

Лабораторная работа 2

на тему

Продвинутые методы безусловной оптимизации

Выполнила:
студентка 1 курса магистратуры
Шахвалиева Юлиана Сергеевна



(подпись)

ОГЛАВЛЕНИЕ

Разностная проверка произведения гессиана на вектор.....	3
Эксперимент 1: зависимость числа итераций метода сопряженных градиентов от числа обусловленности и размерности пространства.....	4
Эксперимент 2: выбор размера истории в методе L-BFGS	6
Эксперимент 3: сравнение методов на реальной задаче логистической регрессии	8

Разностная проверка произведения гессиана на вектор

В рамках данной работы были реализованы два метода, выполняющих умножение гессиана на заданный вектор: честный подсчет и разностная аппроксимация. Второй был реализован с целью проверки корректности реализации первого.

Вышеописанная проверка находится в файле `oracles.py` в функции `testing_hess`. Функция генерирует матрицу A и вектор b со случайными равномерно распределёнными значениями. Данная матрица и вектор выступают в качестве математической функции, на которой будет тестироваться расчет умножения гессиана на заданный вектор. Также случайно (из равномерного распределения) генерируются векторы v – на что впоследствии будет умножаться гессиан, а также x – точка, в которой будет происходить расчет гессиана.

$\nabla^2 f(x)v$ рассчитывается двумя способами, после чего результаты сравниваются. В итоге, реализованная функция `testing_hess` помогла подтвердить корректность реализации процедуры умножения гессиана на заданный вектор.

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

Целью данного эксперимента являлось исследование зависимости необходимого методу сопряженных градиентов для сходимости числа итераций от числа обусловленности оптимизируемой функции и размерности пространства оптимизируемых переменных. Также предполагается сравнение полученных результатов с результатами аналогичного эксперимента, проведенного для градиентного спуска.

Для проведения анализа были сгенерированы квадратичные задачи размера n с числом обусловленности k из равномерного распределения $U(1, 20)$, на них был запущен метод сопряженных градиентов 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 сопряженных градиентов 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.

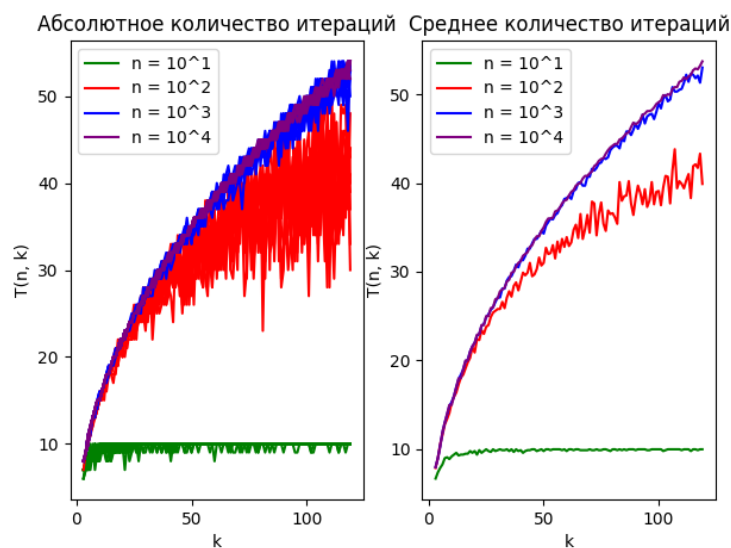


Рисунок 1.1 Абсолютная и средняя зависимость количества итераций метода сопряженных градиентов от числа обусловленности функции.

Из графика видно, что при $n = 10$, количество итераций, необходимое для сходимости, не зависит от числа обусловленности. При $n > 10$ наблюдается прямая зависимость. Также можно сделать вывод, что при увеличении n , удельный рост количества итераций при увеличении числа обусловленности, замедляется.

Для сравнения полученных результатов с результатами аналогичного эксперимента, проведенного для градиентного спуска, прикрепляю графики, полученные для последнего.

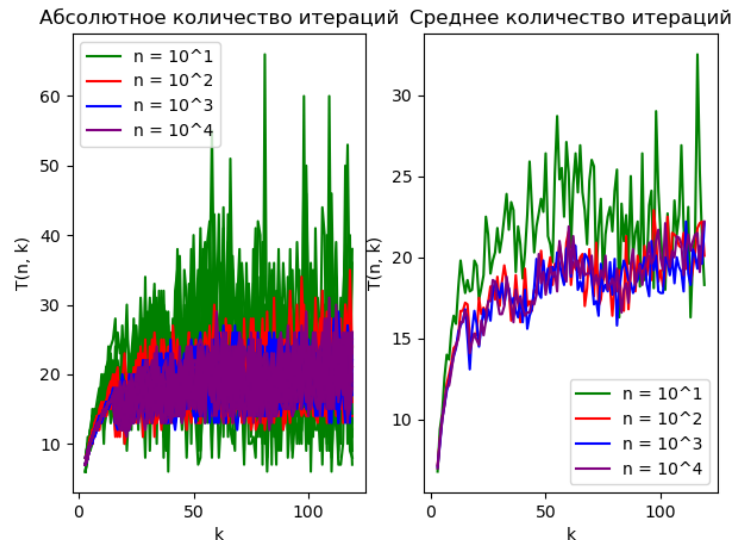


Рисунок 1.2 Абсолютная и средняя зависимость количества итераций метода градиентного спуска от числа обусловленности функции.

Видно, что при $n > 10$ графики абсолютной и средней зависимости количества итераций метода сопряженных градиентов более крутые, чем аналогичные для метода градиентного спуска. То есть при увеличении числа обусловленности функции количество необходимых для сходимости итерации растет сильно быстрее в методе сопряженных градиентов, чем в градиентном спуске.

Исключение составляет случай, когда размерность пространства маленькая, то есть $n = 10$. Тогда методу сопряженных градиентов требуется меньшее количество итераций для сходимости при любом значении числа обусловленности функции, чем градиентному спуску.

Таким образом, можно сделать вывод, что для пространства низкой размерности лучше использовать метод сопряженных градиентов, для остальных – градиентный спуск.

Эксперимент 2: выбор размера истории в методе L-BFGS

Целью данного эксперимента являлось исследование влияния размера истории в методе L-BFGS на его поведение, то есть на скорость сходимости. В качестве тестовой функции была взята логистическая регрессия с l^2 -регуляризатором на данных news20.binary с сайта LIBSVM.

Было построено 6 логистических регрессий с различными размерами истории в методе L-BFGS: 0, 1, 5, 10, 50, 100. Все остальные параметры метода были равными параметрам по умолчанию, начальная точка $x_0 = 0$, коэффициент регуляризации $\lambda = \frac{1}{m}$.

Далее были построены графики зависимости относительного квадрата нормы градиента против номера итерации, а также против реального времени работы.

Полученный результат представлен на рисунке ниже. Слева – график зависимости относительного квадрата нормы градиента против номера итерации, справа - против реального времени работы. На рисунке цвет отвечает за размер истории.

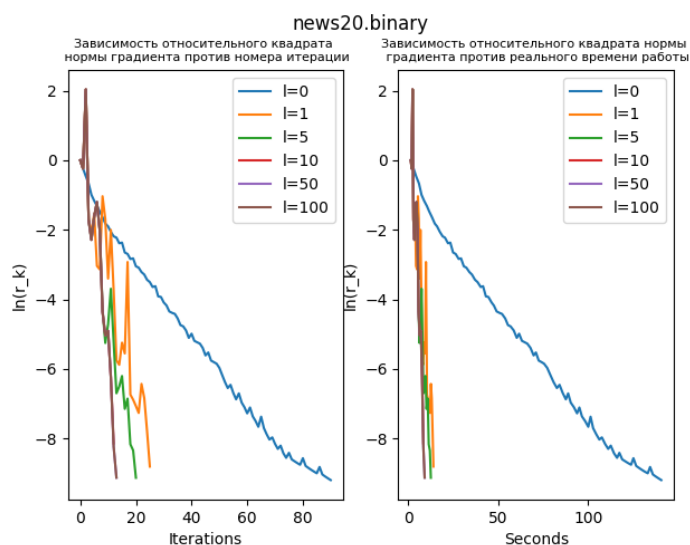


Рисунок 2.1 График зависимости относительного квадрата нормы градиента

На предоставленных графиках видно, что в случае, когда в методе L-BFGS не сохраняется история, он сходится очень медленно: в разы медленнее, чем при наличии истории, даже если длина истории равна единице.

В целом, можно сделать вывод, что при увеличении размера истории, скорость сходимости увеличивается. При этом удельная скорость этого увеличения падает, то есть

разница в скорости сходимости между $l = 5$ и $l = 100$ не такая сильная, как между $l = 0$ и $l = 1$.

Эксперимент 3: сравнение методов на реальной задаче логистической регрессии

Целью данного эксперимента являлось сравнение усеченного метода Ньютона, метода L-BFGS и градиентного спуска на задаче обучения логистической регрессии на реальных данных. В качестве реальных данных были использованы следующие наборы с сайта LIBSVM: w8a, gisette, real-sim, news20.binary, rcv1.binary.

Для каждого набора данных были построены 3 логистические регрессии, которые были обучены с помощью методов, перечисленных выше. Параметры всех методов были равными параметрам по умолчанию, начальная точка – $x_0 = 0$, коэффициент регуляризации $\lambda = \frac{1}{m}$.

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

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

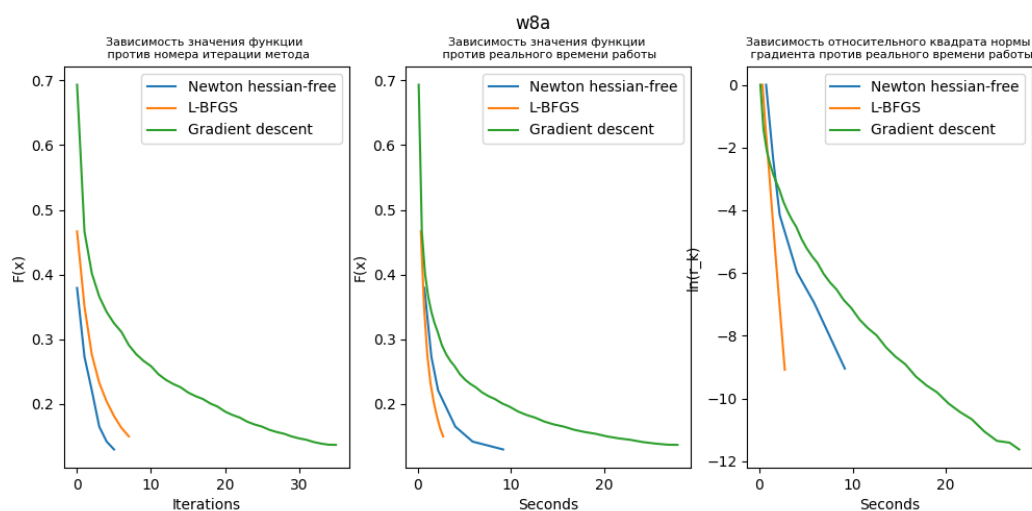


Рисунок 3.1 Графики сходимости для w8a

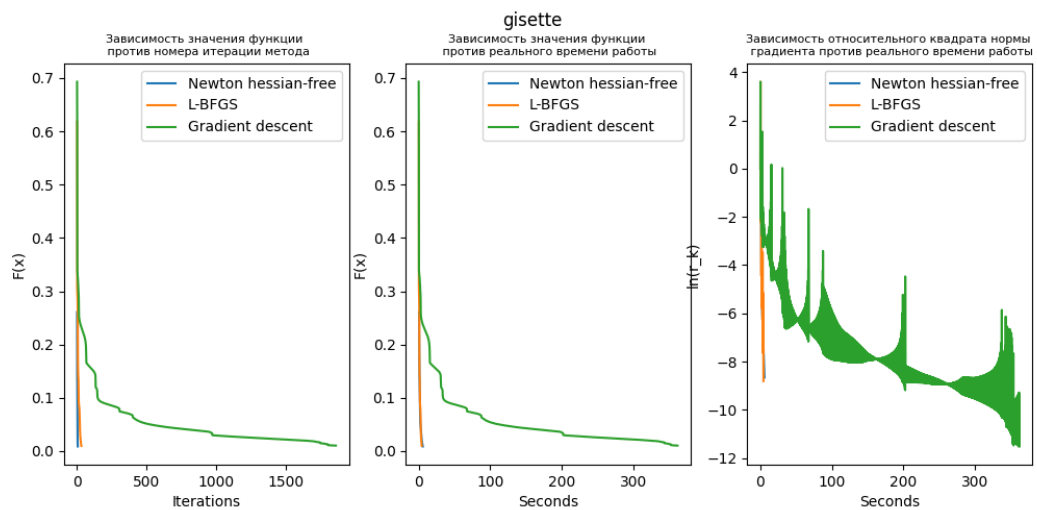


Рисунок 3.2 Графики сходимости для gisette

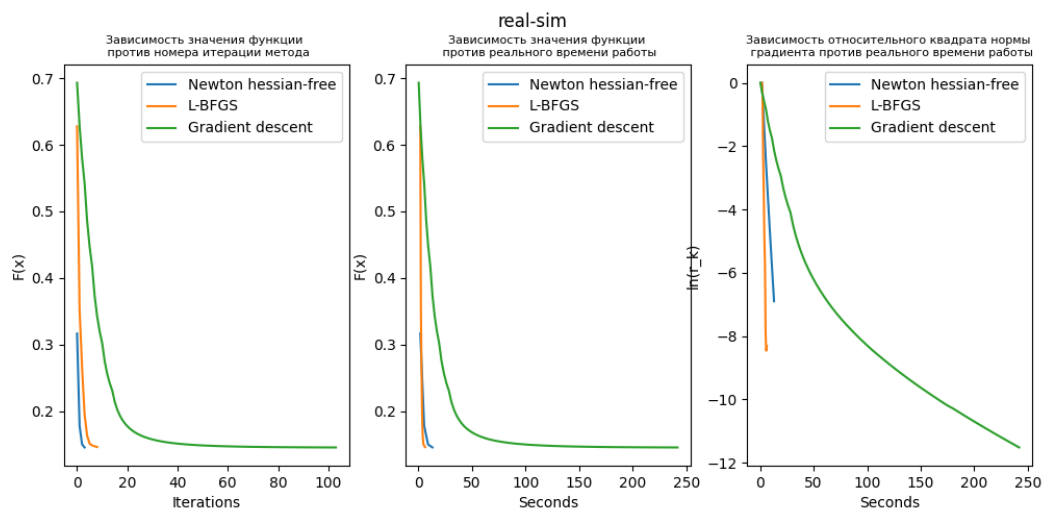


Рисунок 3.3 Графики сходимости для real-sim

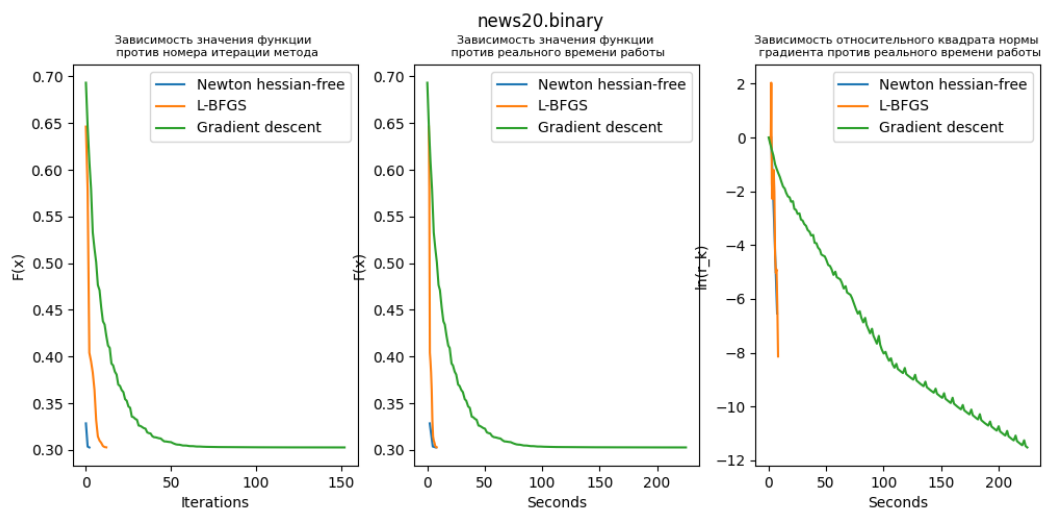


Рисунок 3.4 Графики сходимости для news20.binary

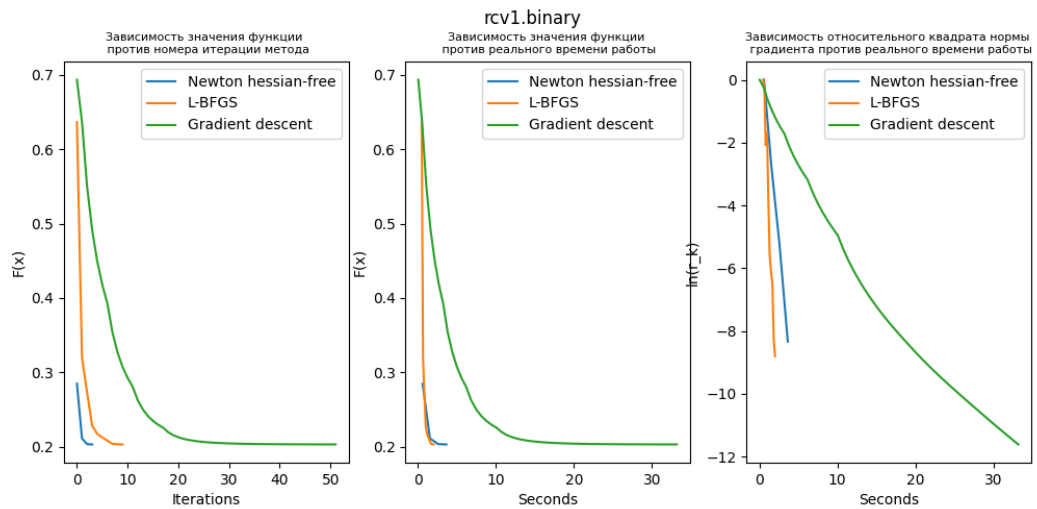


Рисунок 3.5 Графики сходимости для `rsv1.binary`

Как видно из графиков для всех наборов данных градиентный спуск сходится сильно медленнее двух остальных методов. При чем разница колоссальна: для сходимости градиентному спуску нужно в несколько раз больше времени/итераций. Усеченный метод Ньютона и метода L-BFGS показали схожие результаты: они достаточно быстро сошлись на всех рассмотренных наборах данных.