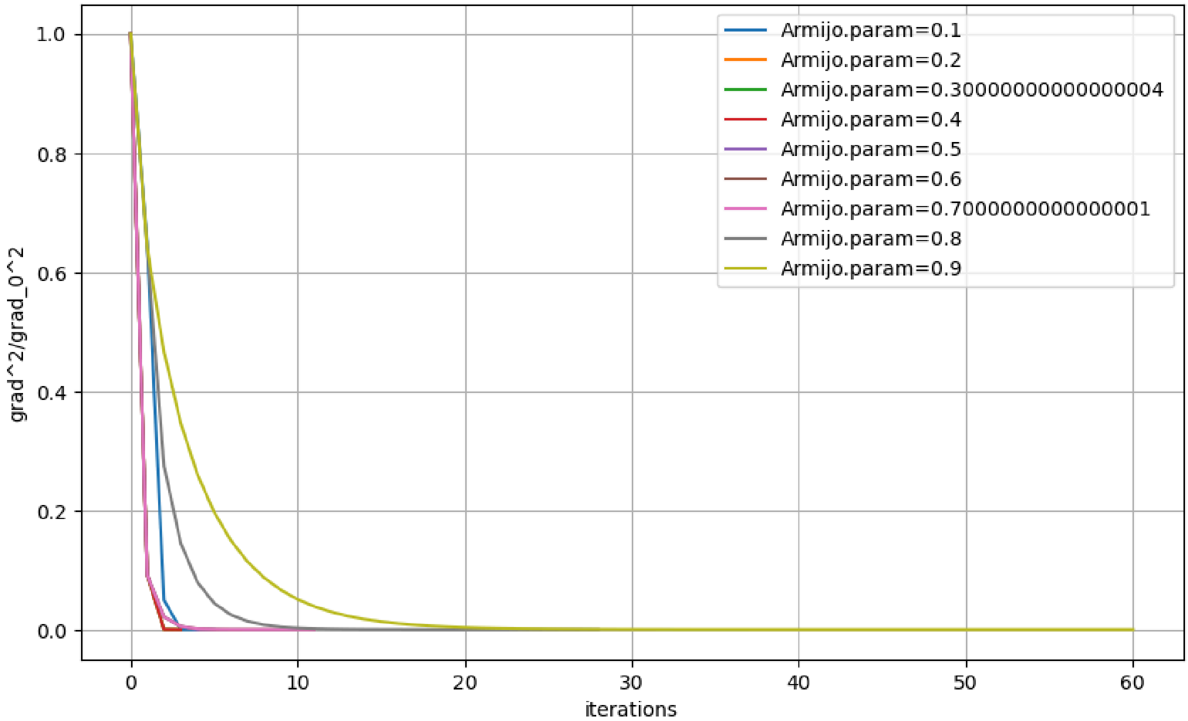

Рис. 26. Зависимость относительного градиента от итераций, вариация параметра c1

Для *c1* с учетом задачи с данными условиями лучше подбирать *c1* равную 0.4, обеспечивающее наибыстрейшую скорость сходимости.

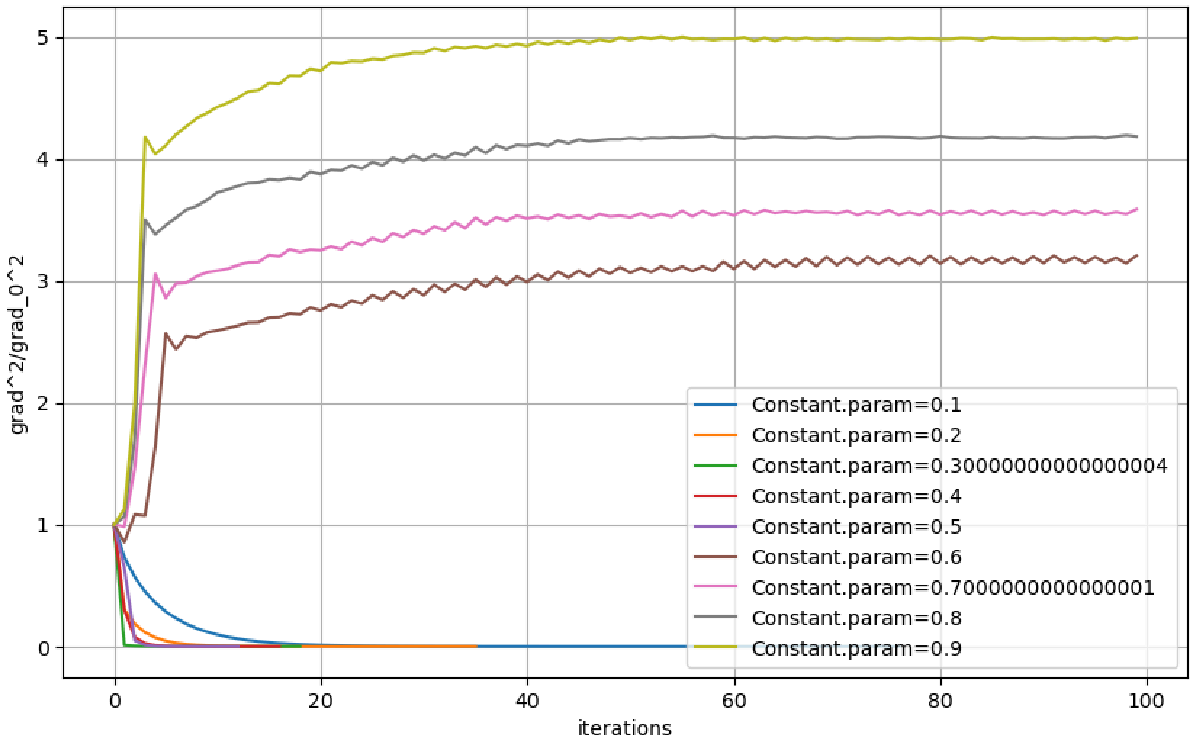

Рис. 27. Зависимость относительного градиента от итераций, вариация параметра *c*

Для константы с учетом задачи с данными условиями лучше подбирать значение равное 0.3, обеспечивающее наибыстрейшую скорость сходимости.

	Градиентный спуск			Метод Ньютона		
Квадратичная функция	C1=0.9	C2=0.9	C=0.009	C1=0.1	C2=0.9	C=0.9
Логистическая регрессия	C1=0.1	C2=0.1	C=0.9	C1=0.4	C2=0.1	C=0.3

В экспериментах наилучший результат, в контексте количества итераций, давали сильные условия Вульфе, чуть хуже условия Армихо, и хуже всего справлялся константный способ определения шага. В последнем случае константный способ для некоторых параметров выходил из области оптимального решения. Поэтому для нахождения шага рекомендуется по возможности использовать сильное условие Вульфе.

В среднем по результатам экспериментов можно сделать вывод, что для сильных условий Вульфе рекомендуется подбирать гиперпараметр *c2* равный 0.9 для квадратичной задачи, и 0.1 для логистической регрессии. Для Армихо потребуется небольшой перебор в пределах [0.1,…0.9]. Для константы требуется сильная корректировка под конкретную задачу, например в квадратичной для схождения потребовалось *c* равная 0.009, а в логистической 0.9 для одного и того же метода поиска.