Третьякова Д. А.

Лабораторная работа №2

Продвинутые методы безусловной оптимизации

1. Эксперимент №1. Зависимость числа итераций метода сопряженных градиентов от числа обусловленности и размерности пространства

Эксперимент заключается в исследовании зависимости числа итераций для метода сопряженных градиентов и для градиентного спуска от числа обусловленности $\kappa \geq 1$ оптимизируемой функции и размерности пространства n.

Числа обусловленности от 1 до 1000 с шагом 50.

Размерность пространства 10, 100, 1000.

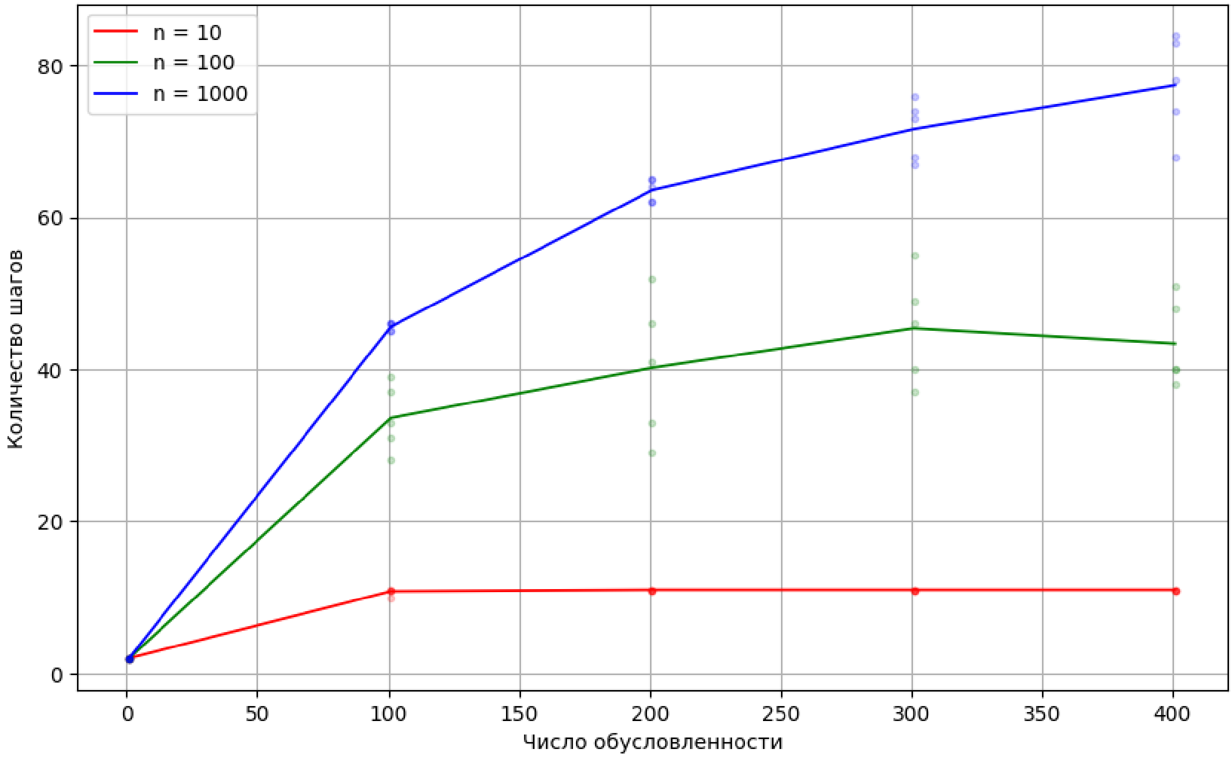


Рисунок 1.1 Метод сопряженных градиентов

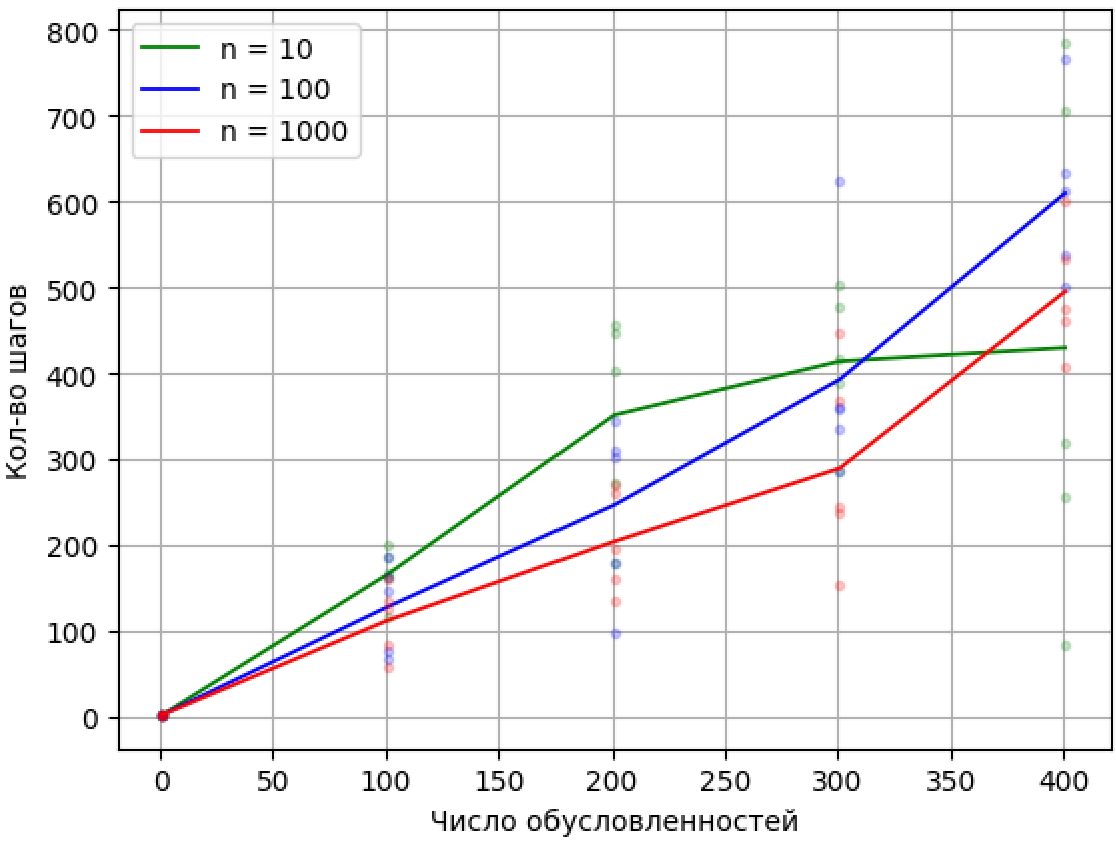


Рисунок 1.2 Градиентный спуск

Эксперимент показал, что метод сопряжённых градиентов значительно лучше справляется с задачами при высоких значениях числа обусловленности, чем градиентный спуск. Даже при больших числах обусловленности в методе сопряженных градиентов количество итераций не превышает размерность пространства.

Зависимость числа итераций от числа обусловленности имеет нелинейный характер, в отличие от градиентного спуска.

2. Эксперимент №2. Выбор размера истории в методе L-BFGS

Цель эксперимента заключается в исследовании поведения метода L-BFGS в зависимости от истории.

2.1 Датасет w8a

Размер истории: 0, Итерации: 27, Время: 1.032865
Размер истории: 5, Итерации: 9, Время: 0.23213
Размер истории: 10, Итерации: 9, Время: 0.148614
Размер истории: 50, Итерации: 9, Время: 0.147475
Размер истории: 100, Итерации: 9, Время: 0.136741
Размер истории: 500, Итерации: 9, Время: 0.143478

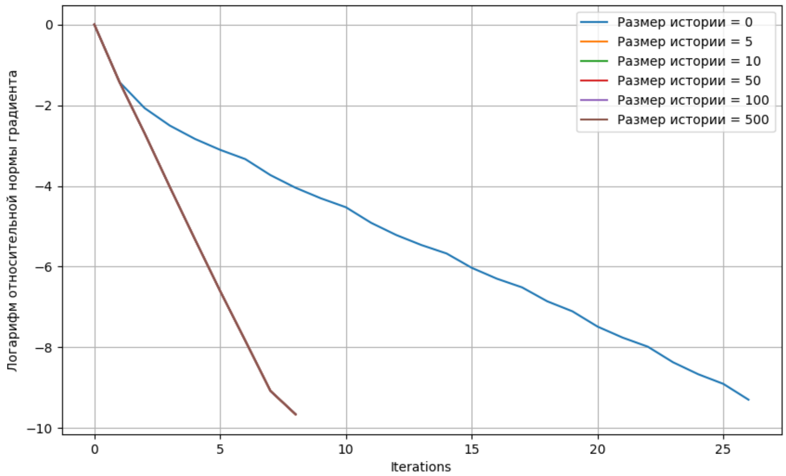
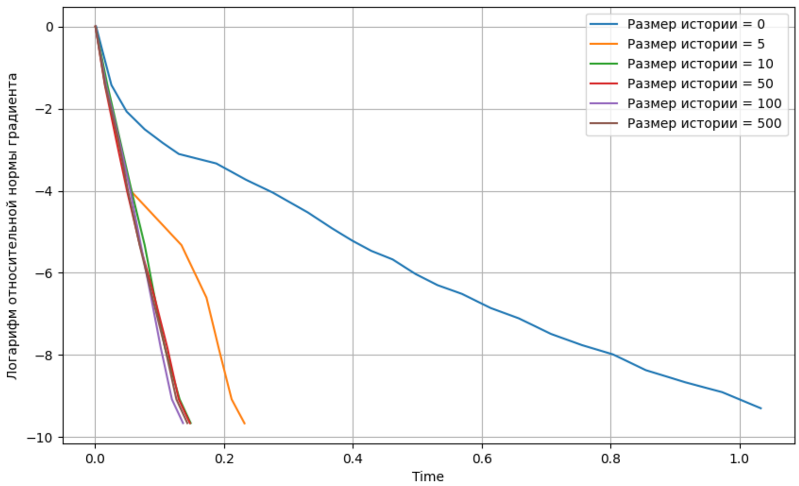


Рисунок 2.1.1 Зависимость логарифма относительной нормы градиента против количества итераций



2.1.2 Зависимость логарифма относительной нормы градиента против реального времени работы

2.2 Датасет gisette_scale

Размер истории: 0, Итерации: 240, Время: 22.308534
Размер истории: 5, Итерации: 22, Время: 1.654134
Размер истории: 10, Итерации: 22, Время: 1.748008
Размер истории: 50, Итерации: 21, Время: 1.605068
Размер истории: 100, Итерации: 21, Время: 1.622715
Размер истории: 500, Итерации: 21, Время: 1.643041

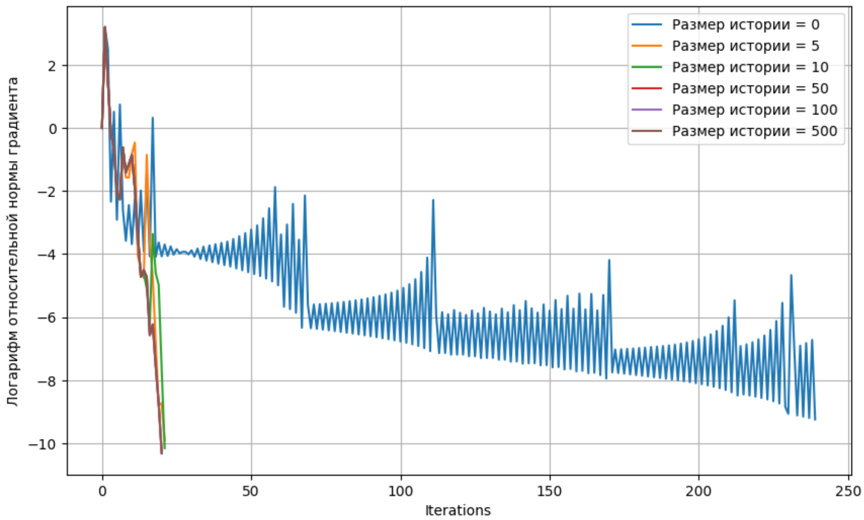
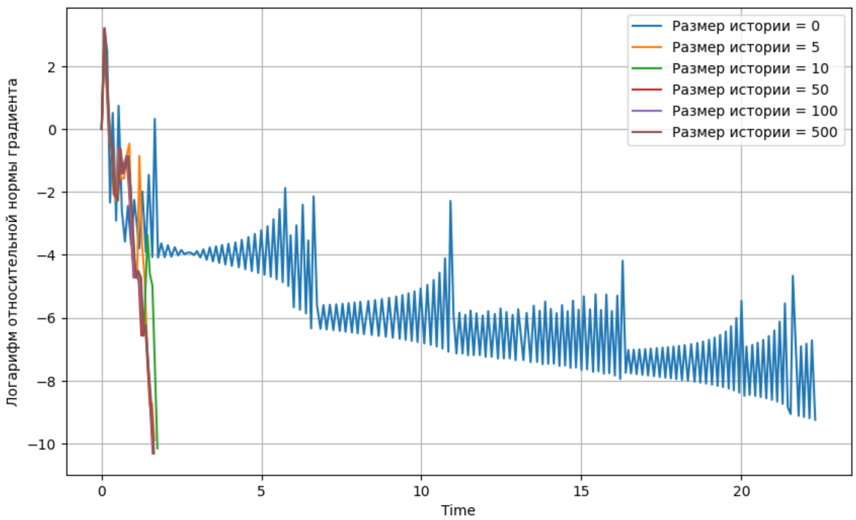


Рисунок 2.2.1 Зависимость логарифма относительной нормы градиента против количества итераций



2.2.2 Зависимость логарифма относительной нормы градиента против реального времени работы

При размере истории, равной нулю, данный метод работает как градиентный спуск. При размере истории не равном нулю, метод сводит количество итераций и затраченного времени к минимуму, что видно на графиках по обоим датасетам.

На рисунке 2.1.2 наглядно видно, что минимальные значения по времени мы получаем, если размер истории равен 10 или более.

3. Эксперимент №3. Сравнение методов на реальной задаче логистической регрессии

3.1 Датасет w8a

Градиентный спуск, итерации: 36, время: 1.494898, 0.042
Метод Гессиан свободного Ньютона, итерации: 7, время: 0.207183, 0.03
Метод L-BFGS, итерации: 9, время: 0.150645, 0.017

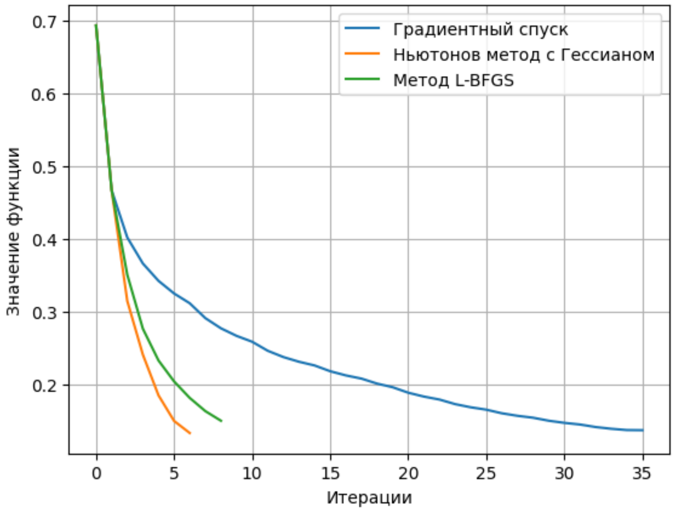


Рисунок 3.1.1 Зависимость значения функции от номера итерации метода

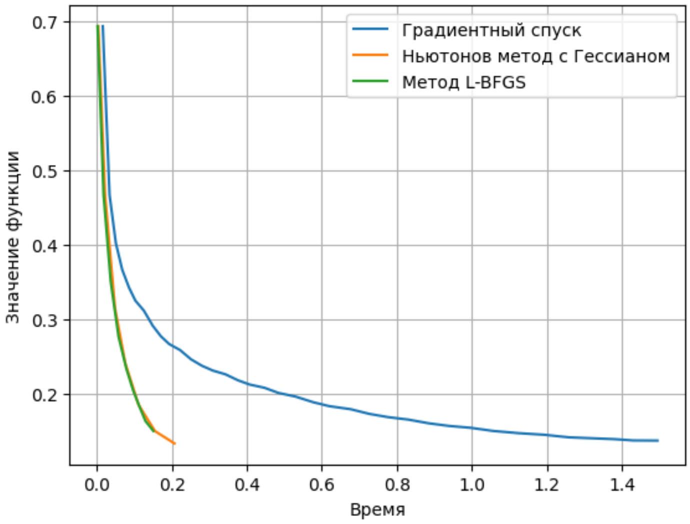


Рисунок 3.1.2 Зависимость значения функции от времени работы метода

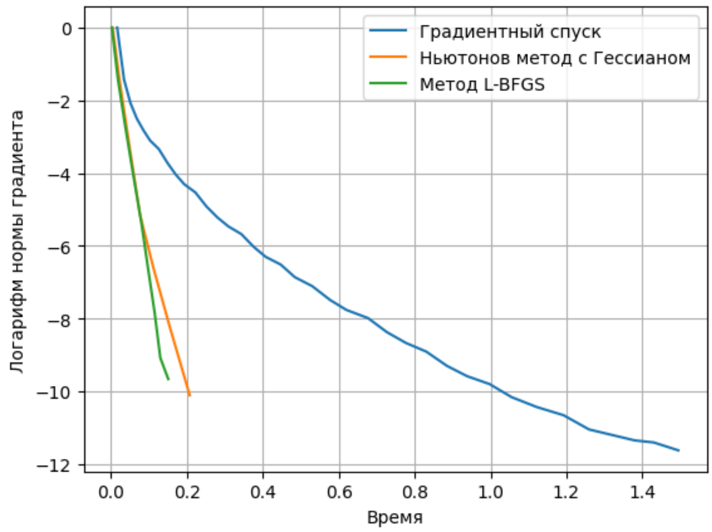


Рисунок 3.1.3 Зависимость логарифма нормы градиента от времени работы метода

3.2 Датасет real-sim

Градиентный спуск, итерации: 104, время: 30.989751, 0.298
Метод Гессиан свободного Ньютона, итерации: 5, время: 0.84617, 0.169
Метод L-BFGS, итерации: 10, время: 0.7855, 0.079

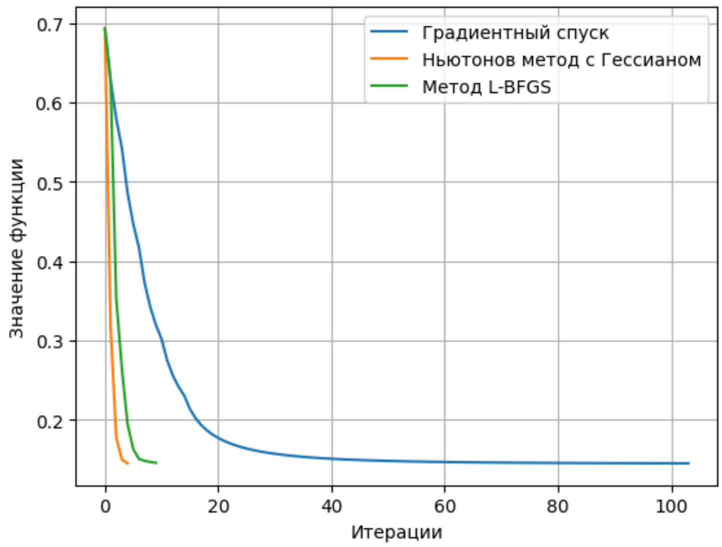


Рисунок 3.1.1 Зависимость значения функции от номера итерации метода

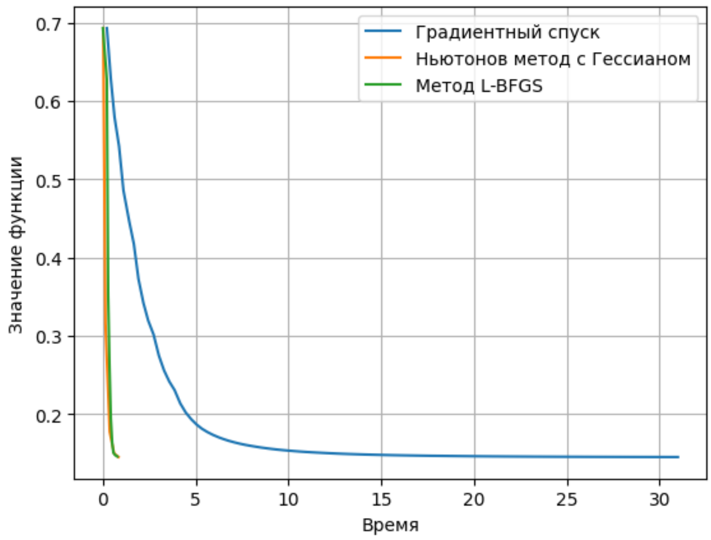


Рисунок 3.1.2 Зависимость значения функции от времени работы метода

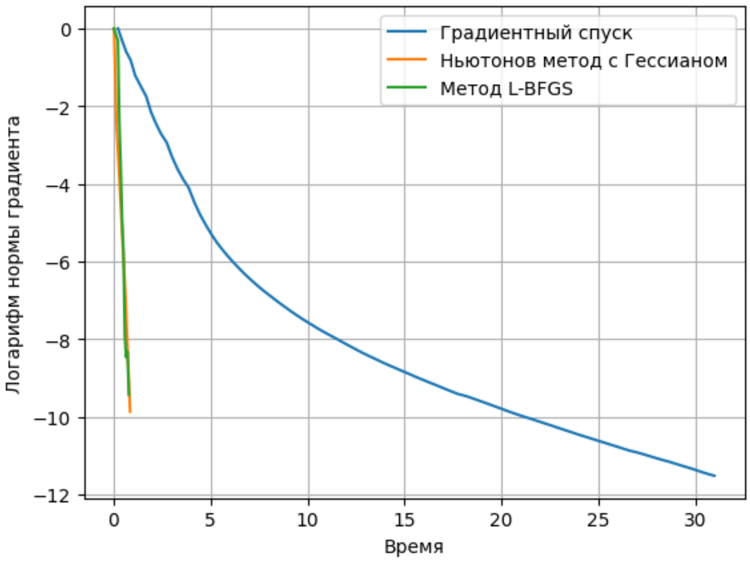


Рисунок 3.1.3 Зависимость логарифма нормы градиента от времени работы метода

3.3 Датасет gisette_scale

Градиентный спуск, итерации: 1832, время: 970.088765, 0.53
Метод Гессиан свободного Ньютона, итерации: 7, время: 8.879354, 1.268
Метод L-BFGS, итерации: 33, время: 13.69326, 0.415

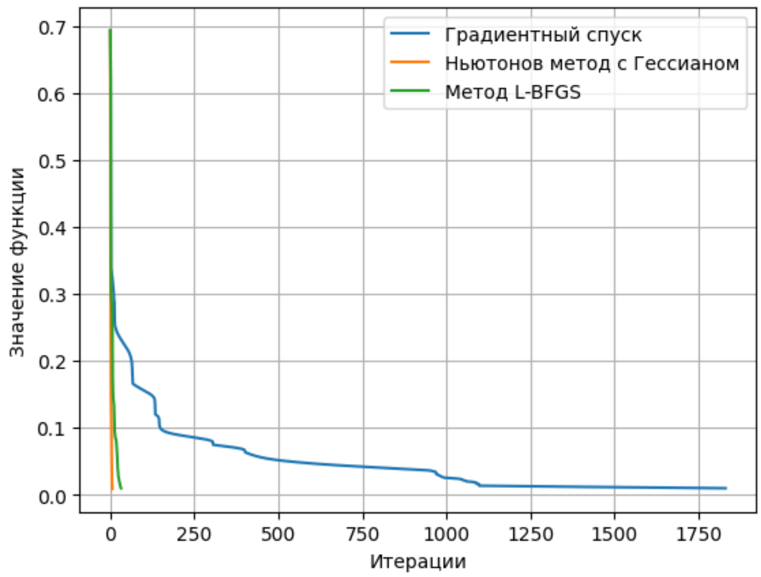


Рисунок 3.1.1 Зависимость значения функции от номера итерации метода

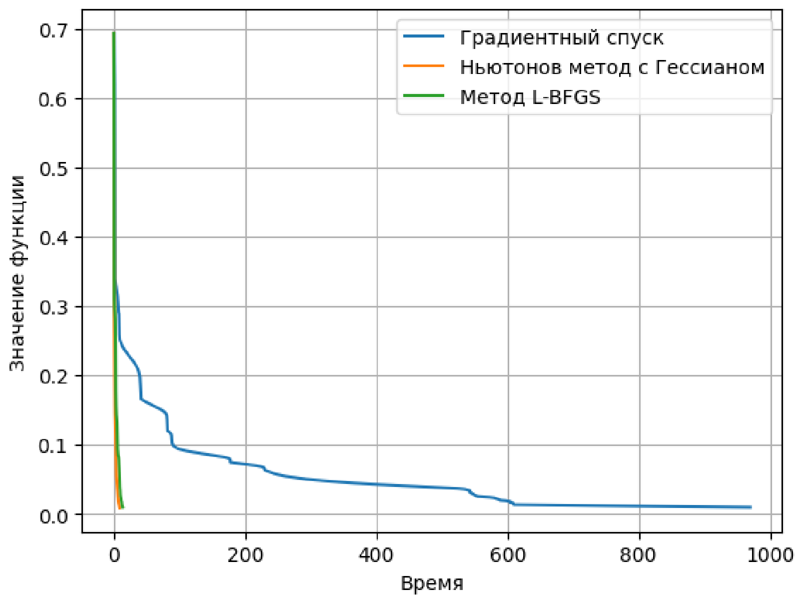


Рисунок 3.1.2 Зависимость значения функции от времени работы метода

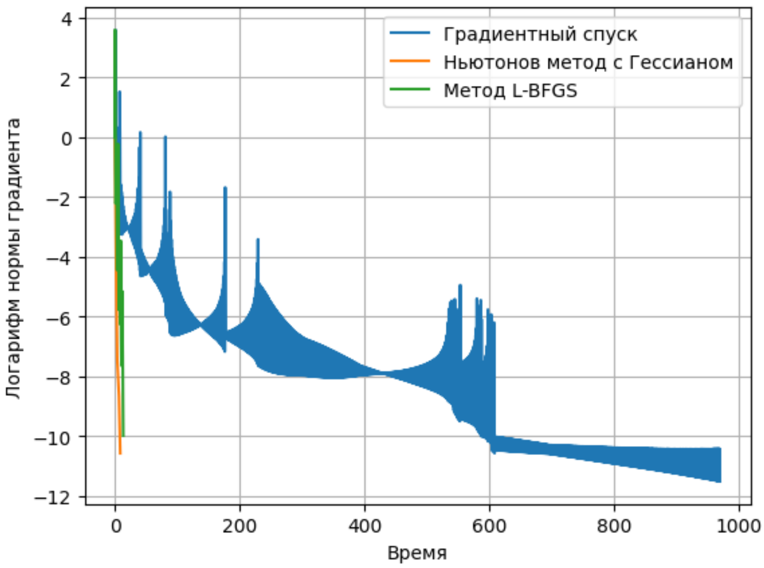


Рисунок 3.1.3 Зависимость логарифма нормы градиента от времени работы метода

3.4 Датасет news20.binary

Градиентный спуск, итерации: 153, время: 208.743946, 1.364
Метод Гессиан свободного Ньютона, итерации: 4, время: 3.933346, 0.983
Метод L-BFGS, итерации: 14, время: 7.421368, 0.53

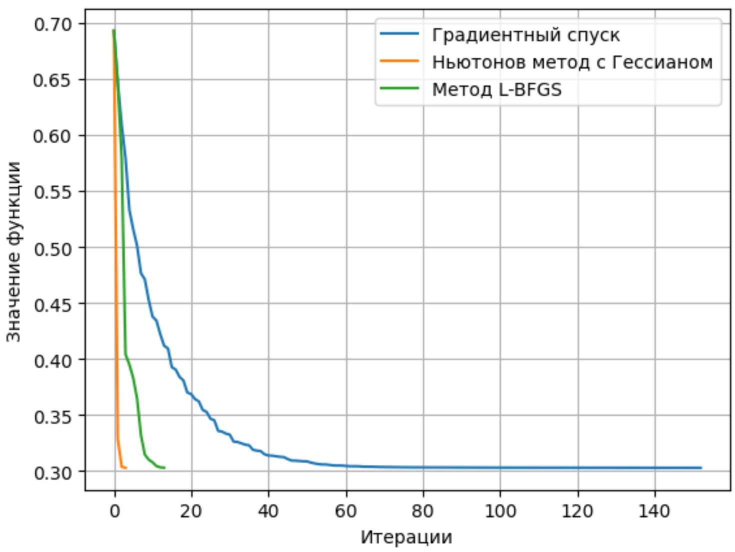


Рисунок 3.1.1 Зависимость значения функции от номера итерации метода

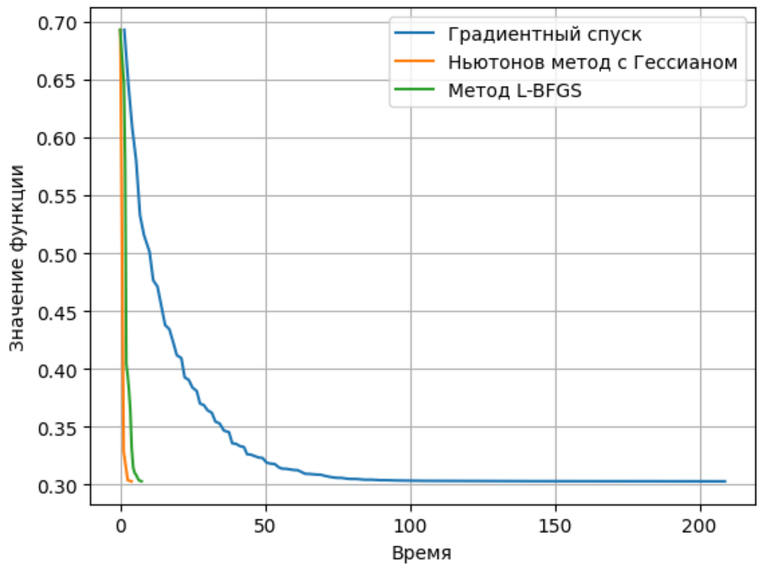


Рисунок 3.1.2 Зависимость значения функции от времени работы метода

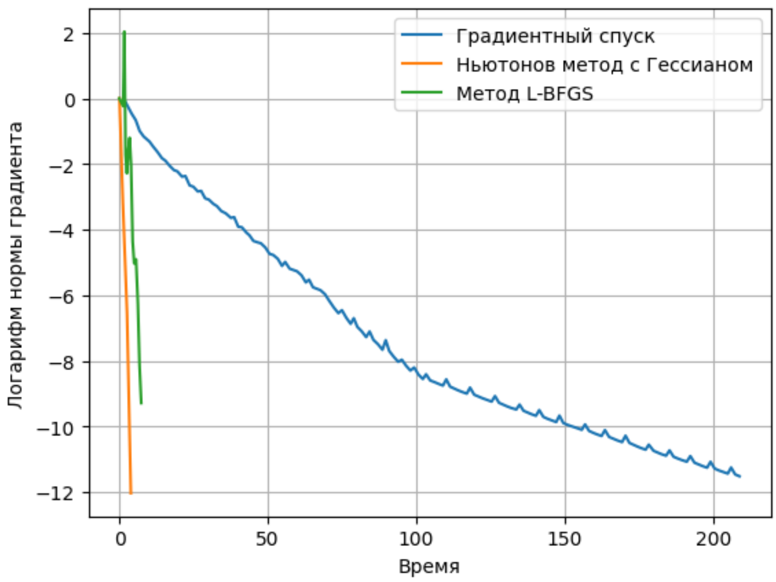


Рисунок 3.1.3 Зависимость логарифма нормы градиента от времени работы метода

3.5 Датасет rcv1_train.binary

Градиентный спуск, итерации: 52, время: 6.16369, 0.119
Метод Гессиан свободного Ньютона, итерации: 5, время: 0.345567, 0.069
Метод L-BFGS, итерации: 11, время: 0.367761, 0.033

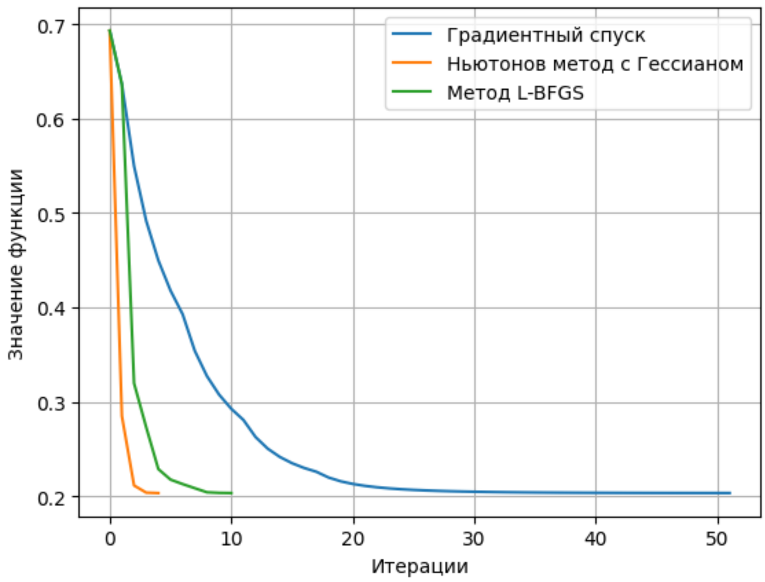


Рисунок 3.1.1 Зависимость значения функции от номера итерации метода

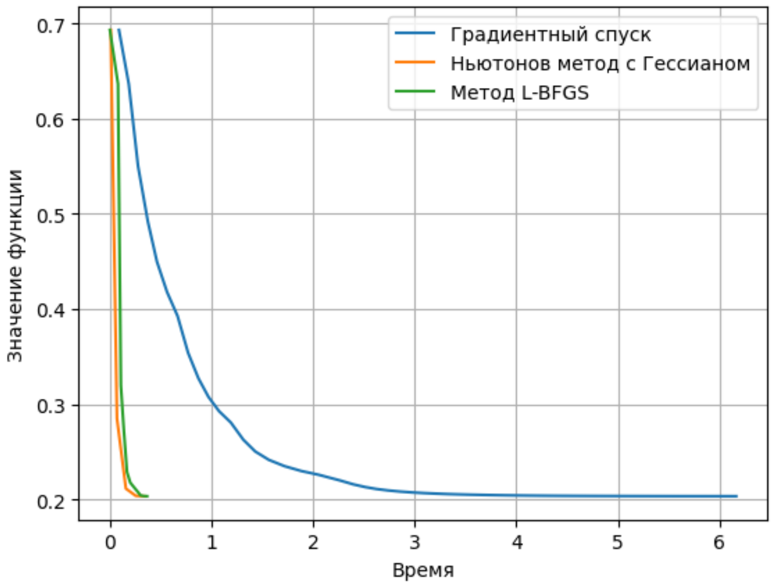


Рисунок 3.1.2 Зависимость значения функции от времени работы метода

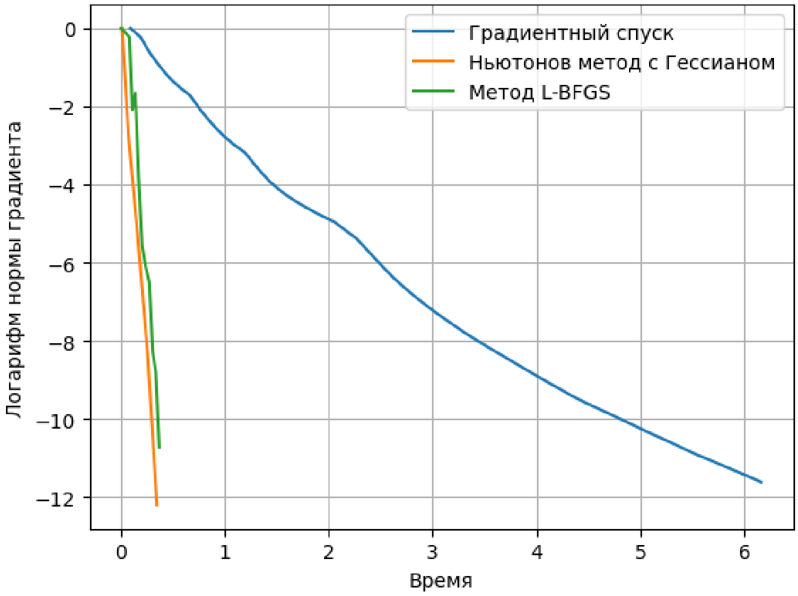


Рисунок 3.1.3 Зависимость логарифма нормы градиента от времени работы метода

Для всех датасетов самым долгим методом является градиентный спуск: чтобы сойтись, ему требуется гораздо большее количество итераций и больше времени. Время, затраченное на одну итерацию градиентного спуска, так же в разы больше времени, затраченного на одну итерацию оставшихся двух методов.

Ньютонов метод с гессианом сходится за чуть меньшее количество итераций, чем метод L-BFGS, однако эти оба метода имеют примерно одинаковое время работы за счет того, что одна итерация ньютонова метода с гессианом требует чуть больше времени, чем итерация L-BFGS метода.