

Отчет по первому практическому заданию

Царькова Анастасия

Формулы для градиента и гессиана функции логистической регрессии:

$$f(x) = \frac{1}{m} \sum_{i=1}^m \ln(1 + \exp(-b_i a_i^T x)) + \frac{\lambda}{2} \|x\|_2^2 = \frac{1}{m} \ln(1 + \exp(-b * Ax)) + \frac{\lambda}{2} x^T x$$

$$\nabla f(x) = -\frac{1}{m} A^T \left(b * \left(\frac{1}{1 - \exp(b * Ax)} \right) \right) + \lambda x = -\frac{1}{m} A^T (b * \text{Expit}(-b * Ax)) + \lambda x$$

$$\begin{aligned} \nabla^2 f(x) &= \frac{1}{m} A^T \text{Diag} \left(\left(1 - \left(\frac{1}{1 - \exp(b * Ax)} \right) \right) * \left(\frac{1}{1 - \exp(b * Ax)} \right) \right) A + \lambda I = \\ &= \frac{1}{m} A^T \text{Diag} (1 - \text{Expit}(b * Ax)) * (\text{Expit}(-b * Ax)) A + \lambda I \end{aligned}$$

Первый эксперимент.

Зависимость поведения метода от обусловленности функции.

В данном эксперименте предлагается проанализировать траекторию градиентного спуска в зависимости от:

1. Обусловленности функции
2. Выбора начальной точки
3. Стратегии выбора шага

Сравнивая графики с линиями уровня функции и траекториями методов.

Зададим объекты на которых будем сравнивать поведения методов:

1. Функции – двумерные квадратичные функции с d нулю (для простоты) $f(x) = x^T A x + b x = x^T A x$.

$$M_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad M_2 = \begin{pmatrix} 1 & 0 \\ 0 & 60 \end{pmatrix}$$

2. Начальные точки

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad x_2 = \begin{pmatrix} 20 \\ 30 \end{pmatrix}$$

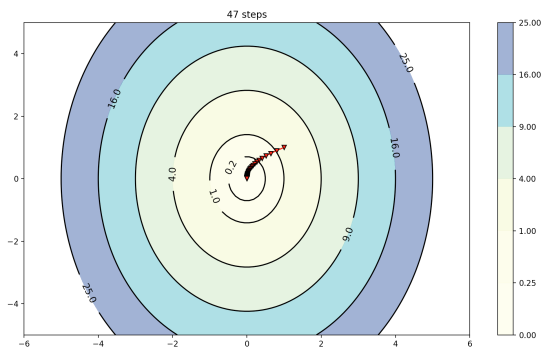
3. Стратегии выбора шага: Armijo, Wolfe, Constant

Результаты эксперимента:

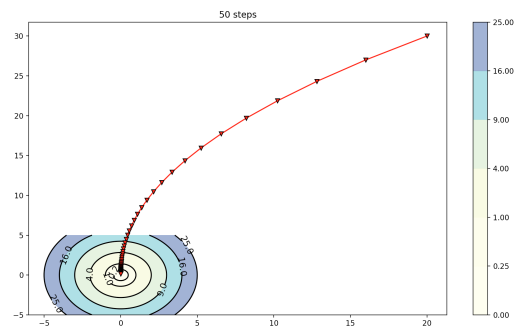
- Constant

Видно что метод работает не очень хорошо. Делает слишком много шагов, но не слишком чувствителен к выбору начальной точки.

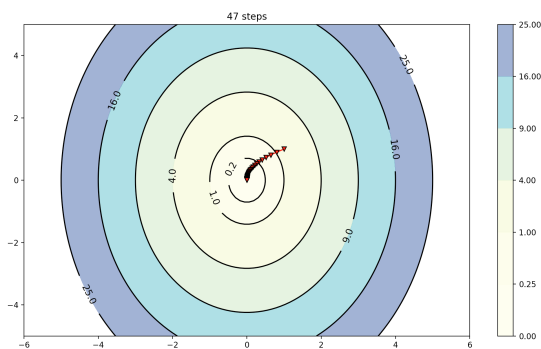
Более того крайне чувствителен к обусловленности функции, на M_2 метод возвращает *computational error*, поэтому я взяла вместо нее $M = \begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix}$ и все равно получила *iterations exceeded*, но полученные графики очень показательны. Кажется что метод вряд ли сойдется при увеличении числа итераций.



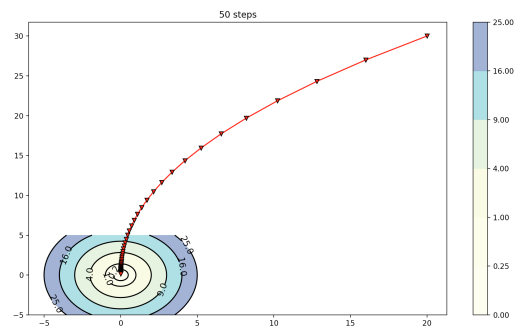
a) M_1, x_1



b) M_1, x_2



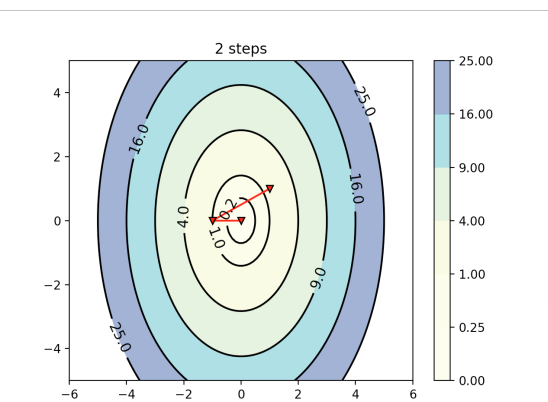
a) M, x_1



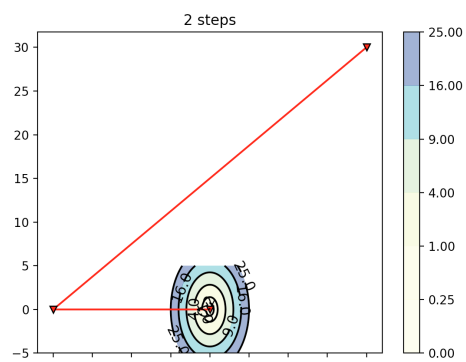
b) M, x_2

- Armijo

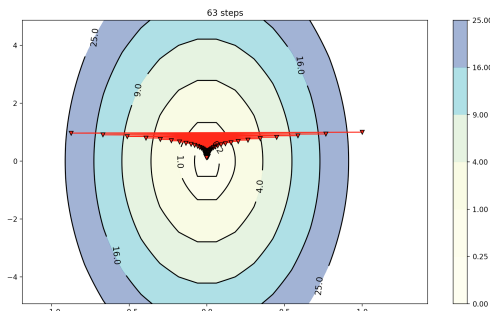
Видно что метод очень чувствителен к обусловленности функции. Так же есть небольшая зависимость о выбора начальной точки.



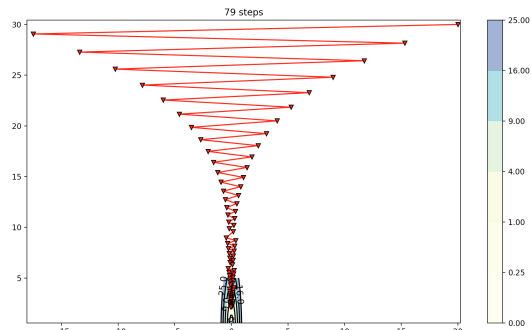
a) M_1, x_1



b) M_1, x_2



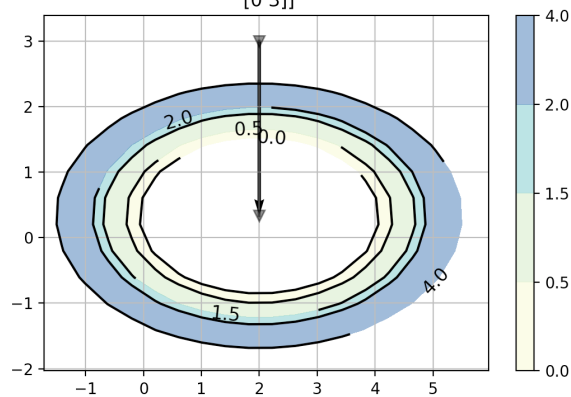
a) M_2, x_1



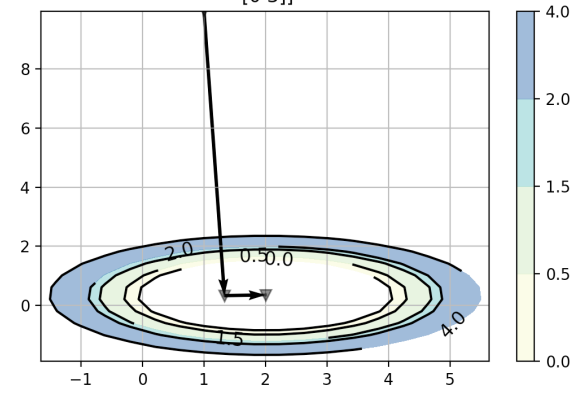
b) M_2, x_2

Зависимость поведения метода от начальной точки

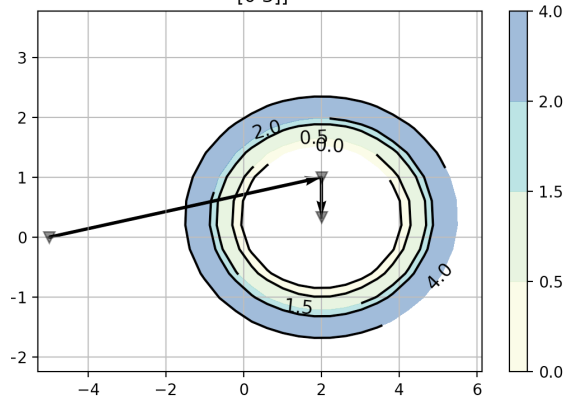
gradient descent with Wolfe linear search, steps: 1
from point [2. 3.]
for Quadratic function with
 $A = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$



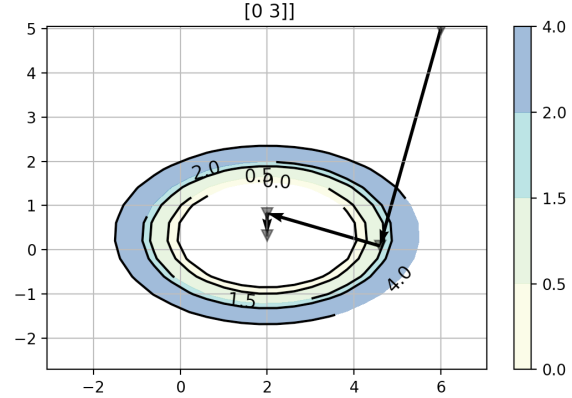
gradient descent with Wolfe linear search, steps: 2
from point [1. 10.]
for Quadratic function with
 $A = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$



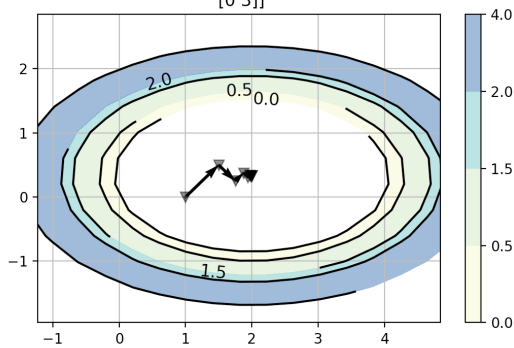
gradient descent with Wolfe linear search, steps: 2
from point [-5. 0.]
for Quadratic function with
 $A = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$



gradient descent with Wolfe linear search, steps: 3
from point [6. 5.]
for Quadratic function with
 $A = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$

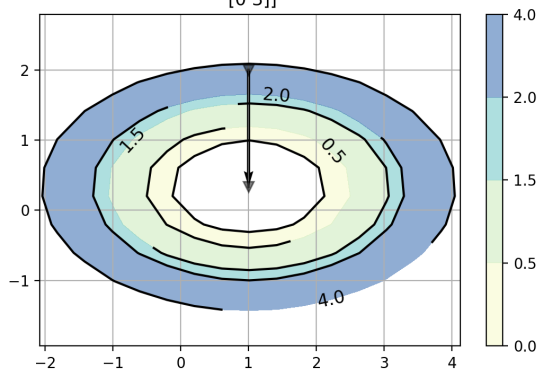


gradient descent with Wolfe linear search, steps: 9
 from point [1. 0.]
 for Quadratic function with
 $A = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$

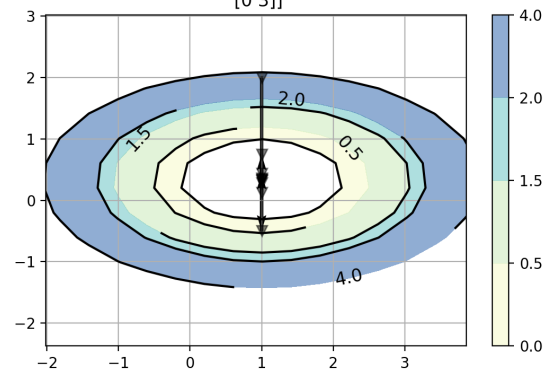


Зависимость поведения метода от стратегии выбора шага.

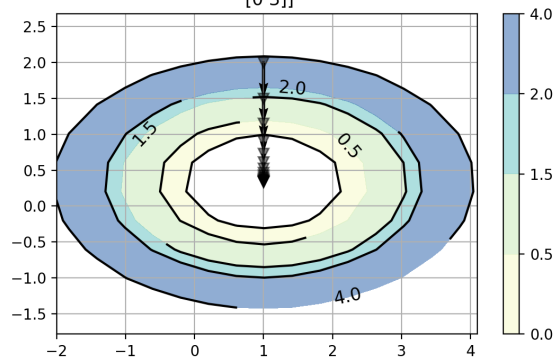
gradient descent with Wolfe linear search, steps: 1
 from point [1. 2.]
 for Quadratic function with
 $A = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$



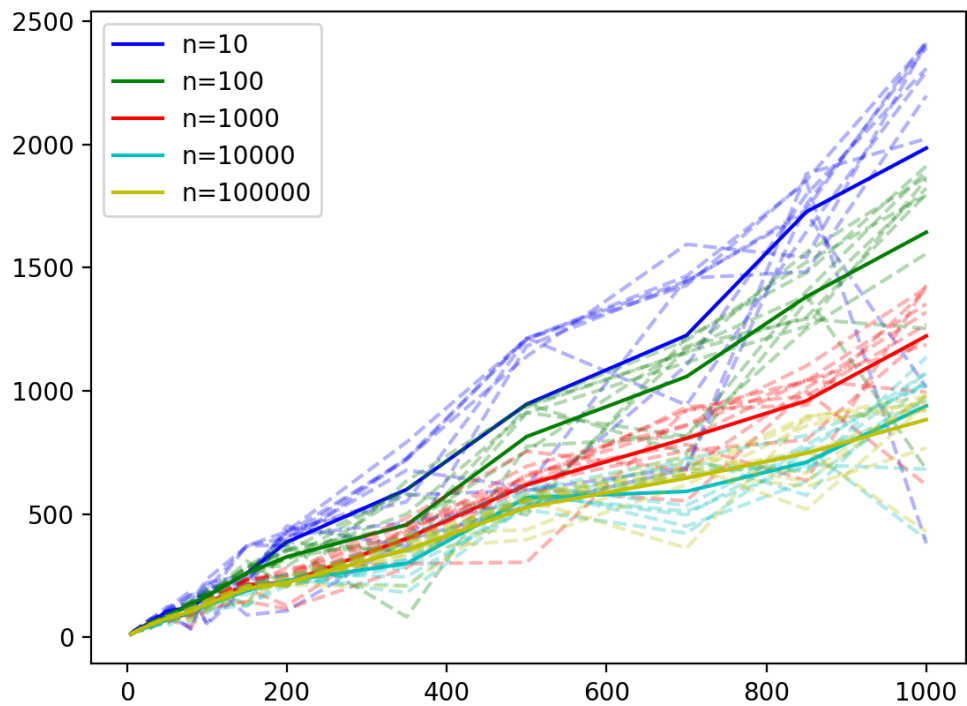
gradient descent with Armijo linear search, steps: 9
 from point [1. 2.]
 for Quadratic function with
 $A = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$



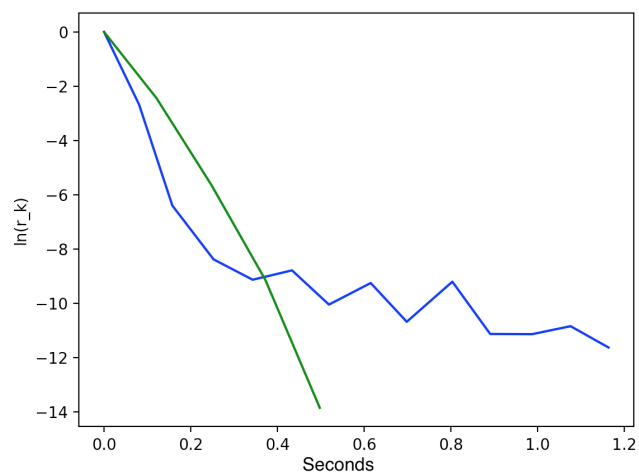
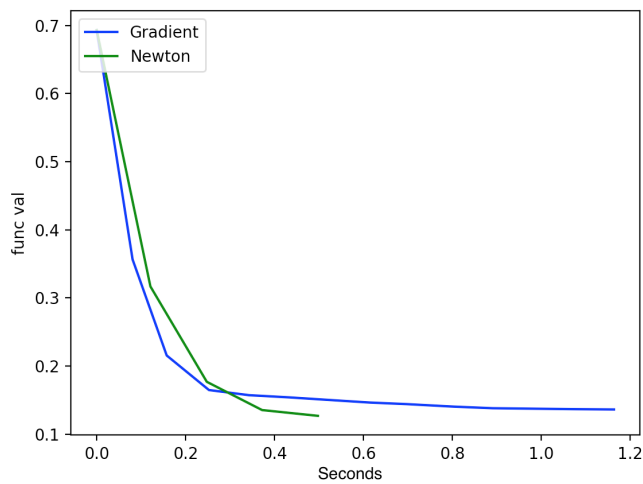
gradient descent with Constant linear search, steps: 17
 from point [1. 2.]
 for Quadratic function with
 $A = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$



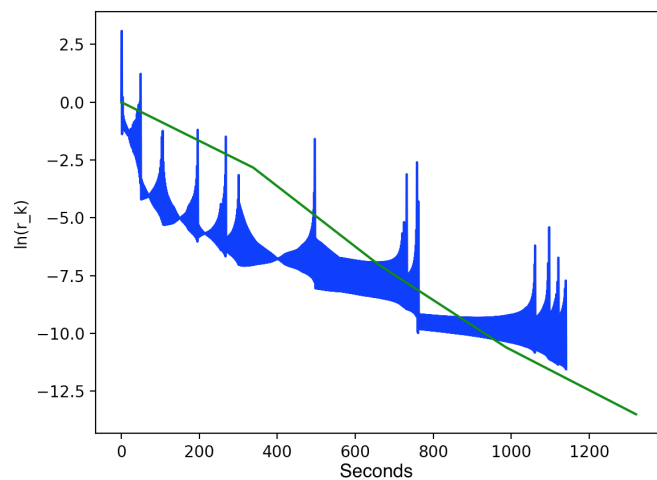
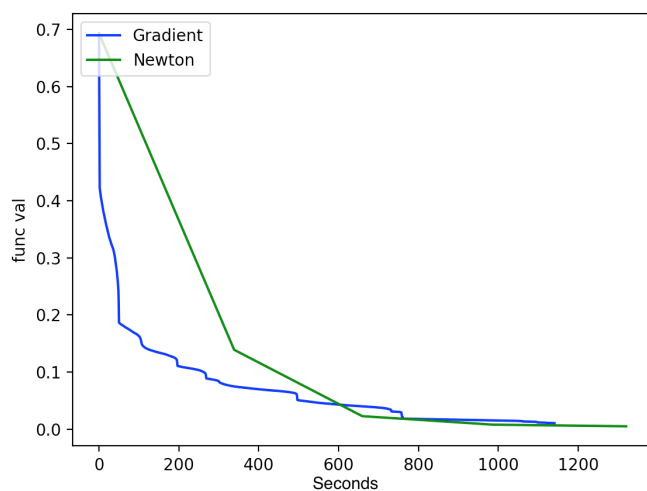
Зависимость числа итераций градиентного спуска от числа обусловленности и размерности пространства.



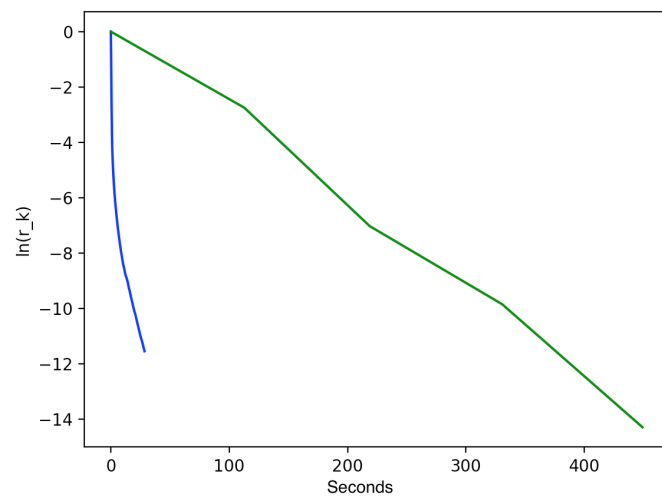
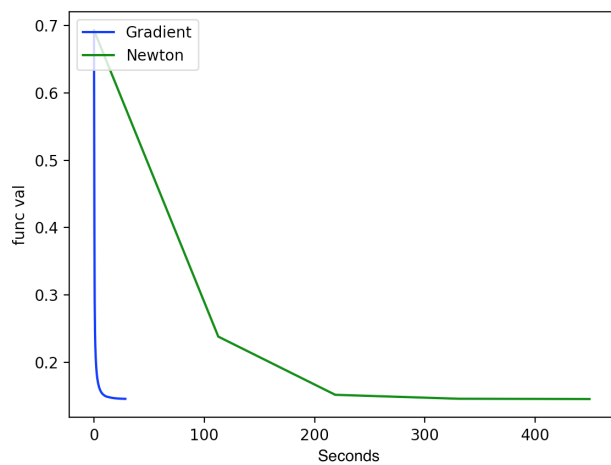
Сравнение методов градиентного спуска и Ньютона на реальной задаче логистической регрессии.
w8a



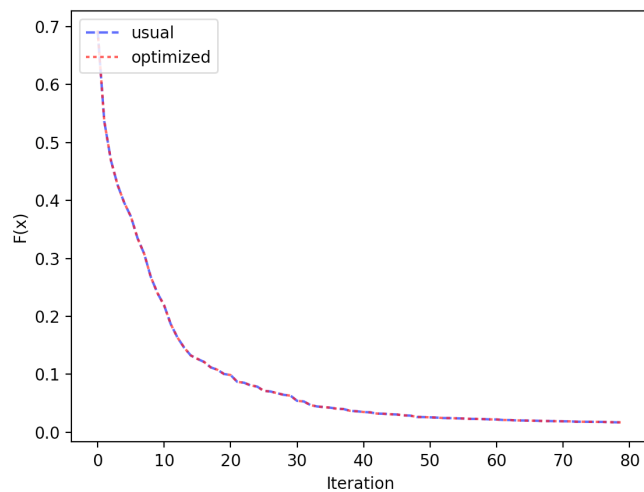
gisette-scale

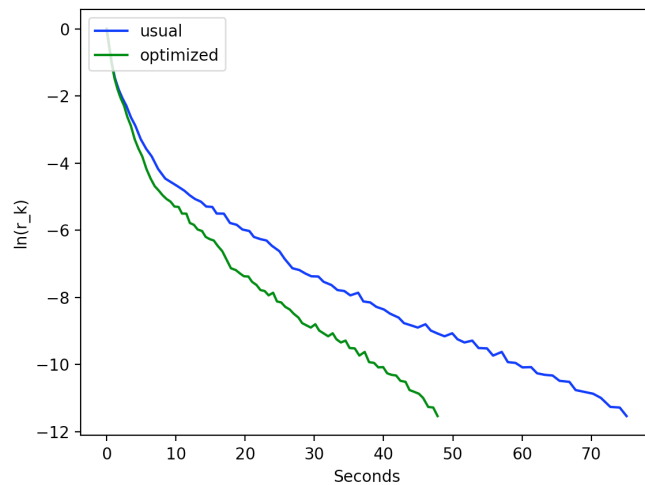
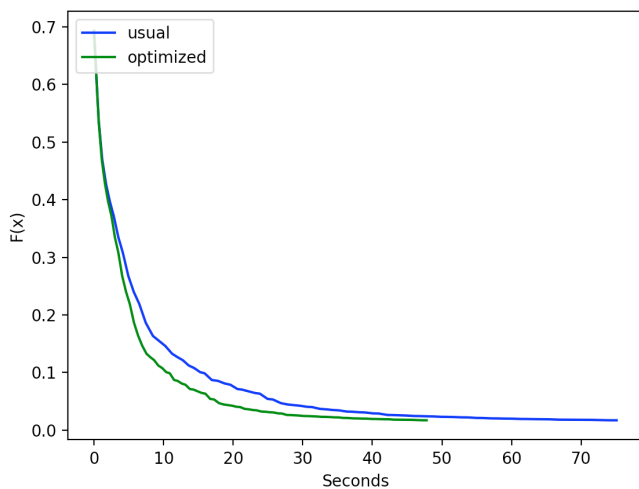


real-sim

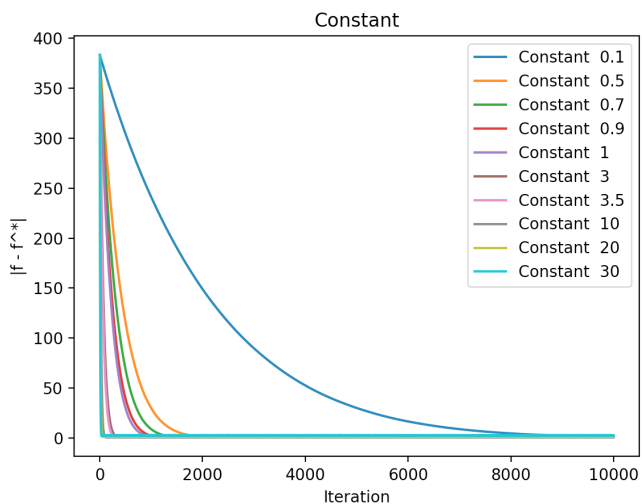
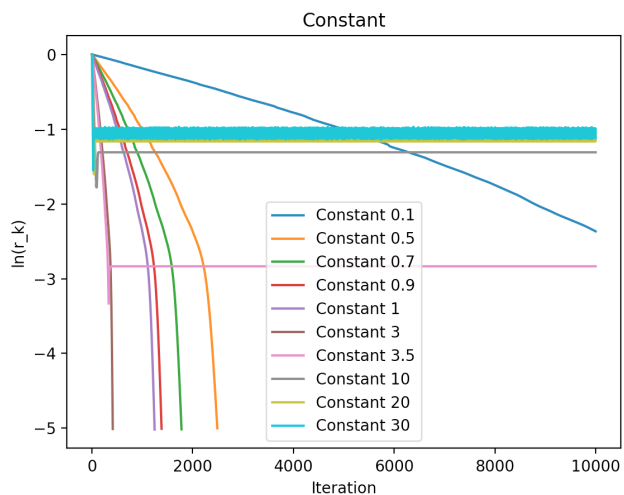
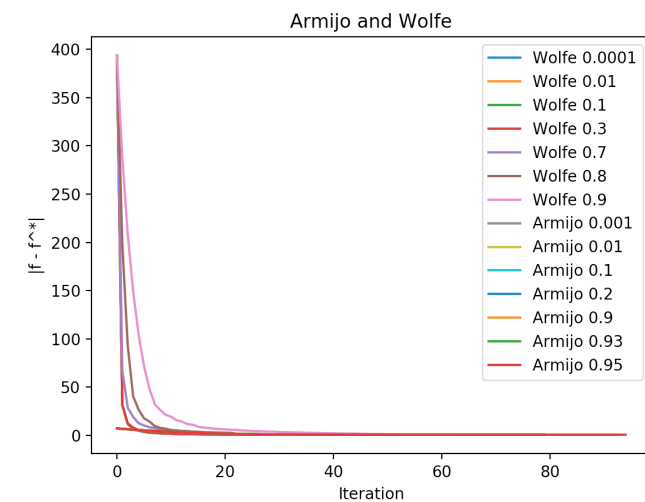
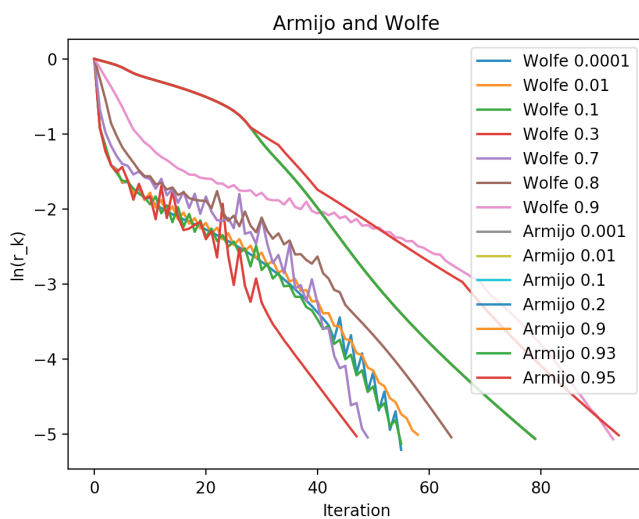


Оптимизация вычислений в градиентном спуске.





Стратегия выбора длины шага в градиентном спуске



Стратегия выбора длины шага в методе ньютона

