Санкт-Петербургский политехнический университет имени
Петра Великого

Институт прикладной математики и механики

Высшая школа прикладной математики и физики

Направление подготовки
«01.03.02 Прикладная математика и информатика»

Курсовая работа «Сравнительный анализ градиентных методов второго порядка»

Дисциплина «Методы оптимизации»

Выполнили студенты группы 5030102/90201: Габдушев Р.М.
Фисюк А.Ю.
Ахметшин Р.И.

Преподаватель: Родионова Е.А.

Санкт-Петербург
2022 г.

Содержание

Список иллюстраций

1. Введение

Проблема оптимизации функций многих переменных имеет приложения во всех сферах, где используются математическое моделирование в любом виде.

Задачи линейного программирования - простейший класс таких задач, широко распространён на практике, хорошо изучен и имеет эффективные методы решения.

Однако, список задач, требующих оптимизации, выходит далеко за эти рамки. И не каждая этих них может похвастаться настолько эффективными методами решения.

В этой работе мы рассмотрим класс задач выпуклой оптимизации, а именно оптимизация функций нескольких переменных, без ограничений.

Целью работы является сравнительный анализ градиентных методов второго порядка:

- Градиентный метод Ньютона
- Градиентный метод Ньютона с дроблением шага (Пшеничного)
- Градиентный метод Бройдена - Флетчера - Гольдфарба - Шанно(БФГШ)

В частности:

- найти ориентировочную вычислительную сложность на один наг алгоритма
- привести графическую иллюстрацию хода работы методов
- провести сравнение скорости сходимости на наборе показательных функций
- вынести рекомендации о выборе метода, в зависимости от особенностей функции

2. Постановка задачи

Дана функция

$$f : \mathbb{R}^n \to \mathbb{R} \tag{1}$$

Найти

$$x_n \in B(\epsilon, x_*) : f(x_*) = \min_{x \in \mathbb{R}^n} f(x) \tag{2}$$

где $\epsilon > 0$ - заданная точность.

3. Описание методов

Градиентные методы основаны на идее аппроксимации минимизируемой функции её разложением в текущей точке, удерживая члены до определённого порядка. В случае методов второго порядка, аппроксимация функции f в точке x_k выглядит так:

$$f(x) \approx f(x_k) + \nabla^T f(x_k)(x - x_k) + (x - x_k)^T H(x_k)(x - x_k)/2 \tag{3}$$

где $\nabla f(x_k)$ - значение градиента, а $H(x_k)$ - матрица Гессе, в точке разложения.

Предполагая, что данная функция принадлежит по крайней мере C^2 и выпукла, делаем вывод, что и её приближение выпукло. Значит оно имеет единственный минимум x_*.[3]

Из (3) элементарно вычисляется оптимизирующий вектор $p_k = x_* - x_k$:

$$p_k = -(H(x_k))^{-1} \nabla f(x_k) \tag{4}$$

3.1. Градиентный метод Ньютона

Метод ньютона предполагает следующее правило построения минимизирующей последовательности:

1. $x_k = x_0$
2. Если условие окончания не выполнено - дальше
3. $p_k = -(H(x_k))^{-1} \nabla f(x_k)$
4. $x_{k+1} = x_k + p_k$
5. $k = k + 1$
6. Возврат к пункту 2

x_0 - начальное приближение

3.2. Градиентный метод Ньютона с дроблением шага (Пшеничного)

Данная модификация метода Ньютона предлагает перед переходом к следующему шагу, дробить длину шага пока не будет выполнено особое условие.

Алгоритм построения последовательности:

1. $x_k = x_0$
2. Если условие окончания не выполнено - дальше
3. $p_k = -(H(x_k))^{-1}\nabla f(x_k)$
4. $\alpha = 1$
5. Если $f(x_k + \alpha p_k) - f(x_k) \leq \alpha\delta\nabla^T f(x_k)p_x$ - переход к пункту 6
6. $\alpha = \alpha\lambda$
7. Возврат к пункту 5
8. $x_{k+1} = x_k + \alpha p_k$
9. Возврат к пункту 2

x_0 - начальное приближение, $0 < \delta \leq \frac{1}{2}, 0 < \lambda < 1$ - параметры дробления.

3.3. Градиентный метод БФГШ

В приложениях исходная функция, как правило, не задаётся формулой, а определяется в ходе измерений. Поэтому матрицу Гессе тоже не получить символьно, а её вычисление требует довольно много обращений к функции. К тому же на каждом шаге необходимо её обращать.

Градиентный метод Бройдена - Флетчера - Гольдфарба - Шанно(БФГШ) относится к квазиньютоновским и помогает обойти эти трудности.

Вместо точного вчисления матрицы Гессе и её обращения, строится последовательность матриц, которая [1] сходится к обращённой матрице гессе.

Схема вычисления[1]:

$$C_{k+1} = (I - \rho_k s_k y_k^T)C_k(I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T, \tag{5}$$

где $\rho_k = \frac{1}{y_k^T s_k}$, I - единичная матрица, $s_k = x_{k+1} - x_k$ - шаг алгоритма на итерации, $y_k = \nabla f_{k+1} - \nabla f_k$ - изменение градиента на итерации. В качестве начального приближения матрицы можно взять единичную.

Также такая матрица обладает свойством самокоррекции, если длина шага выбирается одномерным поиском минимума функции по найденному направлению p_k:

$$f(x_k + \alpha_k p_k) = min_{\alpha \geq 0} f(x_k + \alpha p_k) \tag{6}$$

В итоге алгоритм выглядит следующим образом:

1. $x_k = x_0$
2. $C_k = C_0$
3. Если условие окончания не выполнено - дальше
4. $p_k = -(C_k)\nabla f(x_k)$
5. $\alpha = 1$
6. Поиск оптимального $\alpha_x f(x_k + \alpha_k p_k) = min_{\alpha \geq 0} f(x_k + \alpha p_k)$
7. $s_k = \alpha p_k$
8. $x_{k+1} = x_k + s_k$
9. Обновить C_{k+1} по формуле (5)
10. Возврат к пункту 2

Для всех методов существуют вариации, в которых матрица вычисляется не на каждом шаге, но в нашей работе мы вычисляем её каждый раз.

4. Выбор функций для оптимизации

Для наглядной графической иллюстрации работы методов, мы будем оптимизировать функции двух переменных.

1. Рассмотрим функцию второго порядка:

$$f_1(x, y) = 2x^2 + y^2 + xy \tag{7}$$

   Минимум этой функции методы второго порядка в идеале должны находить за один шаг.

2. Также рассмотрим функцию, состоящую из членов разложения степени выше второй

$$f_2(x, y) = x^4 + (\frac{y}{2})^4 \tag{8}$$

3. Ещё рассмотрим эту же функцию, но с добавлением квадратичных членов:

$$f_3(x, y) = x^4 + (\frac{y}{2})^4 + (\frac{x}{2})^2 + y^2 \tag{9}$$

4. Также, рассмотрим функцию, включающую степени выше 2 с малыми коэффициентами:

$$f_4(x, y) = x^2 + y^2 + 0.05(x^4 + y^4) + 0.0025(x^6 + y^6) + x + y \tag{10}$$

5. Применимость методов

Градиентные методы второго порядка применимы к строго выпуклым, дважды непрерывно дифференцируемым функциям.[2] Покажем, что это верно для выбранных функций, приведя их градиенты и матрице Гессе:

$$\nabla f_1(x, y) = (4x + y, 2y + x) \tag{11}$$

$$H_1(x, y) = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} \tag{12}$$

$$\nabla f_2(x, y) = (4x^3, \frac{1}{4}y^3) \tag{13}$$

$$H_2(x, y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & \frac{3}{4}y^2 \end{pmatrix} \tag{14}$$

$$\nabla f_3(x, y) = (4x^3 + \frac{x}{2}, \frac{y^3}{4} + 2y) \tag{15}$$

$$H_3(x,y) = \begin{pmatrix} 12x^2 + \frac{1}{2} & 0 \\ 0 & 3\frac{y^2}{4} + 2 \end{pmatrix} \tag{16}$$

$$\nabla f_4(x,y) = (0.0015x^5 + 0.02x^3 + 2x + 5, 0.0015y^5 + 0.02y^3 + 2y + 5) \tag{17}$$

$$H_4(x,y) = \begin{pmatrix} 0.0075x^4 + 0.6x^2 + 2 & 0 \\ 0 & 0.0075y^4 + 0.6y^2 + 2 \end{pmatrix} \tag{18}$$

Видим, что все функции выпуклые, и все, кроме f_2 - строго выпуклые. Рассмотрим поведение методов на нестрого выпуклых функциях, а потому не будем исключать f_2.

6. Результаты

Для начала проведём вычисления с использованием малости нормы градиента в качестве критерия остановки.

Вычисления проводились для точности 10^{-6}, параметры дробления метода Пшеничного: $\lambda = 0.9, \epsilon = 0.45$

Далее приведены графики зависимости точности (нормы разности текущего и точного значения) от номера шага и геометрическая интерпретация работы методов (показаны линии уровня и вектора шага).

Здесь и далее, графики в логарифмическом размере выглядят вертикальными, если принимают значение 0 (точное решение).

Зелёный пунктир - заданная точность

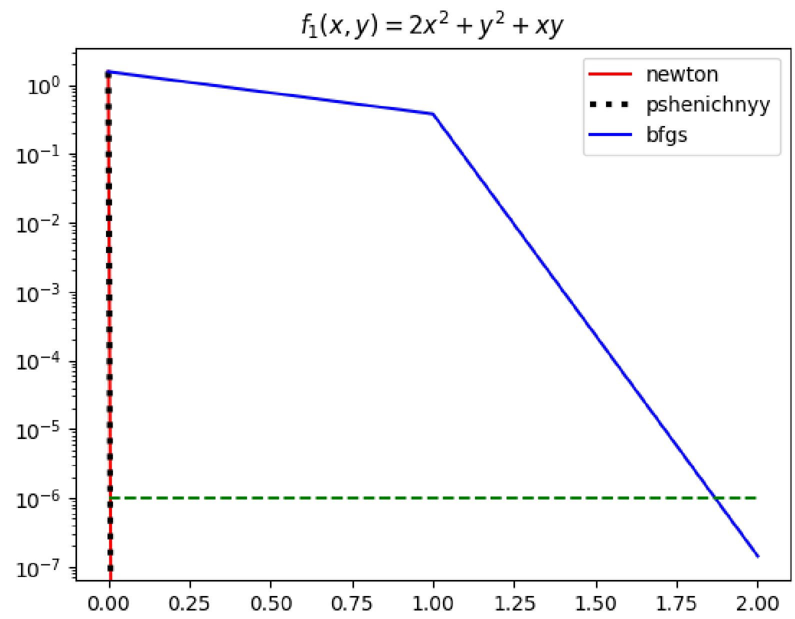

Рис. 1. Точность от шага, f_1, критерий - градиент

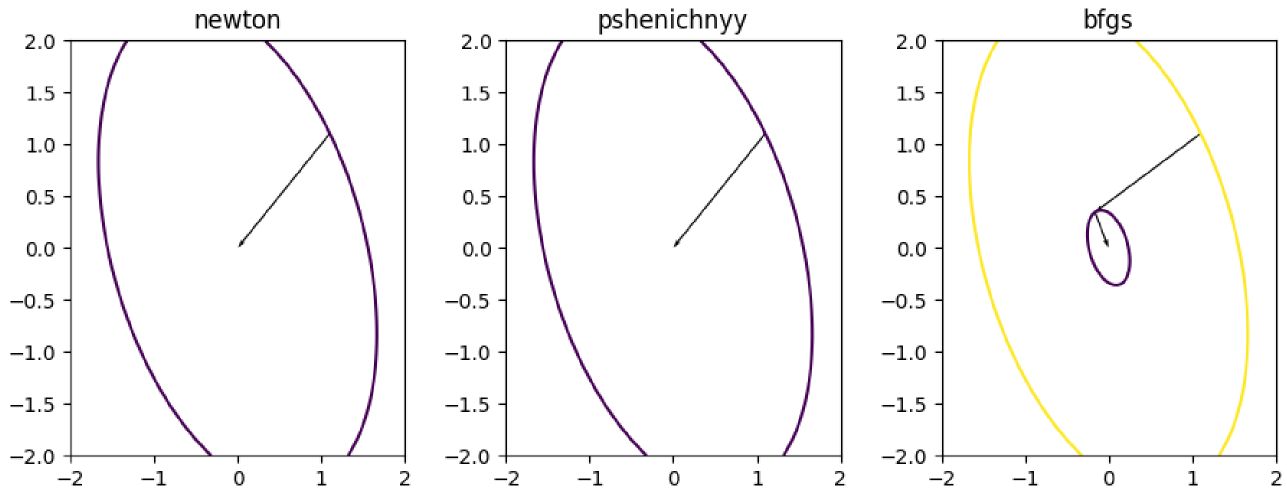

Рис. 2. Иллюстрация работы алгоритмов, f_1

Как видим, заданная точность достигается всеми алгоритмами, остановка происходит вовремя. Как и ожидалось, методы второго порядка решили точно за 1 шаг. Квазиньютоновский метод очень хорошо приблизил решение на втором шаге (сразу после первой корректировки приближения матрицы Гессе)

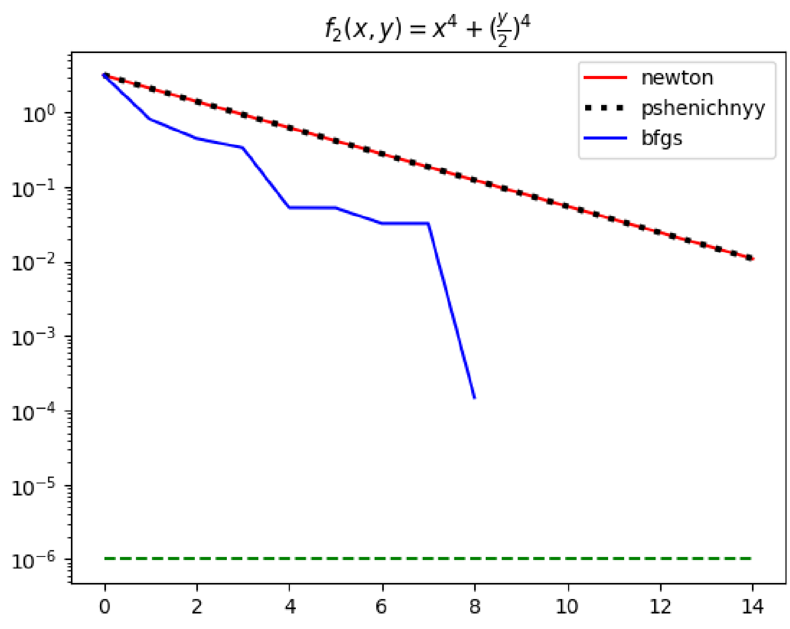

Рис. 3. Точность от шага, f_2, критерий - градиент

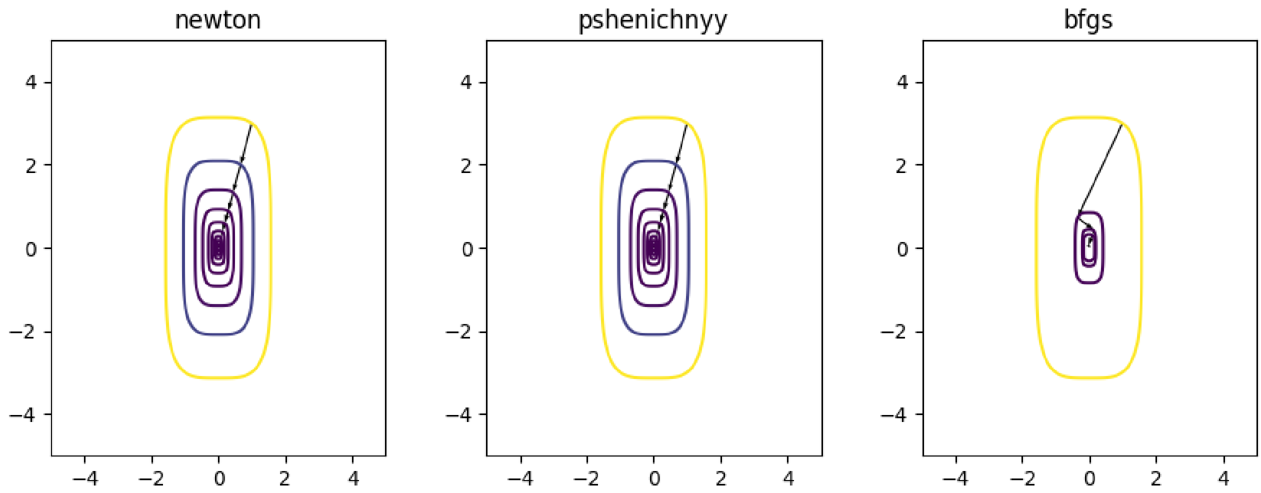

Рис. 4. Иллюстрация работы алгоритмов, f_2

Заметим, что так как в окрестности точки минимума функция ведёт себя как функция четвёртой степени, градиентная оценка приводит к преждевременной остановке алгоритма.

Также, этот опыт показывает, что данная функция плохо аппроксимируется эллиптическим параболоидом, поэтому методы второго порядка медленно сходятся. Квазиньютоновский метод в этом случае оказывается быстрее, благодаря оптимальному выбору длины шага.

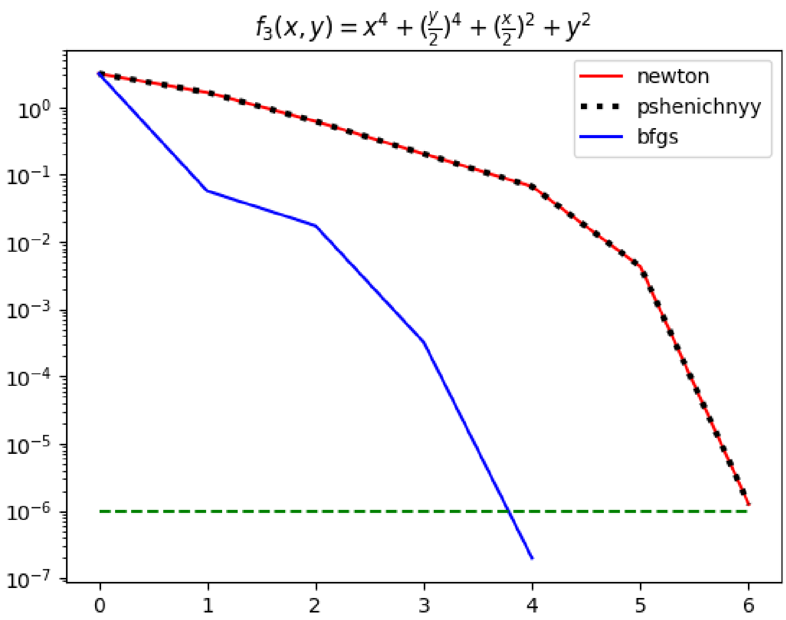

Рис. 5. Точность от шага, f_3, критерий - градиент

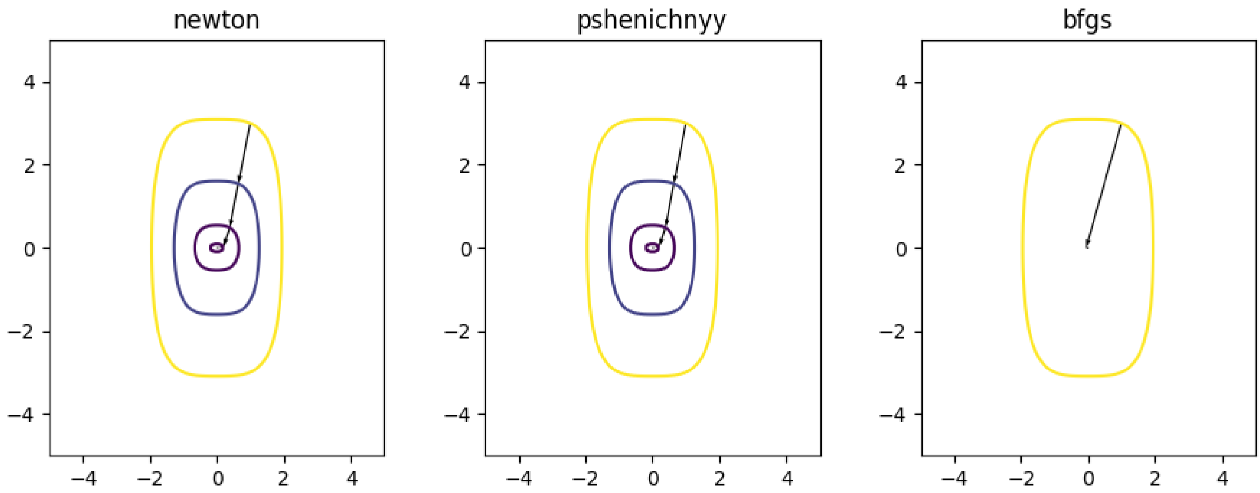

Рис. 6. Иллюстрация работы алгоритмов, f_3

Видим, что при добавлении к f_2 квадратных членов достоверность градиентного критерия улучшилась, хотя и не полностью, так как методы второго порядка хоть и приблизились близко, но не дали заданной точности.

Одновременно с этим, все методы стали сходиться быстрее, но квазиньютоновский метод всё ещё обходит квадратичные.

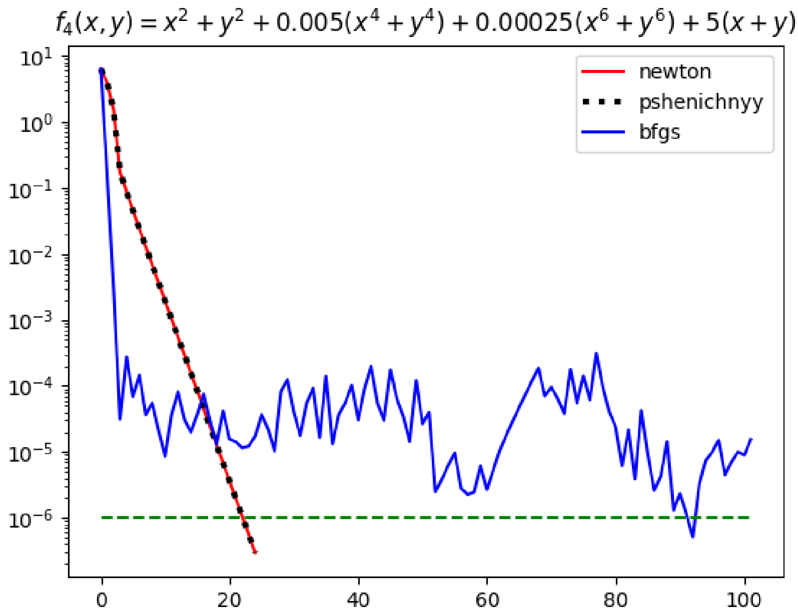

Рис. 7. Точность от шага, f_4, критерий - градиент

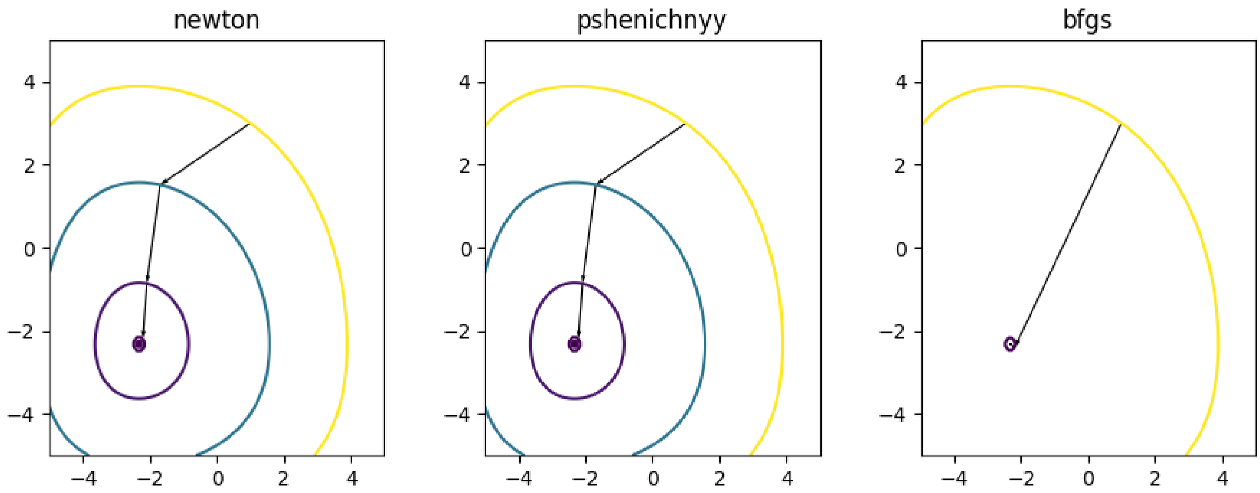

Рис. 8. Иллюстрация работы алгоритмов, f_4

Как видим, на данной функции квадратичные методы сходятся медленнее, чем в случае f_3, а квазиньютоновский и вовсе перестаёт сходиться ближе 10^-4.

Отметим, что эта функция единственная из рассмотренных с минимумом не в нуле. Но, проведя исследования этой же функции с перенесённым в точку минимума началом координат, мы увидели абсолютно те же результаты.

Градиентная оценка здесь приводит к нескольким лишним шагам алгоритма.

Как показал этот опыт, на выбранных нами функциях метод Пшеничного всегда выполняет 0 дроблений и вырождается в метод Ньютона, поэтому далее не будем его рассматривать.

В следующем опыте проверим скорость сходимости методов, в зависимости от начального приближения. Для этого проведём вычисления для 100 точек, равномерно распределённых на окружности с центром в точке минимума и радиусом 2. На следующих графиках изображена зависимость близости решения от номера шага алгоритма. В этом опыте критерием остановки взято расстояние до точного решения, так как ранее было показано, что градиентный критерий даёт разные результаты для разных функций.

На рисунках красным обозначены решения методом Ньютона, синим - методом БФГШ. Зелёным пунктиром отмечена целевая точность $= 10^-6$.

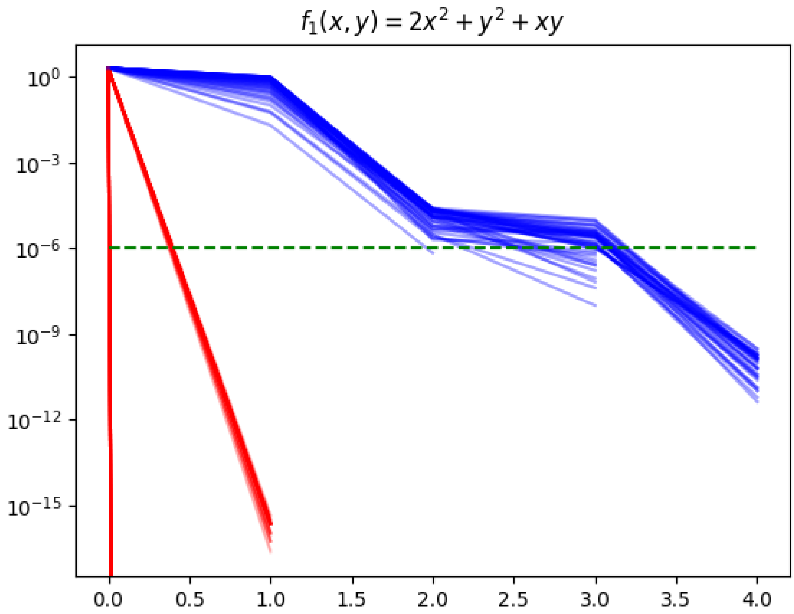

Рис. 9. Выборка начальных приближений, f_1

Для первой функции метод Ньютона всегда находит решение за 1 шаг, в некоторых случаях - точно (вертикальные линии на графике). Квазиньютоновскому методу требуется от 2 до 4 шагов. В среднее число шагов в выборке 3,7.

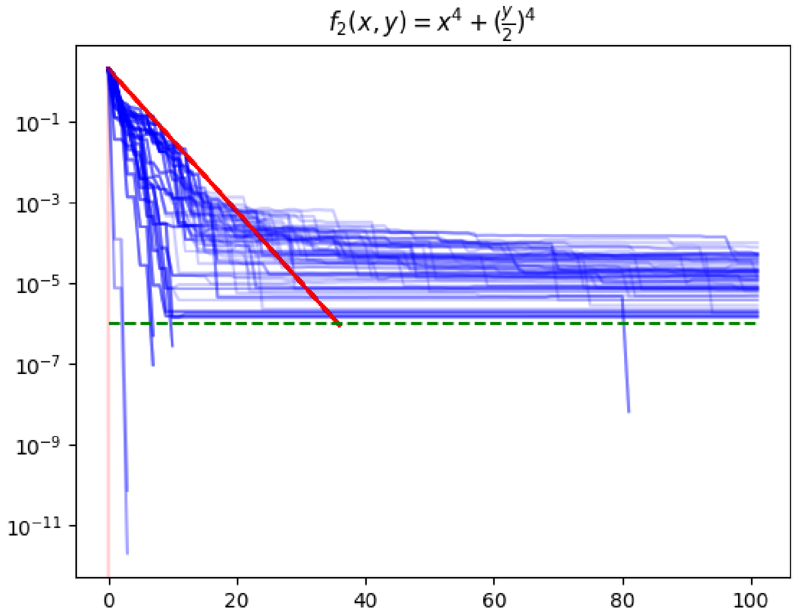

Рис. 10. Выборка начальных приближений, f_2

Здесь становится видно, квазиньютоновский метод плохо справляется с полиномами 4 степени и не сходится в большинстве случаев. Метод ньютона же, сходится за 36 шагов, и обладает зависимостью $n = -5,7log_10(|x_* - x|)$.

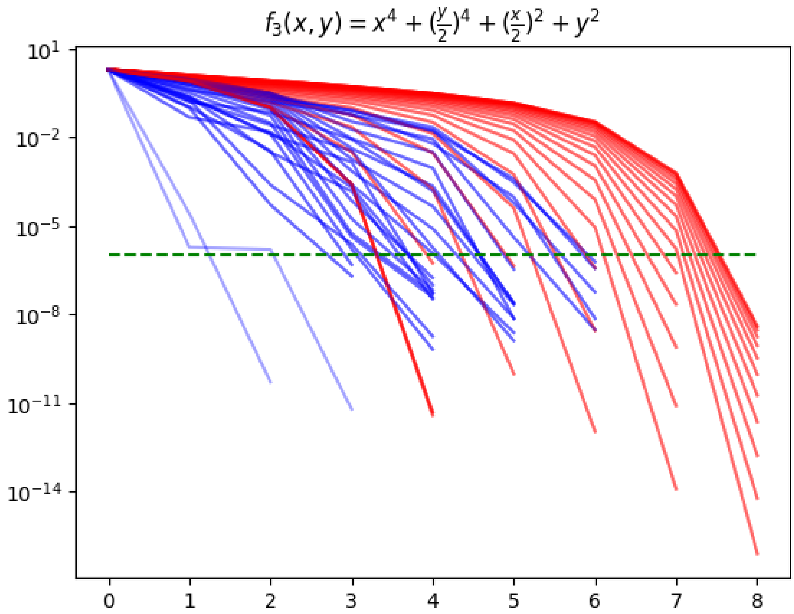

Рис. 11. Выборка начальных приближений, f_3

При добавлении квадратичных членов оба метода стали хорошо сходиться. Но при этом, для данной функции наблюдается заметный разброс в скорости сходимости от начального приближения для обоих методов. Среднее число шагов:

Метод ньютона - 6.76

Метод БФГШ - 4.58

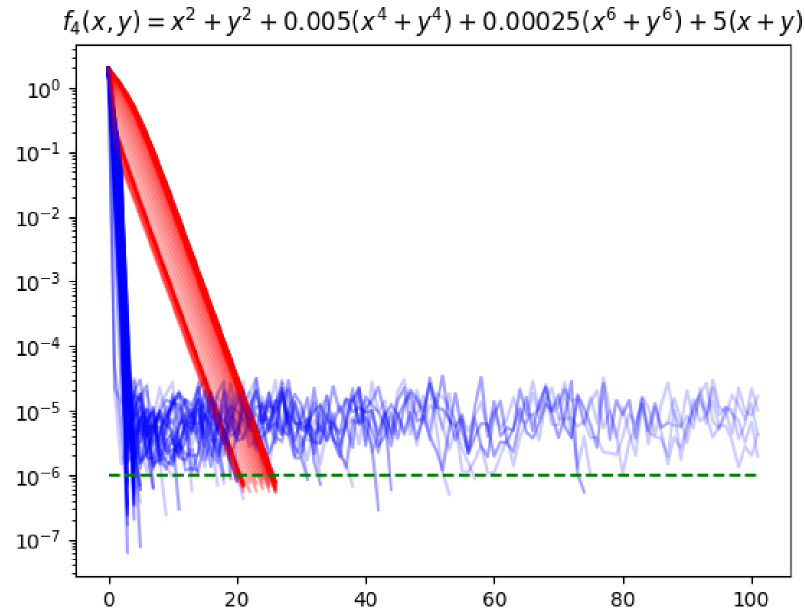

Рис. 12. Выборка начальных приближений, f_4

Для данной функции снова имеем не сходящийся ближе чем 10^-5 метод БФГШ и зависимость $n = a+blog_10(|x_*-x|)$. Однако, в отличие от функции f_2, где БФГШ не ухудшал результат, здесь явно видны скачки точности в худшую сторону.

7. Заключение

В ходе работы было показано, что квазиньютоновские методы могут показывать себя лучше Ньютоновских на определённых задачах, а также позволяют не вычислять вторые производные, что очень полезно, если функция вычисляется трудозатратно. Однако, в ряде других случаев, метод БФГШ вообще не может достичь поставленной точность, тогда как Ньютоновский метод сходится. Таким образом, имеет смысл сначала использовать метод БФГШ, а после достаточного приближения к точке, переходить на метод Ньютона.

Что касается метода Пшеничного, нам не удалось найти строго выпуклые функции, на которых он вёл бы себя отлично от метода Ньютона, но это не значит, что их нет.

8. Список литературы

1. Bonnans, J. Frédéric; Gilbert, J. Charles; Lemaréchal, Claude; Sagastizábal, Claudia A., "Newtonian Methods Numerical Optimization: Theoretical and Practical Aspects (Second ed.), Berlin: Springer, 2006. pp. 51–66

2. Базара М., Шетти К. Нелинейное программирование.-М.: Мир, 1982.- 583 с.

3. Болдырев Ю. Я., Родионова Е. А. Методы оптимизации. Математическое программирование: Учеб. пособие СПб.: Изд-во СПбГТУ, 1999. 82 с.