

Лабораторная работа 2:

Продвинутые методы безусловной ОПТИМИЗАЦИИ.

Сливкина Ксения

1 июня 2023 г.

1 Эксперимент: Зависимость числа итераций метода сопряженных градиентов от числа обусловленности и размерности пространства.

1.1 Описание эксперимента.

В данном эксперименте было необходимо исследовать зависимость числа итераций, необходимое методу сопряженных градиентов для сходимости, от двух параметров: 1) числа обусловленности оптимизируемой функции и 2) размерности пространства n оптимизируемых переменных.

Как и в Лабораторной 1, были рассмотрены значения размерности пространства n равные 10, 100, 1000 и 10000. Для каждого фиксированного n перебирались числа обусловленности k от 1 до 1000 с шагом 100 и строился график зависимости $T(k, n)$ от k . Так как каждый эксперимент необходимо было повторить несколько раз, то также генерировался массив random seed от 1 до 100 с шагом 10 и при повторении эксперимента выбиралось новое random seed.

В качестве матрицы A берется диагональная матрица $A = \text{Diag}(a)$ (создается через `scipy.sparse.diags`), у которой диагональные элементы сгенерированы в пределах $[1, k]$, причем $\min(a) = 1$, $\max(a) = k$ (1-й и 2-й

элементы сгенерированного массива вручную заменены на 1 и k , поэтому даже если данные числа не были сгенерированы случайно, то 1 и k все равно станут соответственно минимумом и максимумом). Вектор b заполняется случайными числами (`np.random.rand`).

1.2 Результаты.

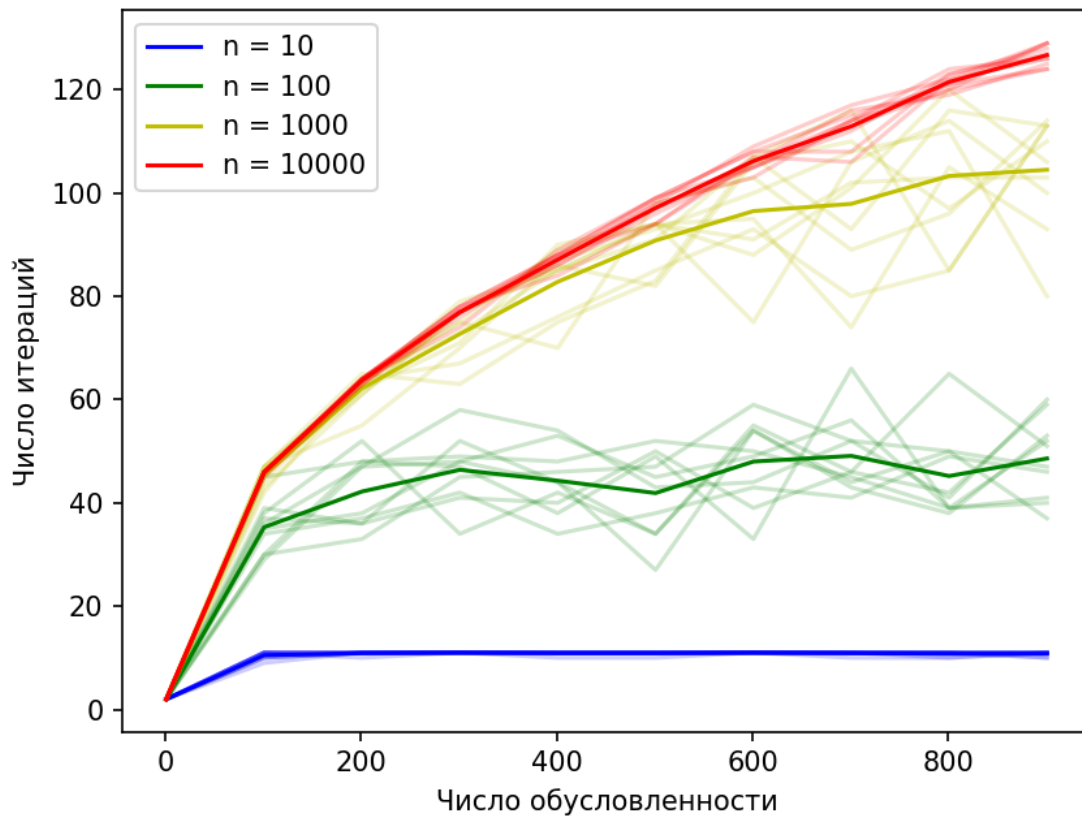


Рис. 1: Зависимость числа итераций от числа обусловленности и размерности пространства

По полученному графику можно сделать следующие выводы: чем выше размерность пространства, тем большее число итераций требуется методу сопряженных градиентов для сходимости, при этом полученное число итераций не превосходит значение размерности. Аналогично, чем выше число обусловленности, тем медленнее сходимость

Сравнение с методом градиентного спуска: при методе градиентного спуска наблюдалась линейная зависимость между числом итераций и числом обусловленности, в случае сопряженных градиентов число итераций возрастает медленнее.

2 Эксперимент: Выбор размера истории в методе L-BFGS.

2.1 Описание эксперимента.

Задачей было исследовать, как влияет размер истории в L-BFGS на поведение метода. Для этого были рассмотрены размеры истории l равные 0, 1, 5, 10, 50 и 100, а результаты расчетов для всех вариантов изображены на графиках.

2.2 Результаты.

Из графиков можно сделать вывод, что чем больше размер истории, тем быстрее работает L-BFGS. Однако, начиная с размера истории, равного 10 и выше, поведение метода уже почти не различается. Поэтому, брать размер истории слишком большим не имеет смысла, 10 - уже оптимально.

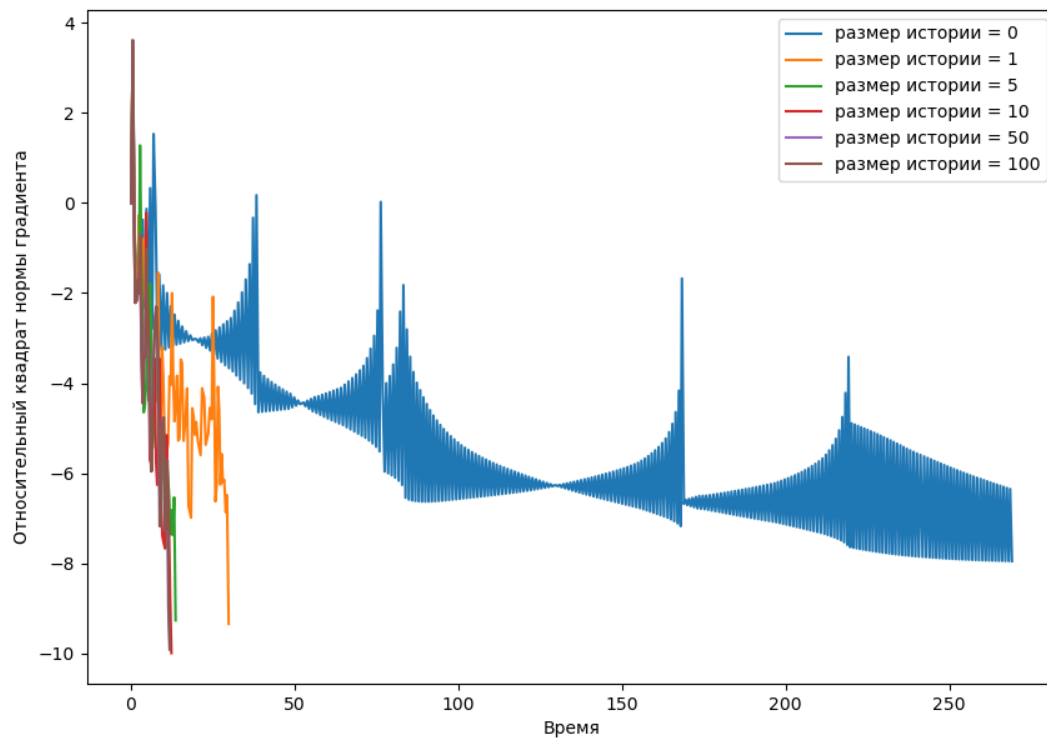


Рис. 2: Зависимость относительного квадрата нормы градиента от времени

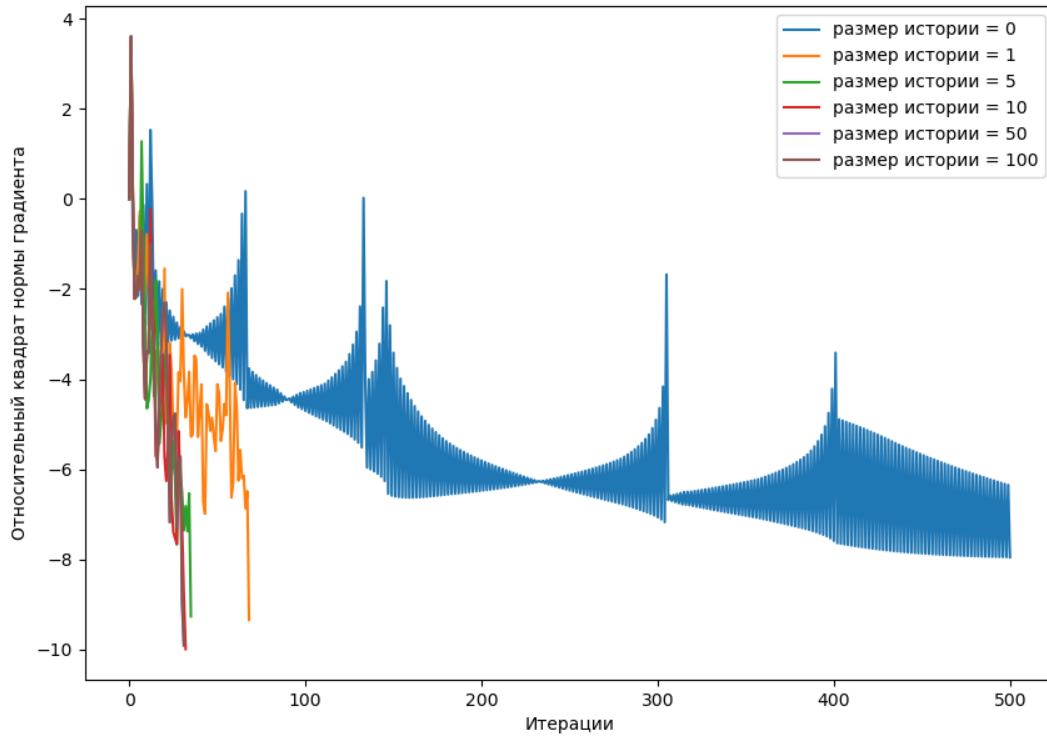


Рис. 3: Зависимость относительного квадрата нормы градиента от количества итераций

3 Эксперимент: Сравнение методов на реальной задаче логистической регрессии.

3.1 Описание эксперимента.

В данном эксперименте сравнивались усеченный метод Ньютона, метод L-BFGS и градиентный спуск на реальной задаче логистической регрессии. В качестве реальных данных использовались следующие наборы данных с сайта LIBSVM: w8a, gisette, real-sim, news20.binary, rcv1.binary. Коэффициент регуляризации взят равным $\lambda = 1/m$, начальная точка $x_0 = 0$. Для каждого датасета были построены 3 графика: а) Зависимость значения функции от номера итерации метода; (b) Зависимость значения функции

от реального времени работы; (с) Зависимость относительного квадрата нормы градиента (в логарифмической шкале) от реального времени работы.

3.2 Результаты.

Можно сделать вывод, что для всех наборов данных усеченный метод Ньютона работает быстрее всего. На втором месте по скорости L-BFGS, на некоторых датасетах, например gisette и real-sim, его результат почти аналогичен усеченному методу Ньютона. Градиентный спуск работает заметно медленнее.

Датасет w8a:

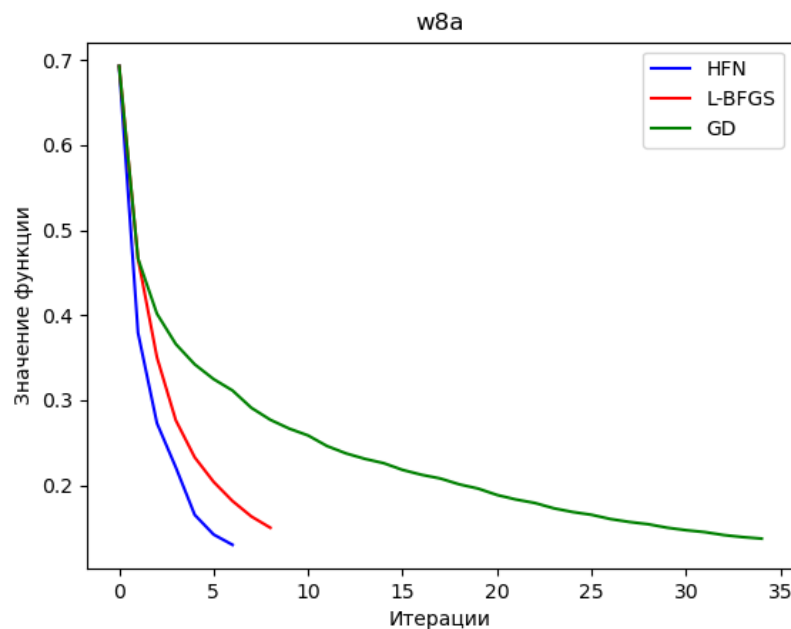


Рис. 4: Значение функции от итерации

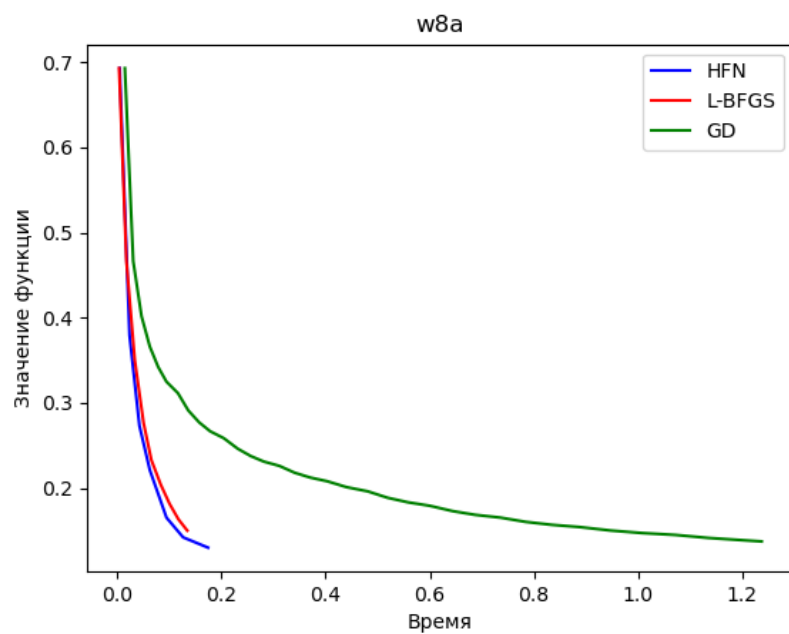


Рис. 5: Значение функции от времени работы

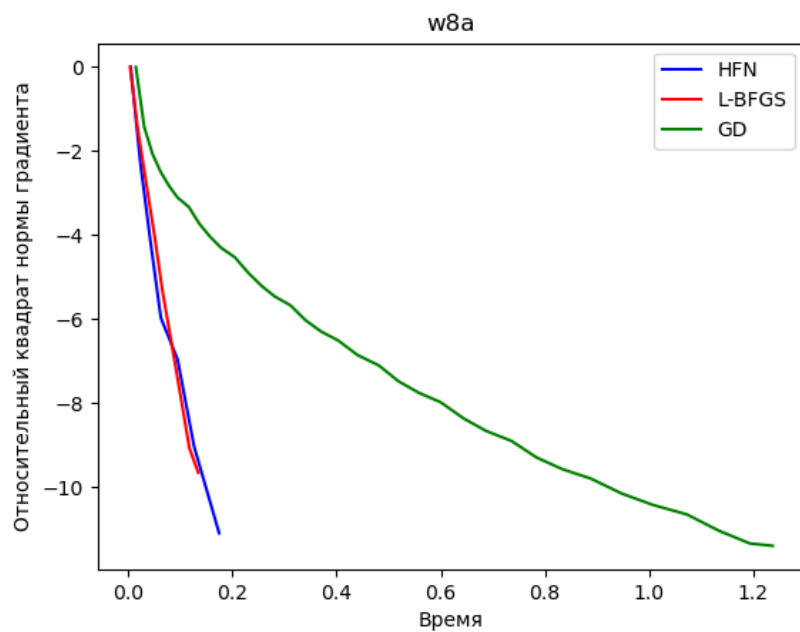


Рис. 6: Относительный квадрат нормы градиента от времени работы

Датасет gisette:

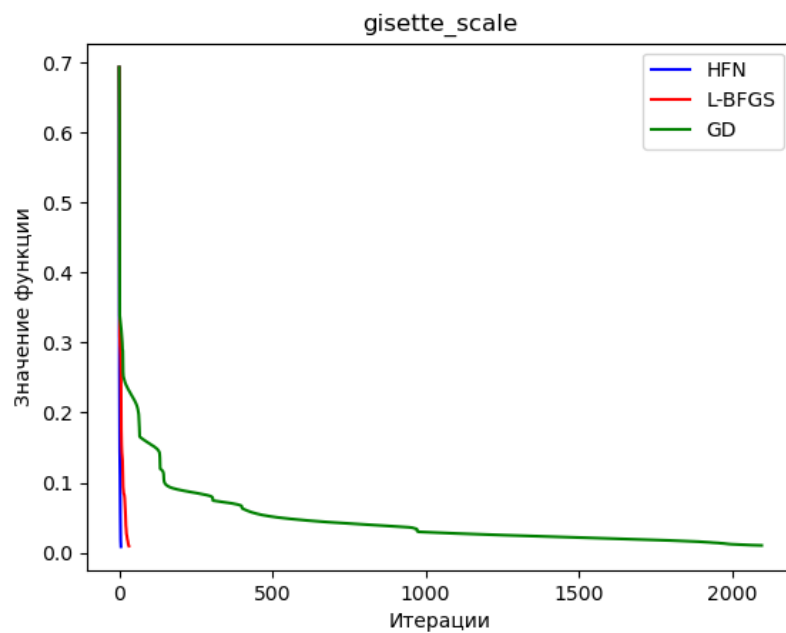


Рис. 7: Значение функции от итерации

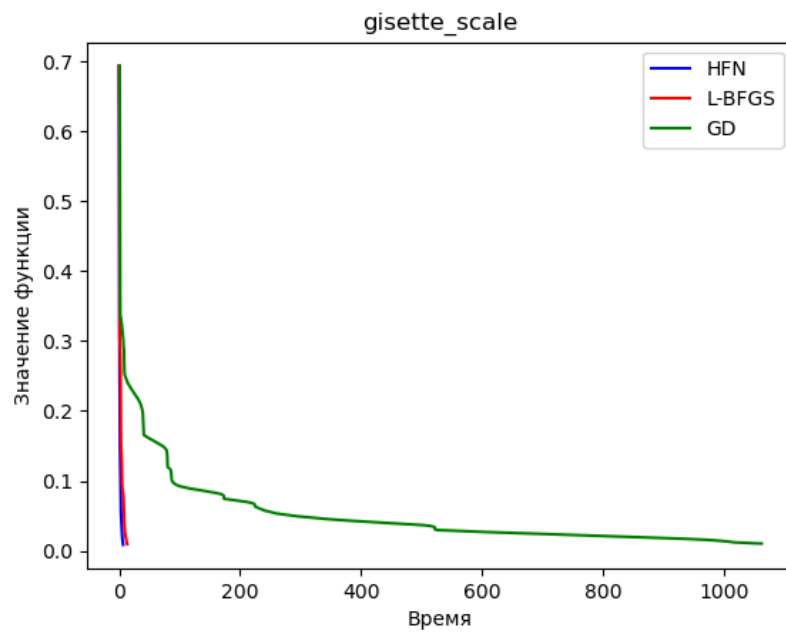


Рис. 8: Значение функции от времени работы

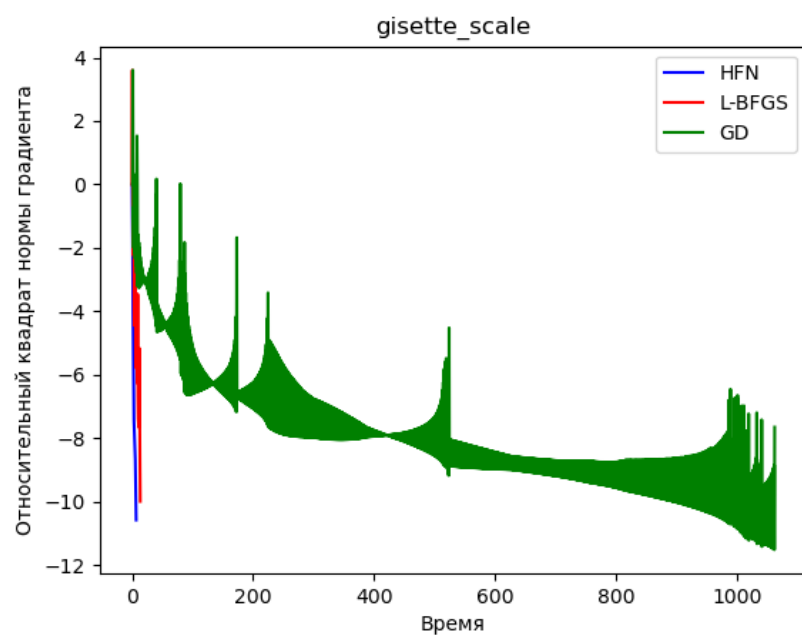


Рис. 9: Относительный квадрат нормы градиента от времени работы

Датасет real-sim:

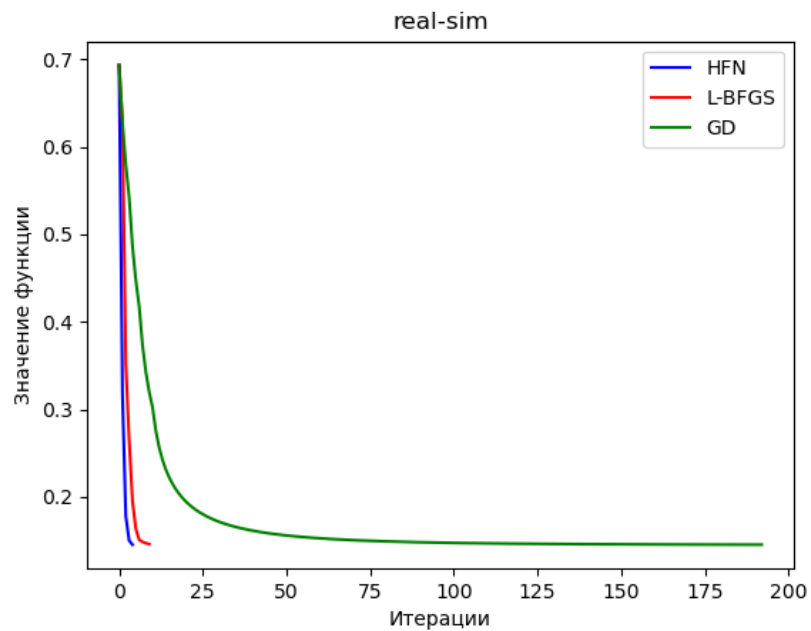


Рис. 10: Значение функции от итерации

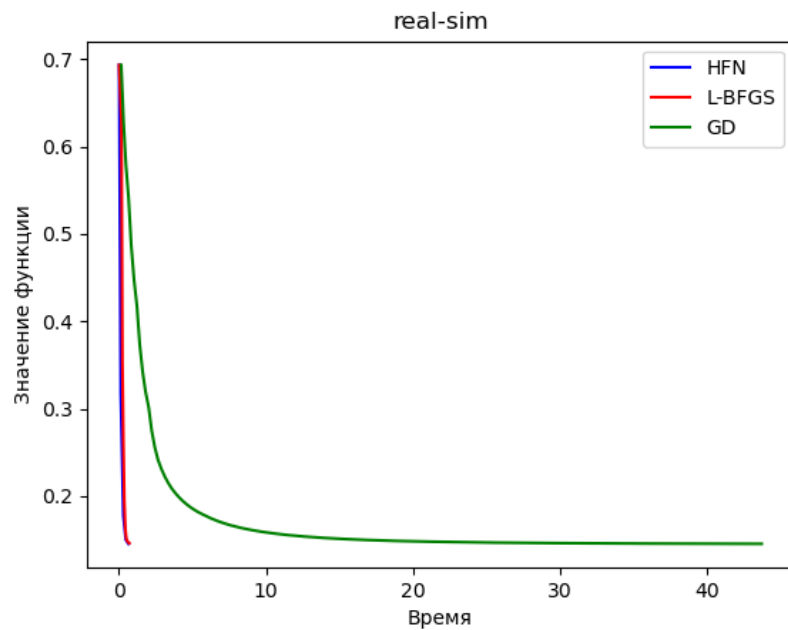


Рис. 11: Значение функции от времени работы

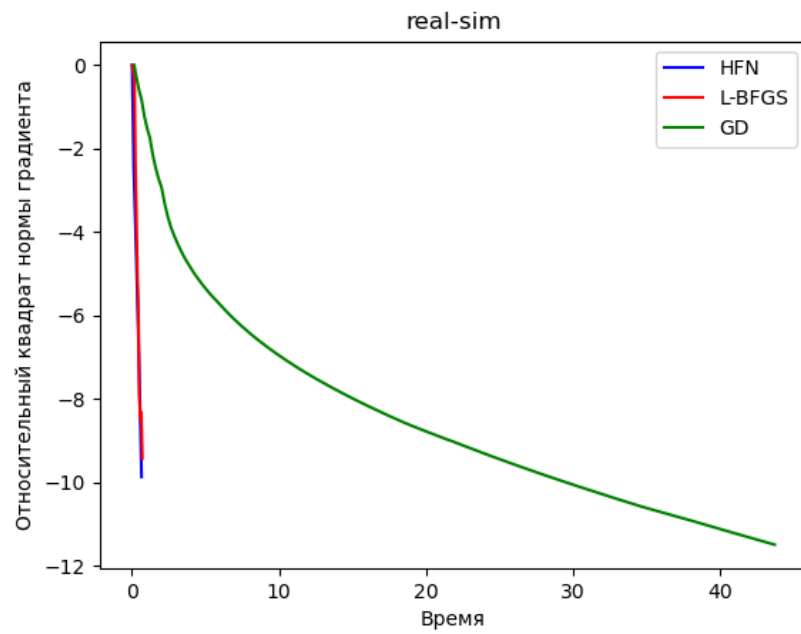


Рис. 12: Относительный квадрат нормы градиента от времени работы

Датасет news20.binary:

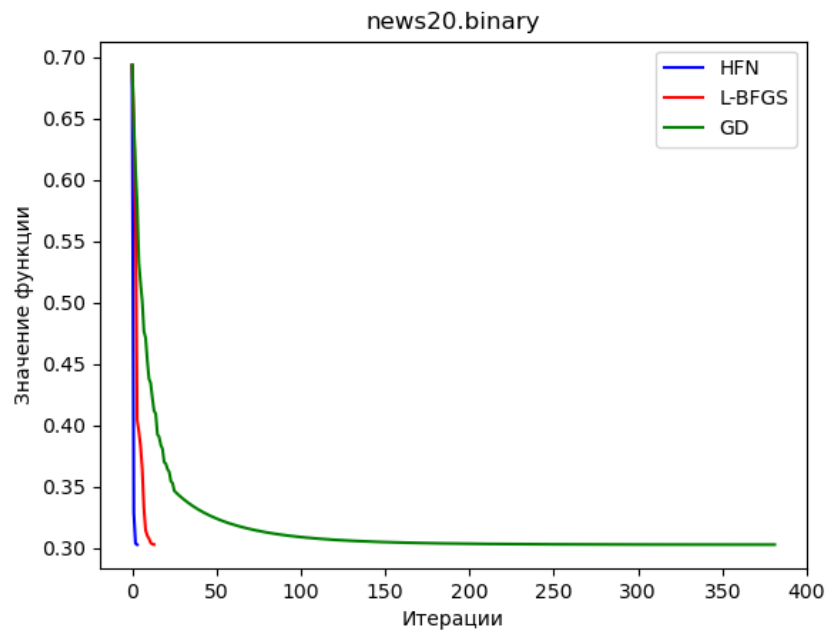


Рис. 13: Значение функции от итерации

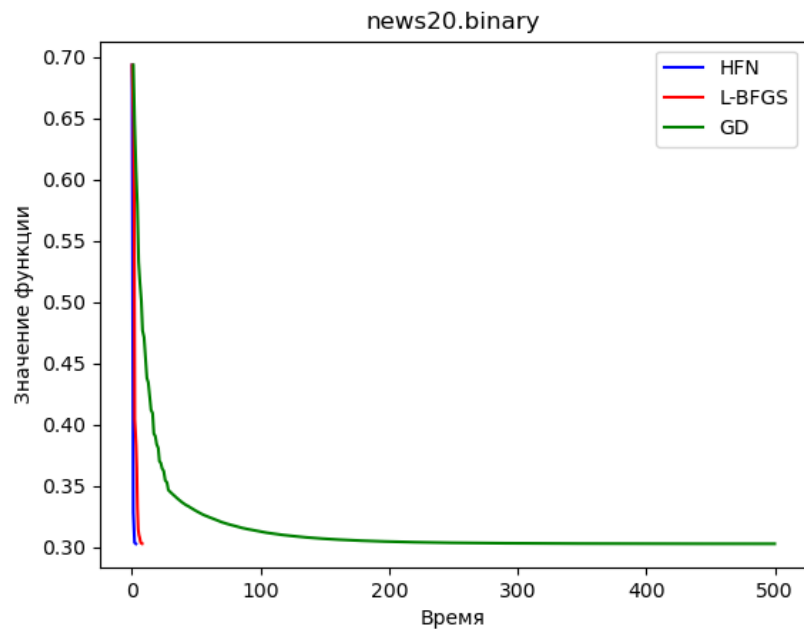


Рис. 14: Значение функции от времени работы

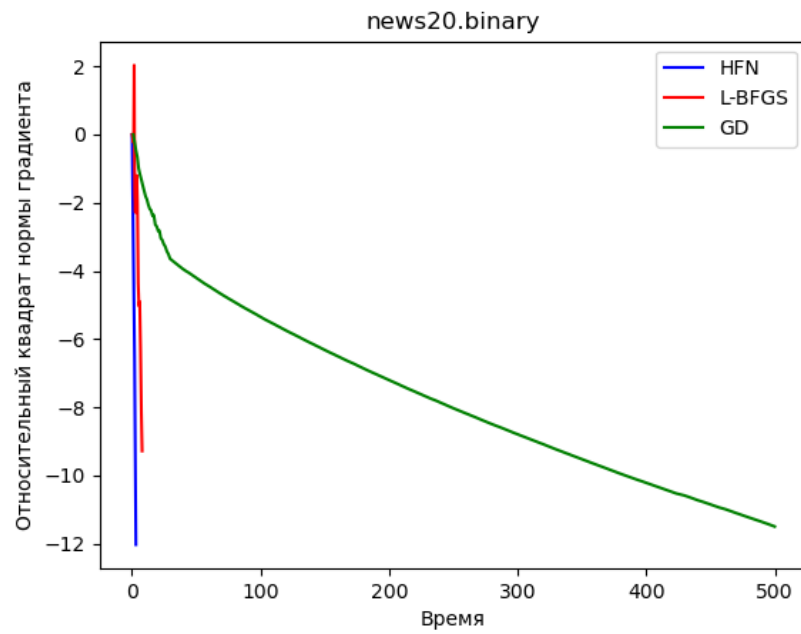


Рис. 15: Относительный квадрат нормы градиента от времени работы

Датасет rcv1.binary:

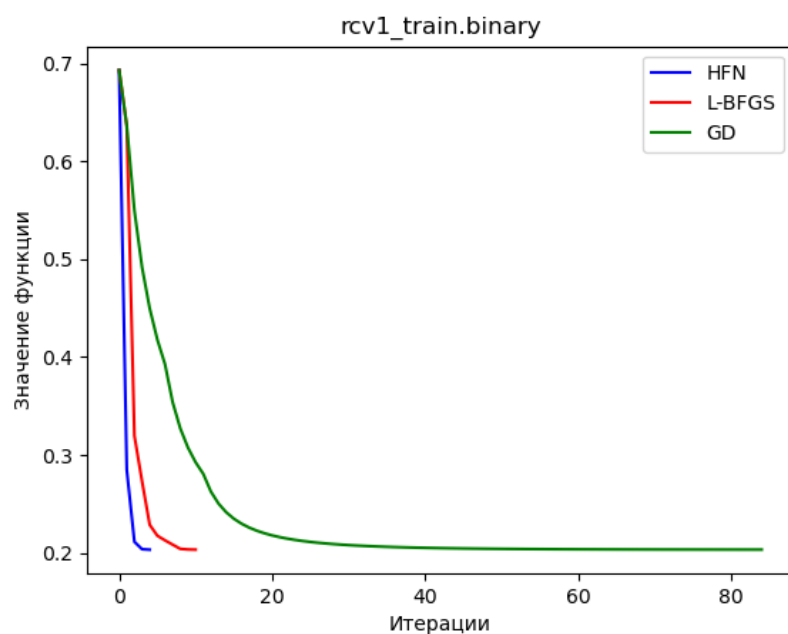


Рис. 16: Значение функции от итерации

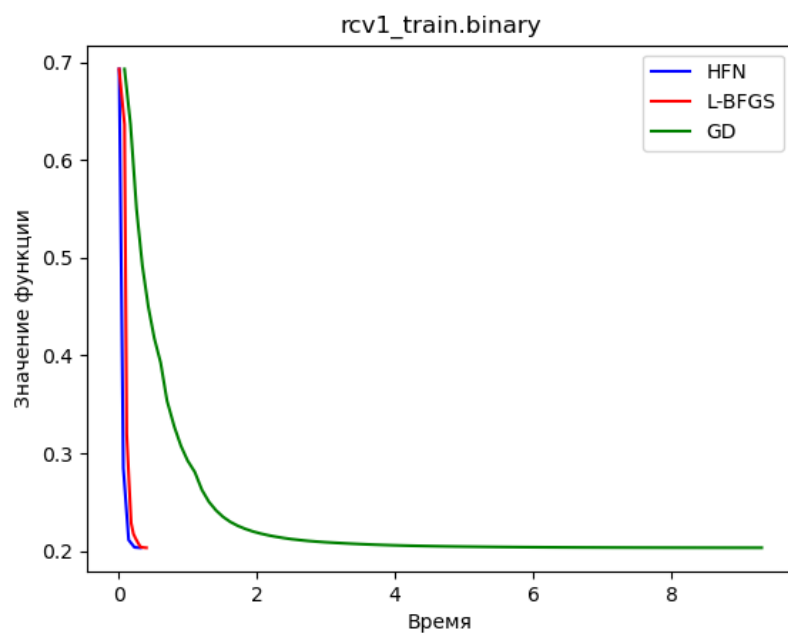


Рис. 17: Значение функции от времени работы

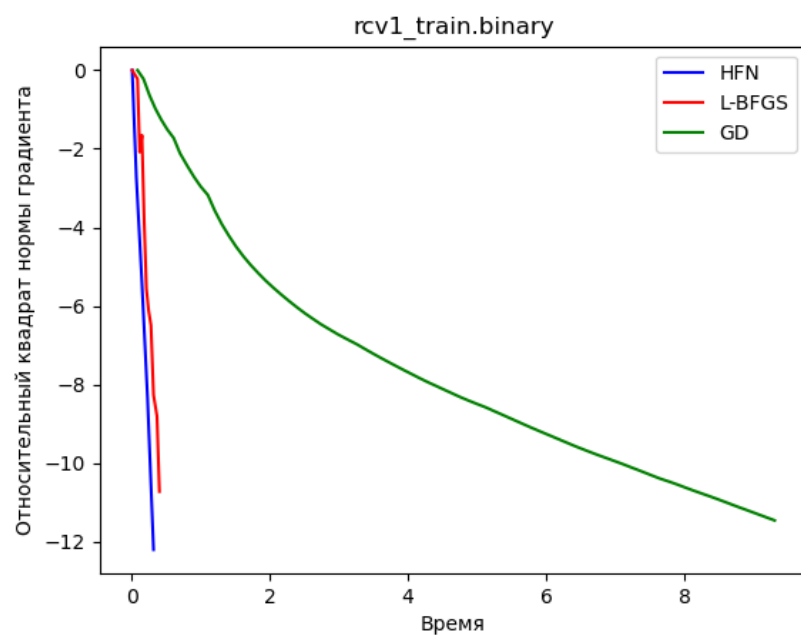


Рис. 18: Относительный квадрат нормы градиента от времени работы