

Национальный исследовательский университет ИТМО
Факультет информационных технологий и программирования
Прикладная математика и информатика

Методы оптимизации

Отчет по лабораторной работе №3

Работу выполнили:

Захаров Кирилл М32391

Мавлютов Эрвин М32391

Шилкин Артем М32391

Преподаватель:

Шохов Максим Евгеньевич

г. Санкт-Петербург

2023 г.

Оглавление

<i>Метод Гаусса-Ньютона</i>	3
<i>Метод Dogleg</i>	4
<i>Метод BFGS</i>	6
<i>Дополнительное задание. Метод L-BFGS.</i>	7
<i>Приложение 1.</i>	17

Постановка задач:

1. Реализация методов Gauss-Newton и Powell Dog Leg для решения нелинейной регрессии. Сравнение эффективности с ранее реализованными методами.
2. Реализация метода BFGS и его исследование его сходимости при минимизации различных функций.
3. Реализация и исследование метода L-BFGS.

Весь проект можно найти по [ссылке](#).

Метод Гаусса-Ньютона

Если у нас есть функция $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, то можно поставить задачу минимизации её модуля

$$S(x) = \frac{1}{2} \|f(x)\|^2 = \frac{1}{2} \sum_{i=1}^m f_i^2(x)$$

называемую также задачей минимизации квадратов. Метод Гаусса-Ньютона ищет точку, в которой градиент функции $S(x)$ равняется нулю путем решения той же задачи, но с линейными приближениями функций $f_i(x)$.

Подробнее, $\nabla S(x) = J^T(x)f(x)$, где $J_{ij}(x) = \frac{\partial f_i(x)}{\partial x_j}$. А

$$\nabla^2 S(x) = J^T(x)J(x) + \sum_{i=1}^m \nabla^2 f_i(x) \cdot f_i(x)$$

Идея метода заключается в том, что можно считать, что

$J^T(x)J(x) \gg \sum_{i=1}^m \nabla^2 f_i(x) \cdot f_i(x)$, т.е. $\nabla^2 S(x) \approx J^T(x)J(x)$, а тогда можно

модифицировать метод касательных $x_{i+1} = x_i - (\nabla^2 S(x))^{-1} f(x)$, не считать гессиан $S(x)$, и получить следующую формулу для пересчета последовательности точек:

$$x_{i+1} = x_i - (J^T(x)J(x))^{-1} J^T(x)f(x).$$

Метод Dogleg

Метод доверительного региона заключается в аппроксимации функции квадратичной моделью:

$$m_k(x) = f(x_k) + (g_k, x) + \frac{1}{2}x^T B_k x,$$

где $g_k = \nabla f(x_k)$, $B_k = \nabla^2 f(x_k)$ в расчете на то, что в некоторой области $\|x - x_k\| \leq \Delta$ (доверительном регионе) она достаточно точно описывает минимизируемую функцию. На каждом шаге мы смотрим на то, насколько хороша наша модель (насколько предсказываемое уменьшение сравнимо с действительным) и увеличиваем или уменьшаем регион в котором минимум считается достоверным. Минимум модели же можно считать аналитически, приравняв $\nabla m_k(x)$ к нулю ($g_k + B_k p = 0$, $p = -B_k^{-1} g_k$, $p = x_{\min} - x$), при этом если новый минимум оказывается в доверительном регионе, он считается новой точкой отсчета, в противном случае возможны варианты выбора x_{k+1} . Для корректной работы алгоритма требуется положительная определенность B_k . Так как в общем случае гессиан может не являться таковым, будем прибавлять достаточно большую диагональную матрицу, чтобы это починить.

Dogleg является модификацией метода доверительного региона, а именно, он определяет поведение алгоритма в случае, когда $\|p^B\| > \Delta$, $p^B = x_{\min} - x_k$. В этом случае мы считаем минимум вдоль

направления антиградиента $p^U = -\frac{g_k^T g_k}{g_k^T B_k g_k} g_k$, если $\|p^U\| \geq \Delta$, то

$x_{k+1} = x_k + \frac{\Delta}{\|p^U\|} p^U$, иначе выбирается τ , такое, что $\|p^U + \tau(p^B - p^U)\| = \Delta$ и $x_{k+1} = p^U + \tau p^B$. Мы реализовали следующую стратегию пересчета Δ :

$$\Delta_{k+1} = \begin{cases} \frac{1}{4}\Delta_k, \rho < \frac{1}{4} \\ \min(2\Delta_k, \Delta_{\max}), \rho > \frac{3}{4} \& \|x_{k+1} - x_k\| = \Delta_k \\ \Delta_k, \text{ else} \end{cases}$$

Метод BFGS

В отличие от предыдущих методов BFGS не вычисляет гессиан напрямую для вычисления минимума функции. Вместо этого он вычисляет его приближенно, поэтому данный метод относят к классу квазиньютоновских методов.

Пусть нам дана функция $f(x)$ и мы решаем задачу поиска минимума данной функции, где $f(x)$ не является выпуклой функцией и имеет непрерывные вторые производные. Воспользуемся идеей метода DogLeg: выберем произвольную точку x_0 , задаем точность поиска H_0 . Для простоты мы считали ее в начальной точке равной единичной матрице. Берем направление, вдоль которого будем искать точку (аналогично методу DogLeg): $p_k = -H_k \cdot \nabla f_k$.

Вычисляем следующую точку по формуле:

$x_{k+1} = x_k + \alpha_k \cdot p_k$, где α_k - минимум функции вдоль заданного направления. α_k можно рассчитать используя сильные условия Вольфе, мы же в своей работе использовали метод дихотомии.

Вся магия этого метода заключается в пересчете матрицы H_k . Сначала определим вектора:

$$\begin{aligned} s_k &= x_{k+1} - x_k \\ y_k &= \nabla f_{k+1} - \nabla f_k \end{aligned}$$

Таким образом, s_k - шаг алгоритма на k -ой итерации, y_k - изменение градиента на k -ой итерации.

Обновляем Гессиан функции на k -ой итерации с помощью такой формулы:

$$H_{k+1} = (I - \rho_k \cdot s_k \cdot y_k^T) H_k (I - \rho_k \cdot y_k \cdot s_k^T) + \rho_k \cdot s_k \cdot s_k^T,$$

где

$$\rho_k = \frac{1}{y_k^T \cdot s_k}$$

Дополнительное задание. Метод L-BFGS.

Как можно заметить в формулах BFGS мы постоянно поддерживаем текущий гессиан функции, что занимает квадрат памяти от размерности пространства. Также при пересчете Гессиана мы вынуждены создать две новые матрицы, что опять таки занимает квадрат памяти, поэтому воспользуемся идеей пересчета нового направления через старые значения. Для этого запомним $m = 7$ предыдущих значений (этот параметр является изменяемым, но как правило его берут меньшим 10), которые мы вычисляли в методе BFGS: s_k, y_k, ρ_k . В основном мы действуем также, как в методе BFGS, но уже считать вектор p_k , вдоль которого мы должны искать минимум функции, будем иначе. Сначала мы зафиксируем наш градиент: $q = \nabla f_k$ и пересчитаем его так:

$$\alpha_i = \rho_i \cdot s_i^T q$$

$$q = q - \alpha_i \cdot y_i$$

Так будем считать m раз от $k - m + 1$ до k в обратном порядке включительно. Теперь мы можем рассчитать приближенное направление как:

$z = \gamma \cdot q$, где $\gamma = \frac{s_k^T y_k}{y_k^T y_k}$. Затем мы пересчитываем наше направление по формуле: $z = z + s_i \cdot (\alpha_i - \beta_i)$ для i от $k - m + 1$ до k включительно, где $\beta_i = \rho_i \cdot y_i^T z$. Затем мы идем вдоль отрицательного направления z . И также аналогично BFGS будем искать точку минимума вдоль него.

Тесты

Для тестирования методов решим задачу минимальных квадратов для приближения точек следующими функциями:

Примечание: поскольку почти всегда методы выдают одну и ту же функцию, для наглядности было построено по одному графику на каждый метод и на каждую минимизируемую функцию.

1) Линейная функция $f(x) = ax + b$

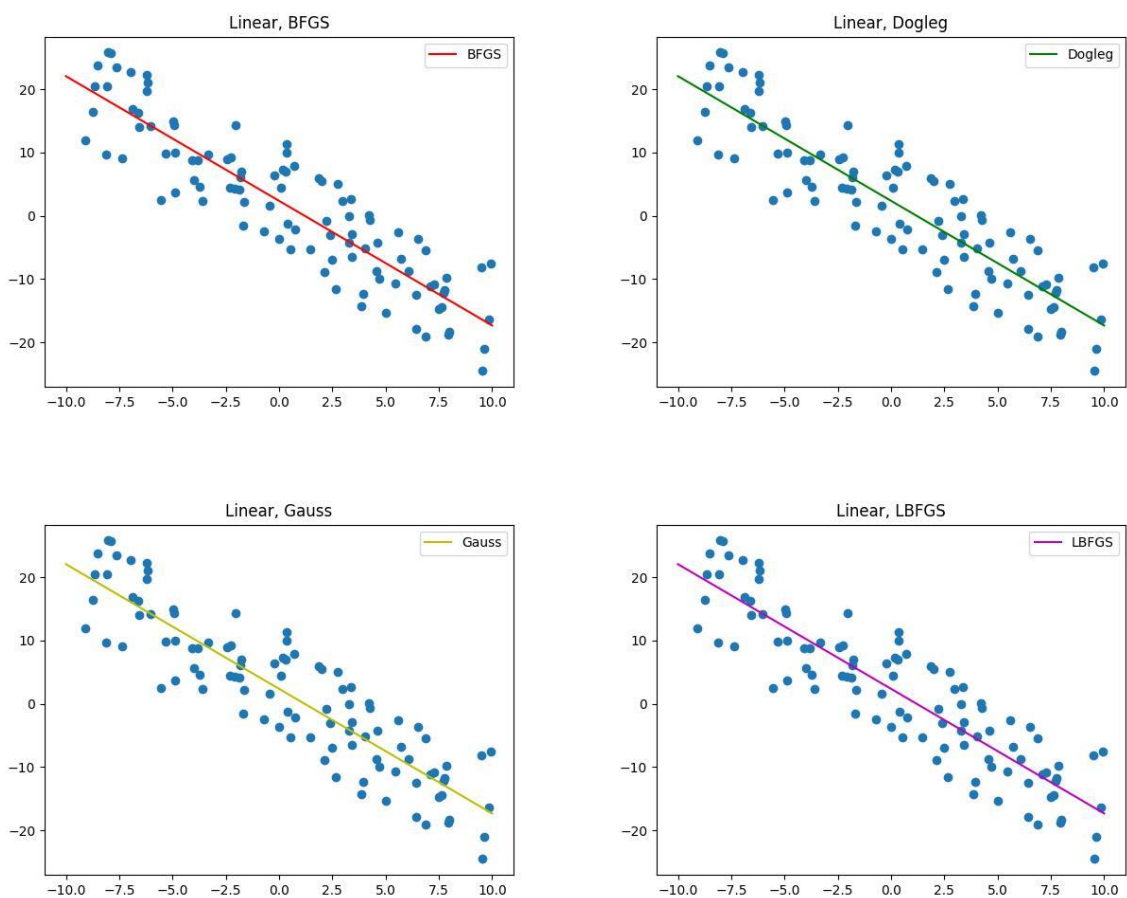


Рис. 1 – Графики аппроксимаций множества точек функцией Linear

2) Квадратичная функция $f(x) = ax^2 + bx + c$

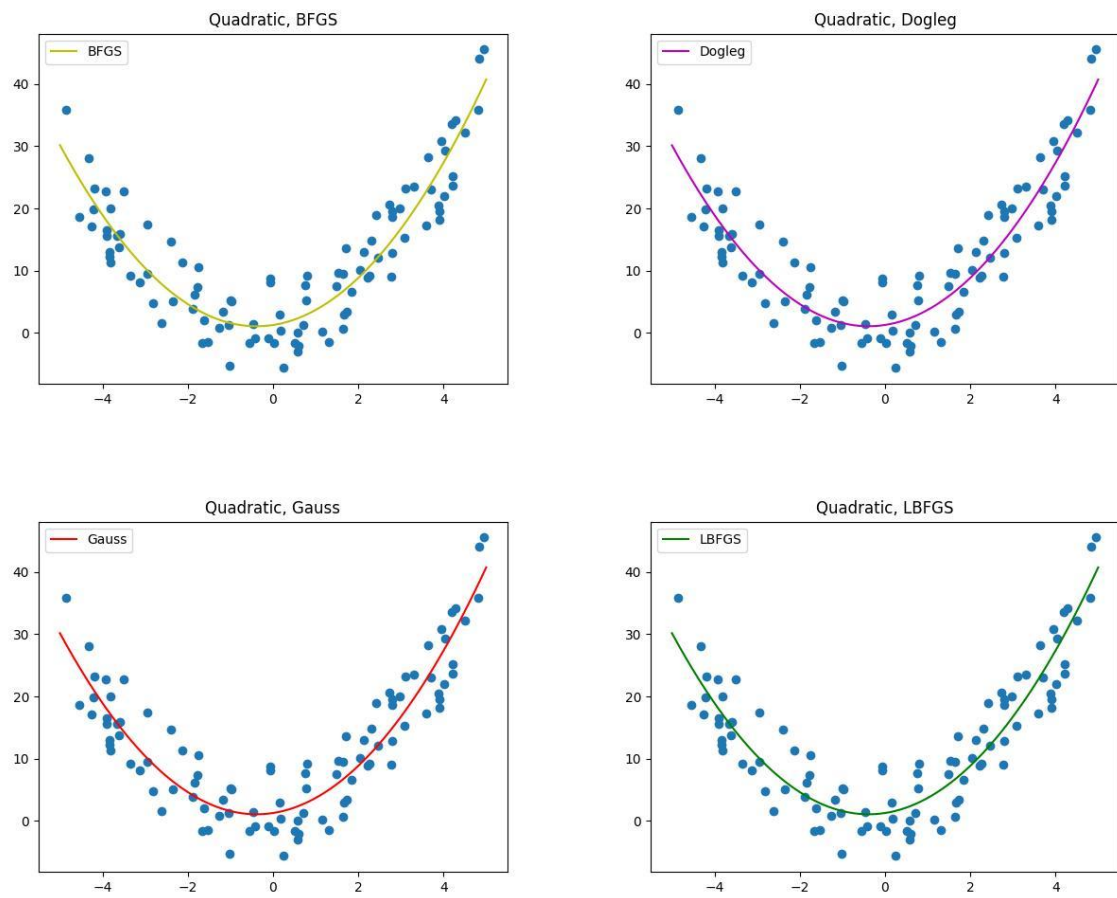


Рис. 2 – Графики аппроксимаций множества точек функцией Quadratic

3) Кубическая функция $f(x) = ax^3 + bx^2 + cx + d$

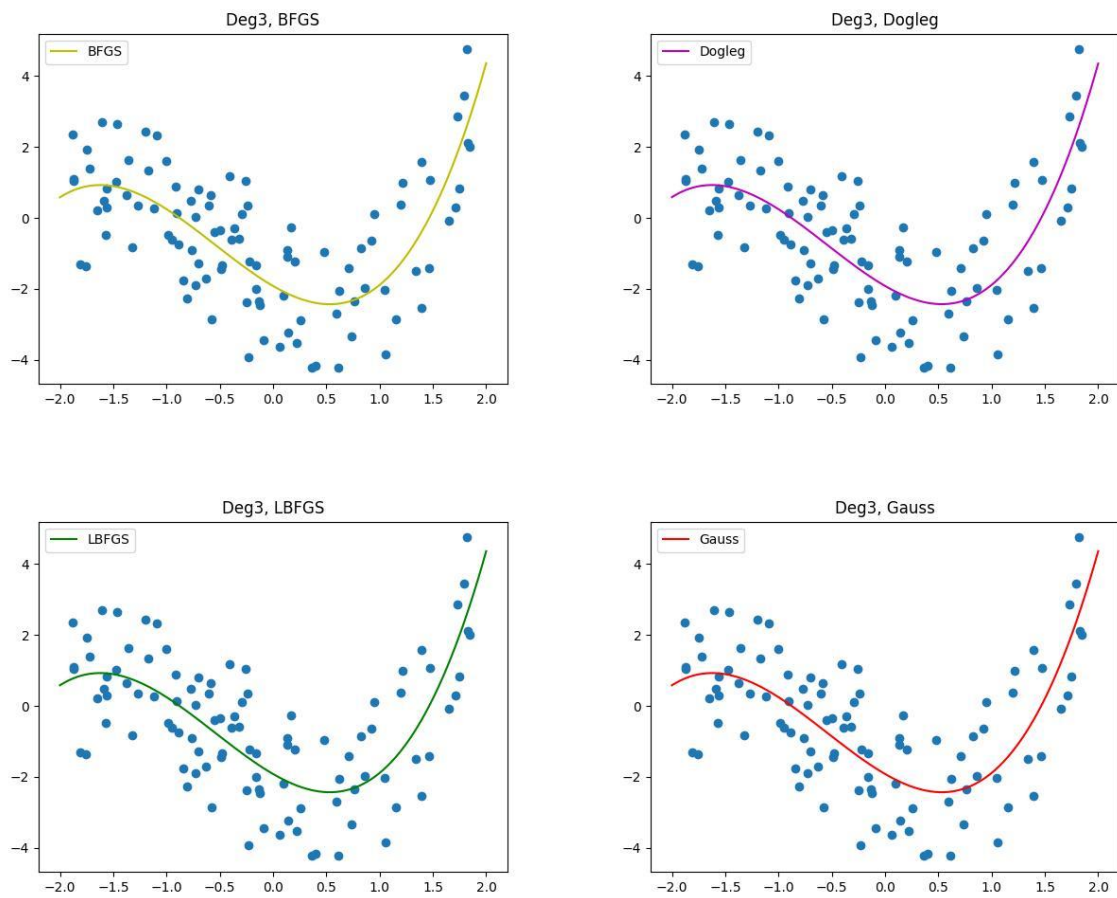


Рис. 3 – Графики аппроксимаций множества точек функцией Deg3

4) Полином 4-й степени $f(x) = ax^4 + bx^3 + cx^2 + dx + e$

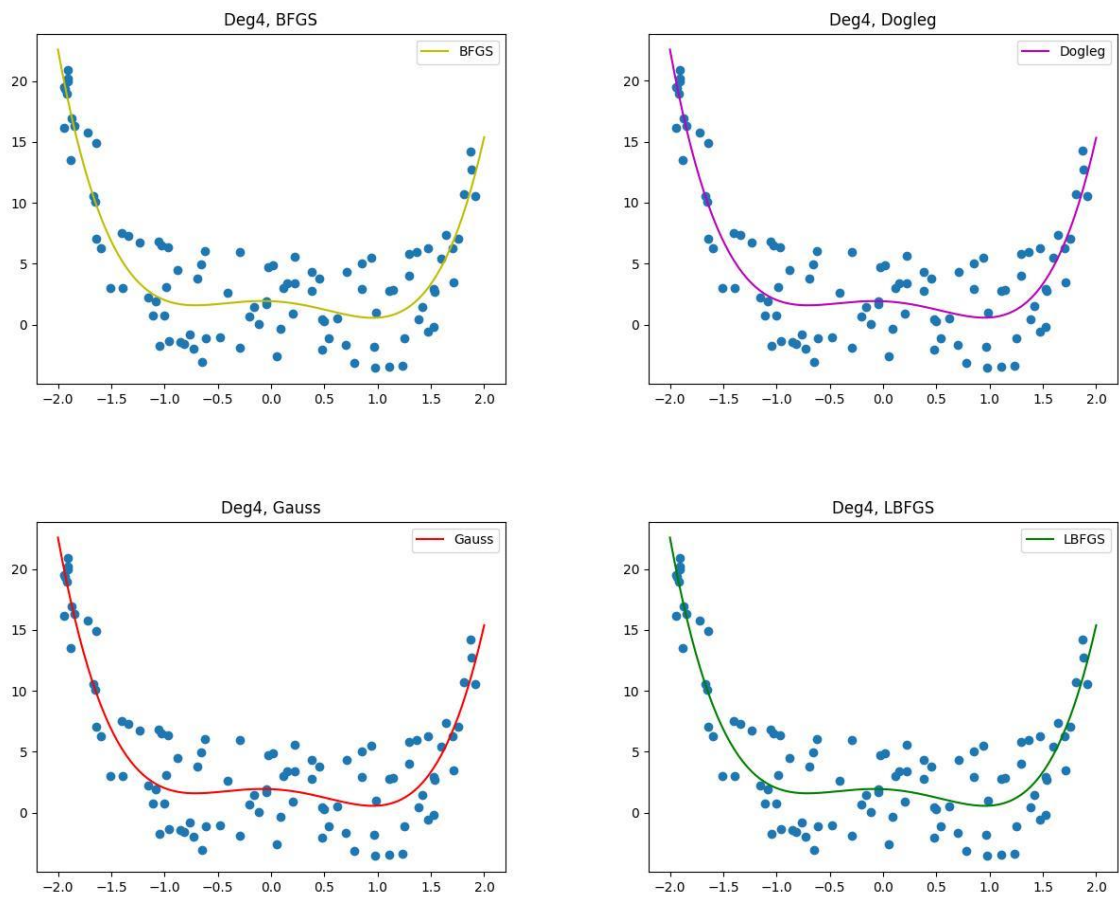


Рис. 4 – Графики аппроксимаций множества точек функцией Deg4

5) Полином 5-й степени $f(x) = ax^5 + bx^4 + cx^3 + dx^2 + ex + f$

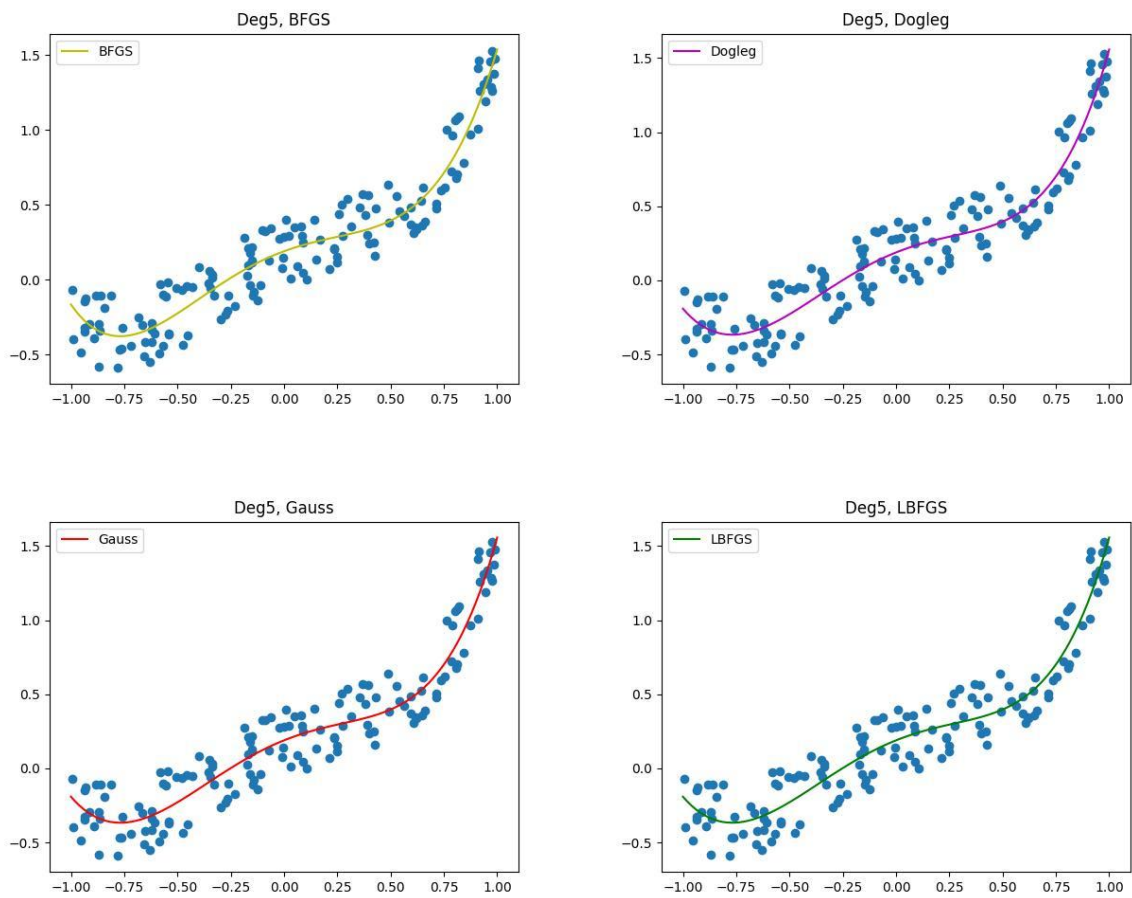


Рис. 5 – Графики аппроксимаций множества точек функцией Deg5

6) Степенная функция (DanWood) $f(x) = a \cdot x^b$

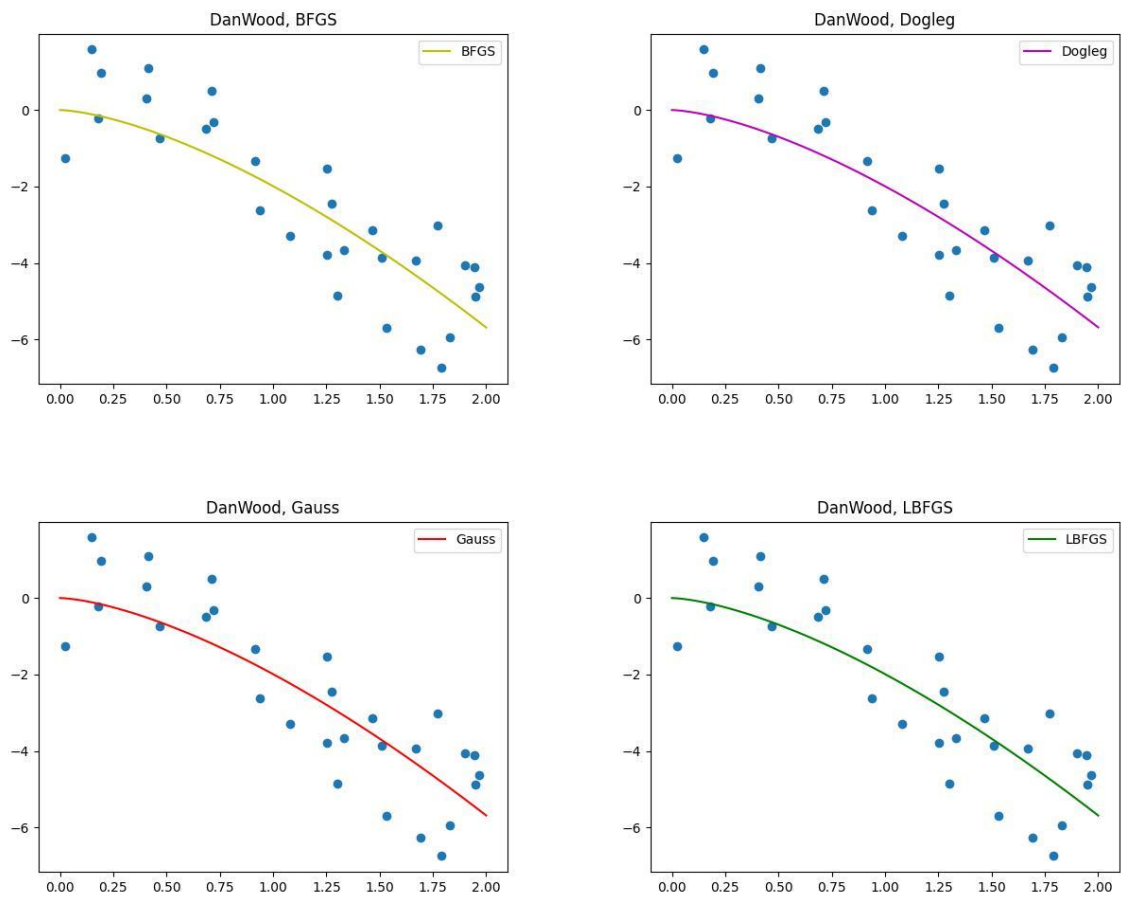


Рис. 6 – Графики аппроксимаций множества точек функцией DanWood

7) Рациональная функция “Misra1d” $f(x) = a \cdot \left(1 - \frac{1}{(1 + 0.5bx)^2}\right)$

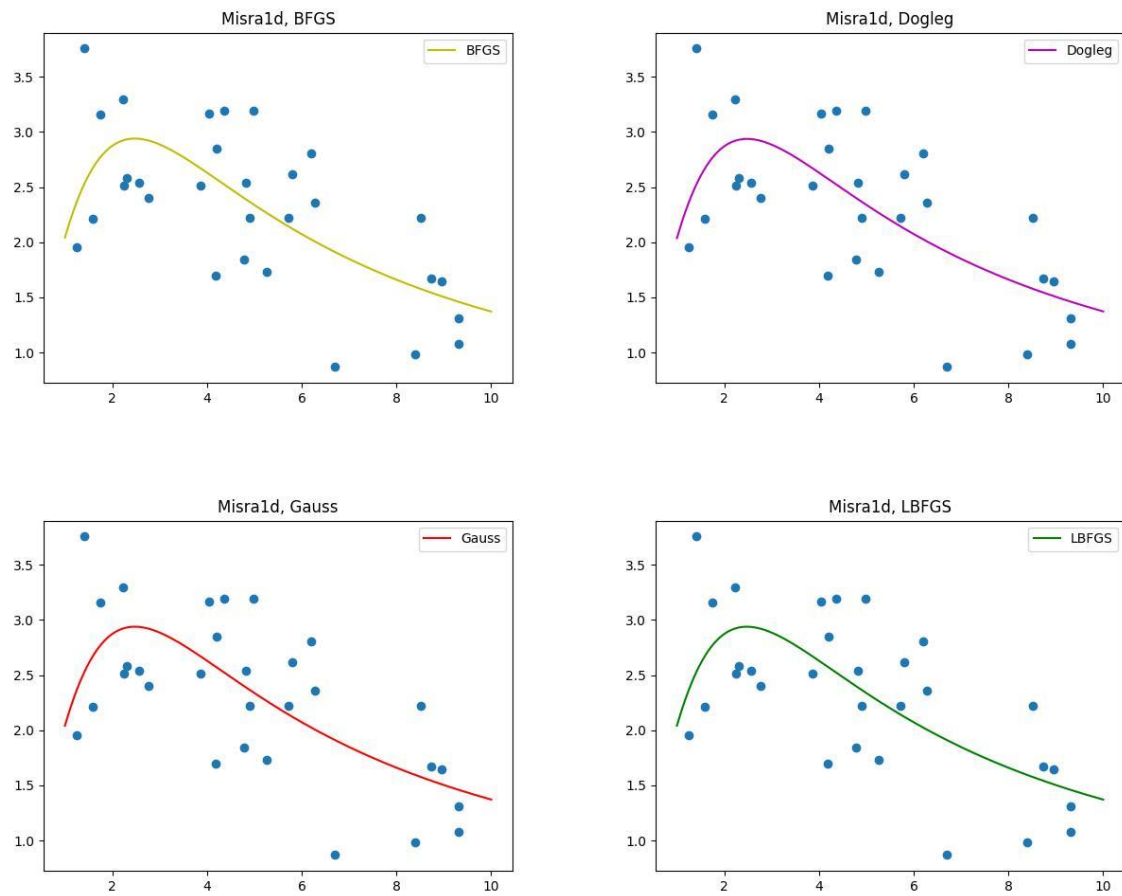


Рис. 7 – Графики аппроксимаций множества точек функцией Misra1d

8) $f(x) = a \sin x + b \cos x$

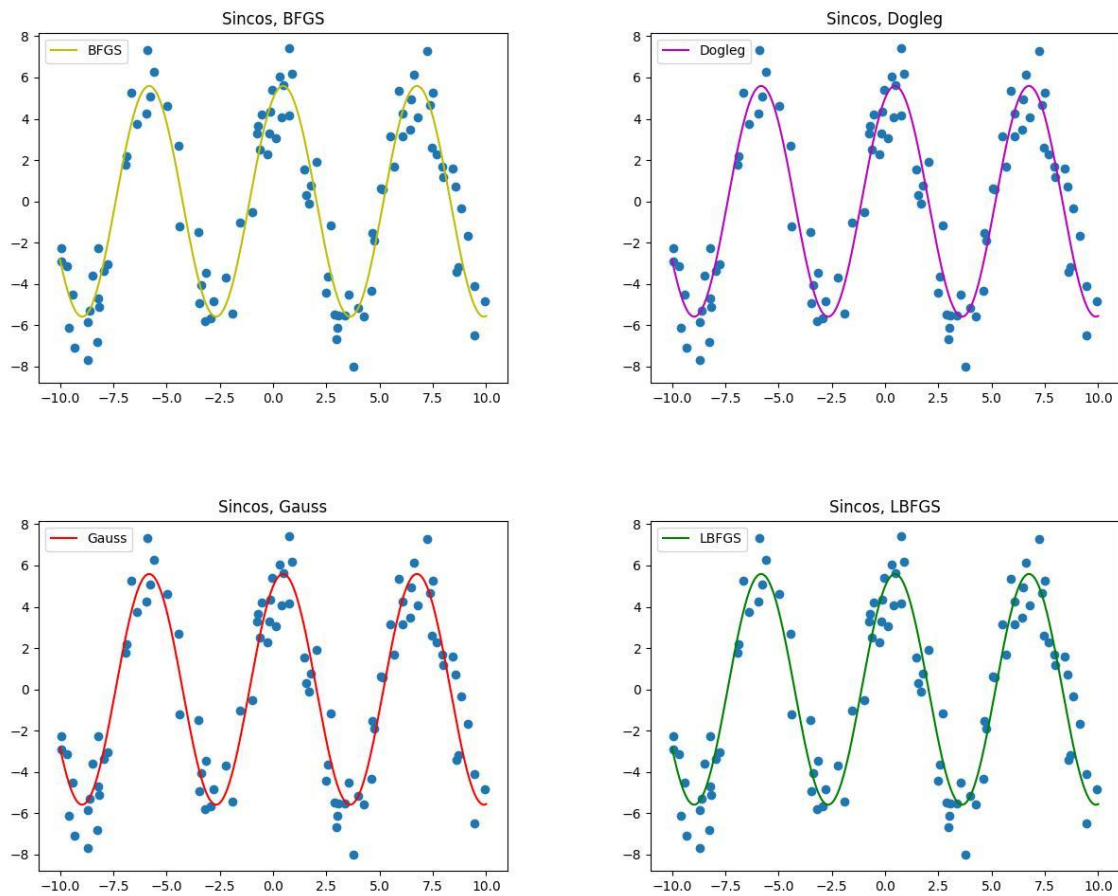


Рис. 8 – Графики аппроксимаций множества точек функцией SinCos

Было замечено, что линейная регрессия требовала гораздо менее тонкой настройки для корректной работы, а сходимость была более надежной по сравнению с реализованными во второй лабораторной методами. Оптимизируемое значение в методе Dogleg оказывается больше (разница порядка 10^{-4} , так как вычисление гессиана порождает довольно большую погрешность (при уменьшении соответствующих малых приращений сходимость улучшается). Можно заметить, что все алгоритмы сходятся за малое число итераций, что компенсирует тяжеловесность вычислений. На линейных функциях Ньютон-Гаусс, ожидаемо, сходится за 1-2 итерации (иногда за 2 из-за дальности стартовой точки и погрешности вычисления производной).

9) Квадратичные функции $f(x) = x^T Ax$

В качестве второго теста мы запускали методы Dogleg, BFGS, LBFGS на квадратичных функциях, сгенерированных в рамках первой лабораторной. Можно заметить, что с увеличением числа обусловленности сходимость Dogleg'a ухудшилась (вероятно, это связано с возрастающей значимостью погрешностей при вычислении гессиана при увеличении числа обусловленности). Интересно, что Dogleg сходится за почти константное количество шагов, тогда как число шагов BFGS и LBFGS возрастает. Связано это с тем, что данные методы считают Гессиан приближенно и в стартовой точке полагают его равным единичной матрице, что для квадратичной функции с большим числом обусловленности, очевидно, является неверным. Таким образом, данные методы тратят некоторое количество шагов на то, чтобы сделать правильную аппроксимацию Гессиана.

Для того чтобы убедиться, что LBFGS использует меньше памяти чем BFGS мы сделали следующее: создали квадратичную функцию с большим числом переменных и небольшим числом обусловленности (порядка 10). Мы получили, что для размерности равной 2000 использование памяти методом LBFGS 404 Мб против 463 Мб BFGS. Также при размерности равной 2500 LBFGS использовал 647 Мб против 847 Мб у BFGS.

Приложение 1.

Метод	Время (мс)	Память (Мб)	Шагов сходимости	Арифм. операций	Значение минимиз. функции
1					
dogleg	9391,66	222,43	4	8405556	2,5E-07
bfgs	94,90	222,41	1	1012795	3,7E-24
lbfgs	60,41	222,41	1	2622	3,7E-24
2					
dogleg	7240,97	222,43	4	8405556	5,4E-07
bfgs	54,56	222,42	14	14168513	1,3E-03
lbfgs	23,14	222,42	4	11166	1,2E-05
2,718281828					
dogleg	9875,54	222,43	4	8405556	8,9E-07
bfgs	74,71	222,41	17	17204435	1,4E-03
lbfgs	31,43	222,41	6	19431	1,1E-06
3					
dogleg	6392,95	222,42	4	8405556	1,0E-06
bfgs	334,23	222,42	20	20240372	1,2E-03
lbfgs	44,55	222,42	6	19431	3,4E-06
4					
dogleg	9419,46	222,42	4	8405556	1,6E-06
bfgs	421,94	222,42	21	21252354	1,3E-03
lbfgs	35,56	222,42	7	24330	4,5E-06
5					
dogleg	8450,43	222,42	4	8405556	2,3E-06
bfgs	88,18	222,42	23	23276309	1,2E-03
lbfgs	59,49	222,42	8	29235	7,9E-06
6					
dogleg	6519,15	222,42	4	8405556	3,1E-06
bfgs	183,89	222,42	24	24288291	1,0E-03

lbfgs	41,54	222,42	9	34137	4,9E-06
7					
dogleg	9308,28	222,42	4	8405556	3,9E-06
bfgs	191,65	222,42	26	26312234	9,0E-04
lbfgs	44,34	222,42	10	39039	6,2E-06
7,389056099					
dogleg	6316,90	222,41	4	8405556	4,1E-06
bfgs	113,03	222,41	26	26312243	1,2E-03
lbfgs	47,56	222,41	11	43941	2,1E-06
8					
dogleg	7803,42	222,42	4	8405556	4,9E-06
bfgs	107,26	222,42	27	27324213	8,2E-04
lbfgs	59,88	222,42	11	43941	5,4E-06
9					
dogleg	8477,63	222,42	4	8405556	5,7E-06
bfgs	119,69	222,42	27	27324213	8,3E-04
lbfgs	51,35	222,42	12	48843	4,9E-06
10					
dogleg	8876,70	222,42	4	8405556	8,6E-06
bfgs	115,52	222,42	25	25300267	1,2E-03
lbfgs	74,84	222,42	13	53745	2,4E-06
11					
dogleg	8043,38	222,42	4	8405556	9,3E-06
bfgs	128,89	222,42	25	25300264	1,1E-03
lbfgs	54,95	222,42	13	53742	4,7E-06
20,08553692					
dogleg	6312,72	222,41	4	8406066	2,3E-05
bfgs	179,08	222,41	45	45539838	3,1E-04
lbfgs	88,82	222,41	18	78255	3,7E-06
54,59815003					
dogleg	6323,65	222,41	4	8406576	1,9E-04

bfgs	220,37	222,41	60	60719532	7,3E-05
lbfgs	126,34	222,41	31	141984	2,8E-06
148,4131591					
dogleg	6386,35	222,41	4	8406576	1,3E-03
bfgs	400,51	222,41	113	114354185	3,1E-05
lbfgs	149,83	222,41	37	171393	6,3E-07
403,4287935					
dogleg	9861,40	222,43	4	8407086	8,4E-03
bfgs	1514,53	222,41	225	227695443	1,0E-05
lbfgs	483,91	222,41	62	293997	5,3E-06
1096,633158					
dogleg	9477,88	222,42	4	8406576	6,0E-02
bfgs	1460,24	222,41	314	317761145	5,6E-06
lbfgs	241,11	222,41	54	254769	3,4E-08
2980,957987					
dogleg	9040,56	222,42	5	10507486	4,0E-01
bfgs	3567,06	222,41	353	357228263	2,1E-06
lbfgs	608,20	222,41	61	289086	1,7E-08
8103,083928					
dogleg	8469,82	222,42	5	10507486	2,8E+00
bfgs	1531,22	222,41	371	375443921	5,1E-07
lbfgs	376,01	222,41	65	308697	1,2E-08

Таблица 1 – усредненная статистика для исследования сходимости квадратичных 100-мерных функций с разным числом обусловленности.