



МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М. В. ЛОМОНОСОВА
Факультет вычислительной математики и кибернетики
Кафедра математических методов прогнозирования

Отчет ко второму практическому заданию по МОМО: Продвинутые методы безусловной оптимизации

Студент 517 группы:
Оспанов А.М.

Москва, 2016

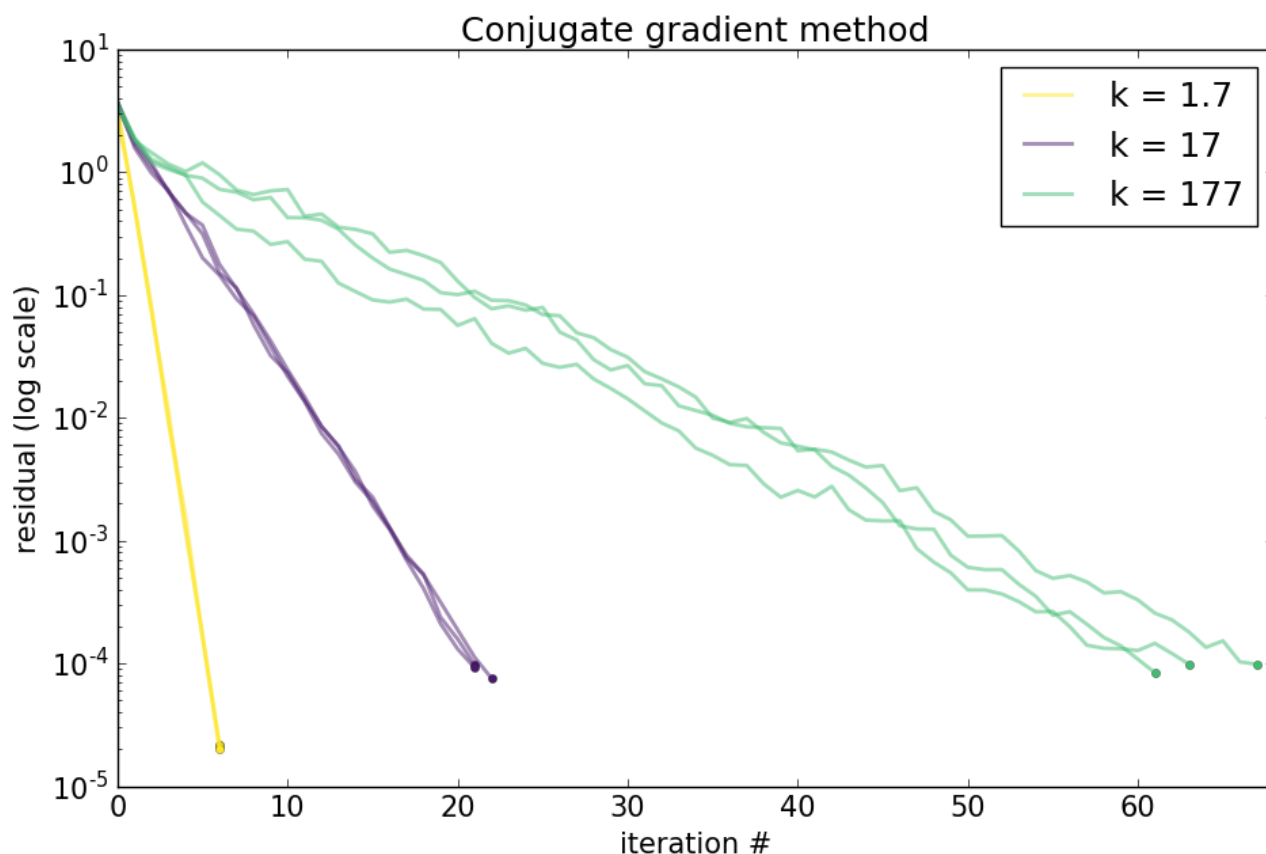
1 Введение

В данной работе будут реализованы модули `optim.py`, `lossfuncs.py`, `special.py` с методами безусловной оптимизации и функциями потерь для логистической регрессии: метод сопряженных градиентов для решения системы линейных уравнений, нелинейный метод сопряженных градиентов, метод LBFGS и неточный метод Ньютона. Будут проведены исследования скорости сходимости метода сопряженных градиентов (CG) для различных чисел обусловленности матрицы системы. Также нелинейный метод сопряженных градиентов (NCG), метод LBFGS и неточный метод Ньютона (HFN) будут сравниваться между собой по невязке от вызова оракула и от времени работы циклов.

Код написан на языке Python 3 с использованием библиотеки `numpy`.

2 Скорость сходимости метода сопряженных градиентов в зависимости от числа обусловленности κ матрицы системы

Скорость сходимости метода будут исследоваться при $\kappa \in \{1.7, 17, 177\}$



На графике видно, что чем больше κ , тем меньше скорость сходимости. Это объясняется особенностью метода CG: чем меньше различных собственных значений или чем меньше кластеров образуют собственные значения, тем быстрее метод сходится. Также очевидно, что чем меньше число обусловленности, тем меньше кластеров и/или меньше различных

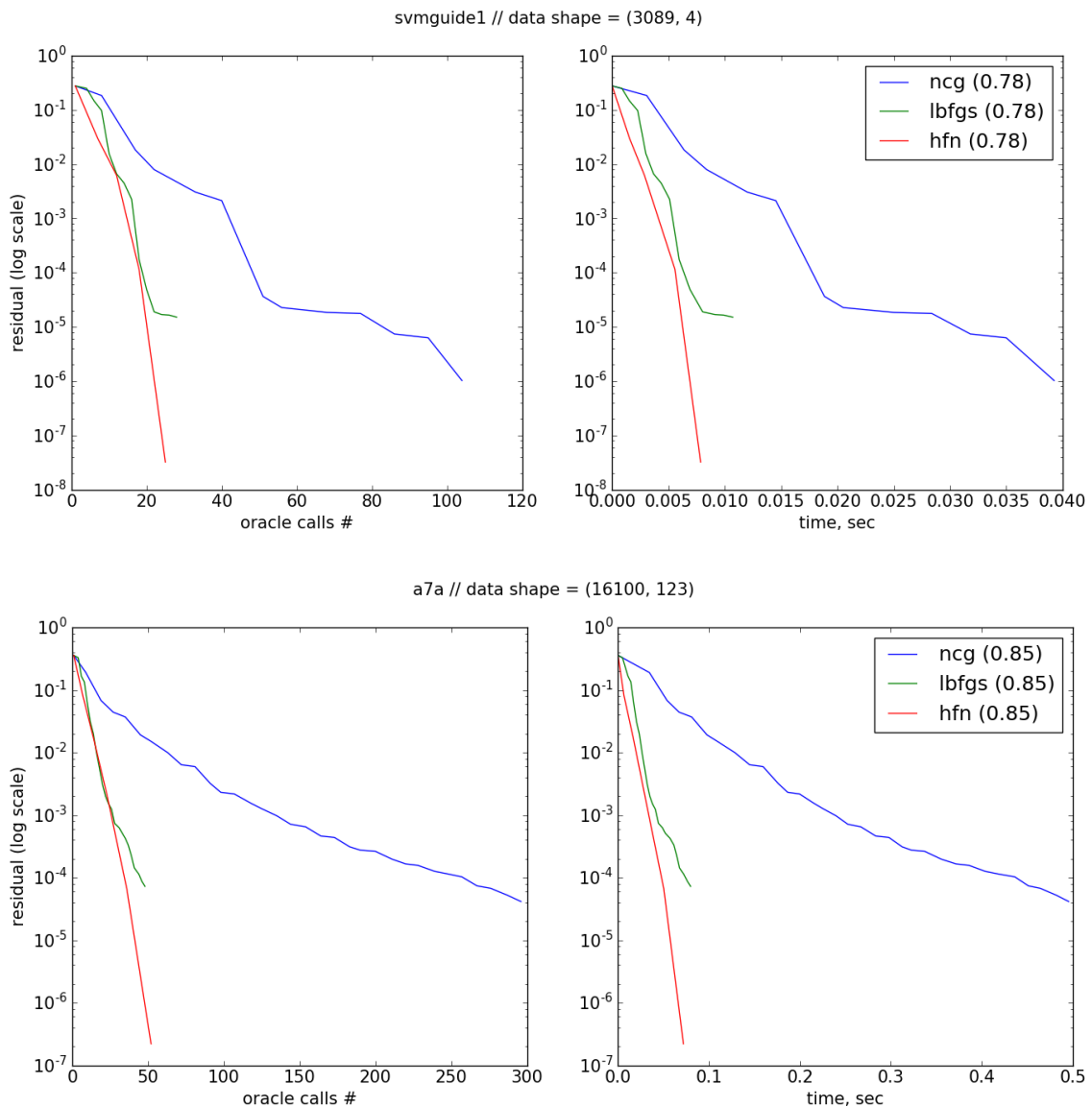
собственных значений. Т.о. скорость сходимости метода обратно пропорциональна числу обусловленности. Также стоит отметить, что скорость сходимости метода линейна.

3 Сравнение методов NCG, LBFGS и HFN

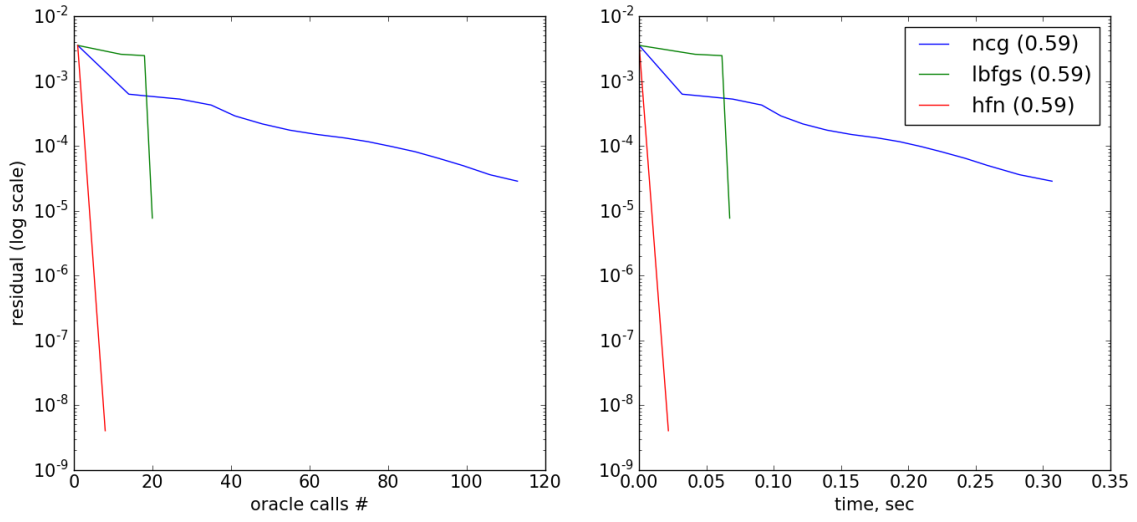
Сравниваться три реализованных метода будут на задачах обучения двухклассовой логистической регрессии на реальных данных. В качестве данных были выбраны:

Data set	(data size, feature size)
svmguide1	(3089, 4)
a7a	(16100, 123)
madelon	(2000, 500)
gisette scale	(6000, 5000)
leu	(38, 7129)
rcv1	(20242, 47236)

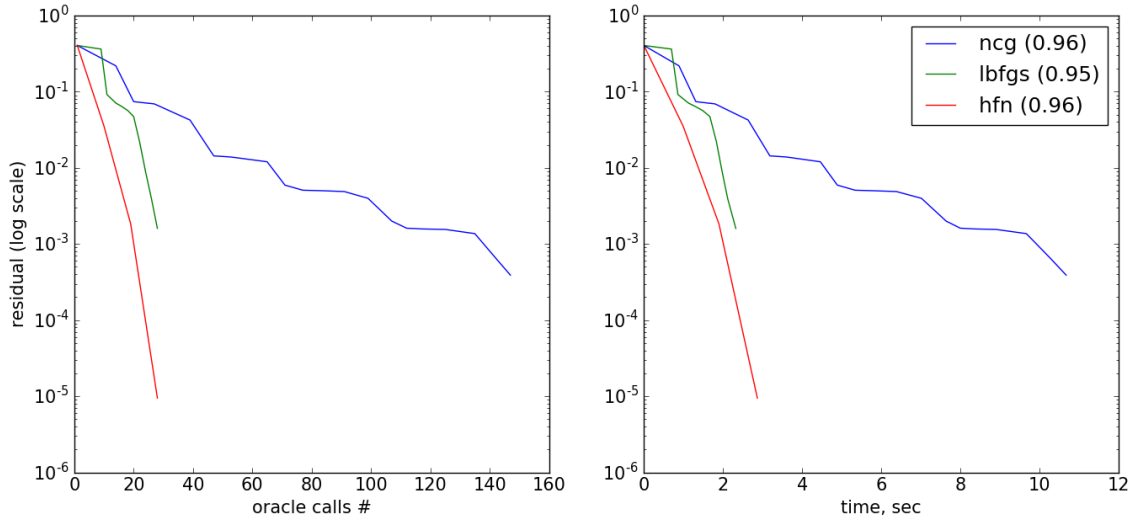
Приведем все графики скоростей сходимости



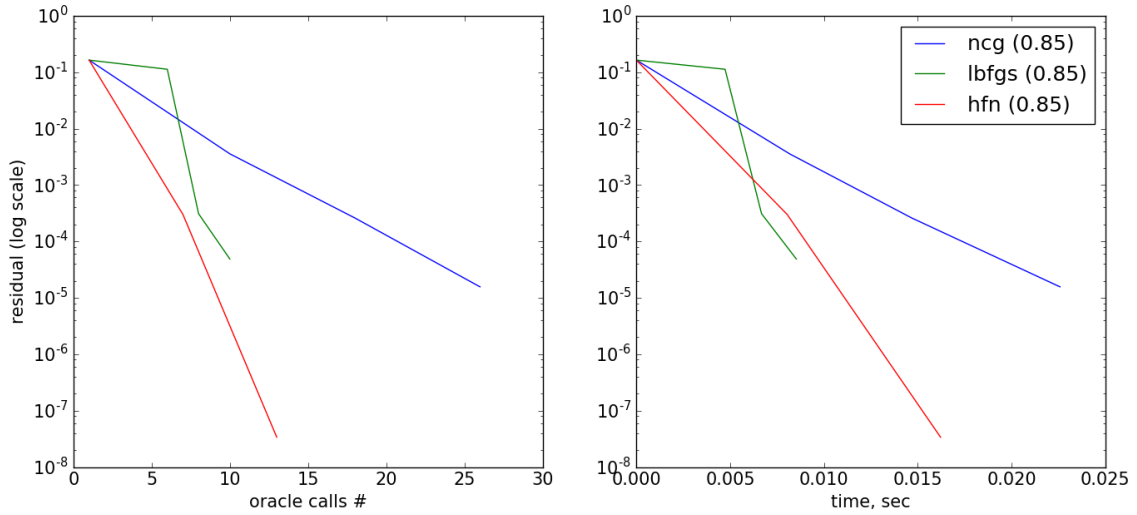
madelon // data shape = (2000, 500)

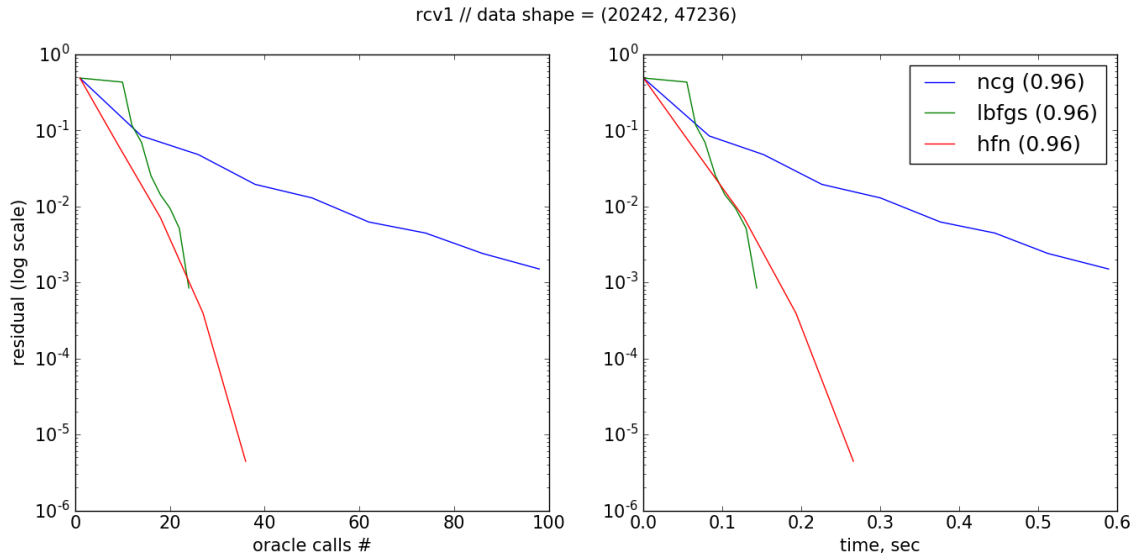


gisette_scale // data shape = (6000, 5000)

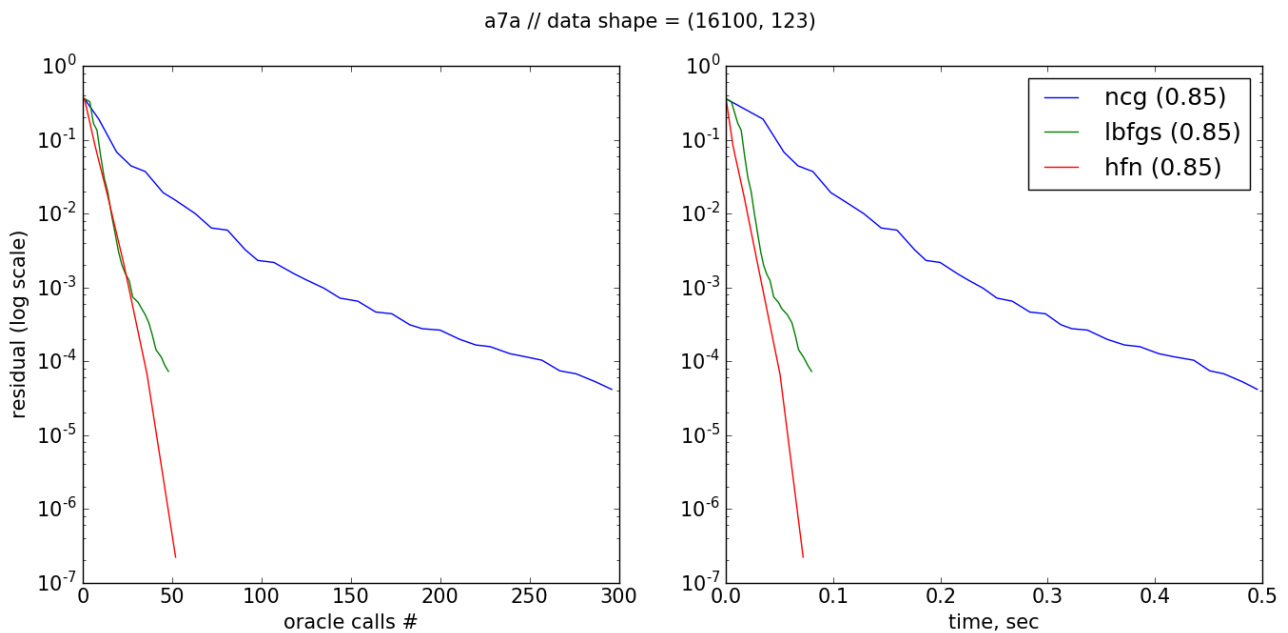


leu // data shape = (38, 7129)

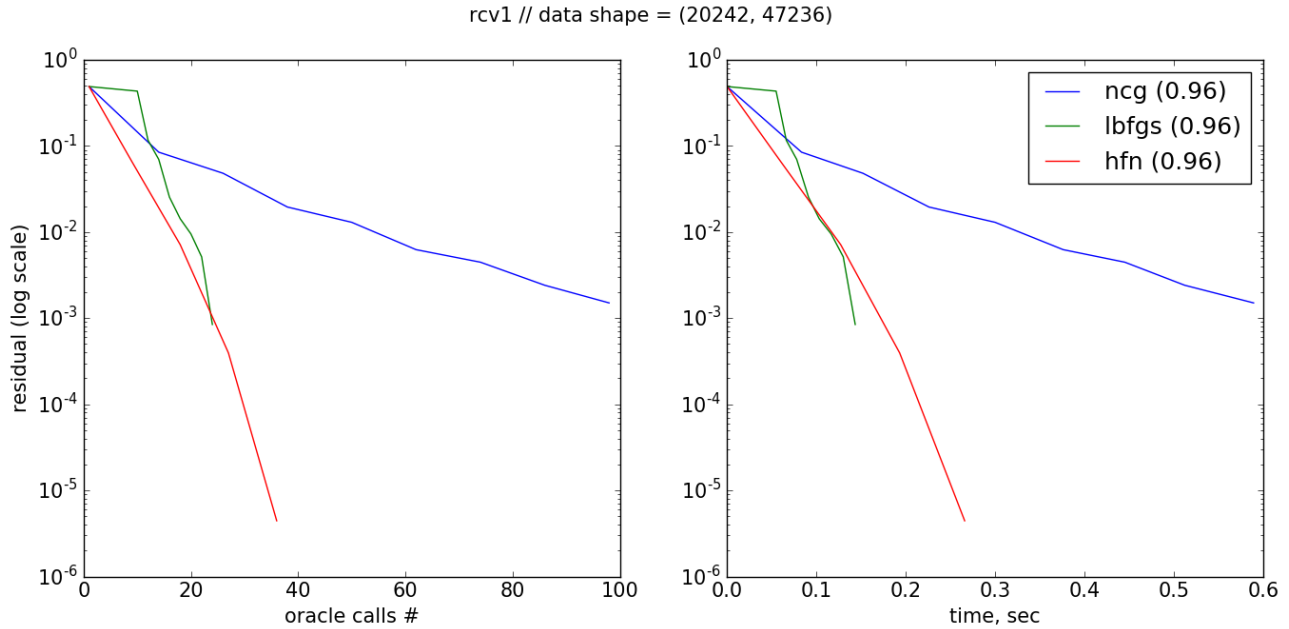




Из графиков видно, что метод NCG имеет сублинейную скорость сходимости, а HFN и LBFGS - суперлинейные скорости сходимости. В целом видно, что соотношения скоростей сходимости методов не меняются при увеличении количества признаков. Поэтому подробно рассмотрим только данные с маленьким и большим количеством признаков.



Из графика понятно, что методы HFN и LBFGS на данных с маленьким количеством признаков работают одинаково и по количеству вызовов оракула и по времени работы. Но HFN дает большую точность. А метод NCG работает дольше в ≈ 5 раз, но при этом имеет невязку примерно как у LBFGS.



Как было замечено раньше, соотношения скоростей и точностей сохраняются. Соответственно, метод HFN является быстрым и самым точным. Метод LBFGS сходится быстрее всех, но менее точно. И метод NCG также выдает невязку примерно равную с LBFGS, но сходится дольше и также примерно в 5 раз.

Из двух типов данных можно понять, что строгой зависимости между размерностью данных и скоростью работы методов не видно. Данные могут быть большой размерности, но сходиться быстрее, чем данные с меньшей размерностью. Скорость сходимости, скорее всего, зависит от природы данных.

4 Заключение

В работе сравнивались методы безусловной оптимизации. Были исследованы их скорости сходимости. В результате можно сказать:

1. Чем число обусловленности матрицы системы меньше, тем быстрее сходится метод CG;
2. Метод HFN является достаточно быстрым и самым точным, в то время как NCG является самым не точным и самым медленным;
3. Нет особой разницы в работе методов с данными большой и маленькой размерностей.