Санкт-Петербургский политехнический университет
Высшая школа прикладной математики и вычислительной физики, ФизМех

Направление подготовки
«01.03.02 Прикладная математика и информатика»
Специальность «Системное программирование»

Лабораторная работа №4
тема "Многомерная минимизация без ограничений"
дисциплина "Методы оптимизации"

Выполнили студенты гр. 5030102/00201	Гвоздев С.Ю., Золин И.М. Хламкин Е.В.
Преподаватель:	Родионова Е.А.

Санкт-Петербург

2023

Содержание

1 Постановка задачи

Пусть дана задача двумерной минимизации $f(x) = 4x_1 + x_2 + 4\sqrt{1 + 3x_1^2 + x_2^2}$. Необходимо:

1. Реализовать градиентный метод первого порядка наискорейшего спуска с поиском шага по методам золотого сечения и пробных точек
2. Нарисовать линии уровней функции цели
3. Показать в ходе вычислительного эксперимента ортогональность звеньев градиентной ломанной на двух последовательных итерационных шагах для точности 0.01
4. Реализовать ДФП-метод - градиентный метод второго порядка
5. Решить задачу с точностью от 10^{-1} до 10^{-4}
6. Выполнить сравнительный анализ алгоритмов методов

2 Исследование применимости методов

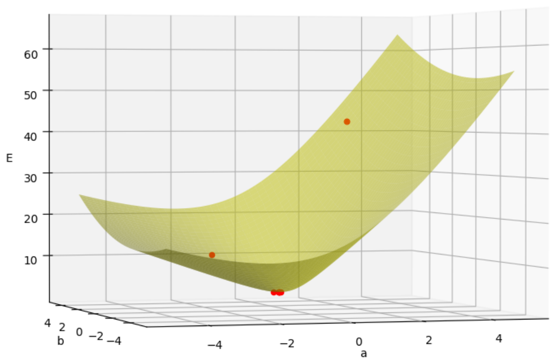

Рис. 0: График целевой функции.

Необходимо выполнение теоремы о скорости сходимости градиентного метода наискорейшего спуска. Для других методов существуют аналогичные теоремы, отличающиеся только значением скорости сходимости.

Theorem 1 *Если функция $f(x)$ дважды непрерывно дифференцируема и существуют такие числа $0 < m \leq M$, что*

$$m||x||^2 \leq x^T H(y) x \leq M||x||^2, \quad \forall\ x, y$$

тогда $x_k \longrightarrow x^*$, $f(x_k) \longrightarrow f(x^*)$ при любой начальной точке x_0, где x^* - оптимальная.

Используем функцию $f(x) = 4x_1 + x_2 + 4\sqrt{1 + 3x_1^2 + x_2^2}$.

$$\frac{\partial f(x)}{\partial x_1} = 4 + \frac{12x_1}{\sqrt{1 + 3x_1^2 + x_2^2}}$$

$$\frac{\partial f(x)}{\partial x_2} = 1 + \frac{4x_2}{\sqrt{1+3x_1^2+x_2^2}}$$

$$\frac{\partial^2 f(x)}{\partial x_1^2} = \frac{12(1+3x_1^2+x_2^2)-36x_1^2}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}} = \frac{12(1+x_2^2)}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}}$$

$$\frac{\partial^2 f(x)}{\partial x_1 x_2} = \frac{-12x_1x_2}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}}$$

$$\frac{\partial^2 f(x)}{\partial x_2^2} = \frac{4(1+3x_1^2+x_2^2)-4x_2^2}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}} = \frac{4(1+3x_1^2)}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}}$$

$$\det(H) = \frac{48(1+x_2^2)(1+3x_1^2)-144x_1^2x_2^2}{(1+3x_1^2+x_2^2)^3} = \frac{48(1+3x_1^2+x_2^2)}{(1+3x_1^2+x_2^2)^3} = \frac{48}{(1+3x_1^2+x_2^2)^2}$$

По критерию Сильвестера:

$$\begin{cases} \frac{12(1+x_2^2)}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}} > 0 \\ \frac{48}{(1+3x_1^2+x_2^2)^2} > 0 \end{cases} \Rightarrow \texttt{выполнено} \quad \forall x_1, x_2$$

Это означает, что гессиан - положительно определенная матрица. В таком случае все его собственные числа положительны.

Пусть имеется два различных собственных числа λ_1 и λ_2 и соответствубщие им собственные вектора x_1, x_2. Если $x_1 \neq x_2$, то любой вектор z пространства можно разложить по базису x_1, x_2. (Матрица Гессе- симметрическая и положительная определенная, поэтому ортогональность собственных векторов обеспечена) (Что делать если $x_1 = x_2$ - далее)

Тогда для собственного вектора x (x_1 или x_2):

$$x^T H x = x^T(Hx) = x^T \lambda x = \lambda x^T x = \lambda ||x||^2$$

Теперь возьмем произвольный z:

$$z^T H z = (\alpha_1 x_1 + \alpha_2 x_2)^T H(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1^2 \lambda_1 ||x_1||^2 + \alpha_2^2 \lambda_2 ||x_2||^2 + 2\alpha_1\alpha_2\lambda_1\lambda_2(x_1, x_2)$$

Но $(x_1, x_2) = 0$, так как это базисные вектора. Получаем:

$$z^T H z = \alpha_1^2 \lambda_1 ||x_1||^2 + \alpha_2^2 \lambda_2 ||x_2||^2$$

Но

$$||z|| = ||\alpha_1 x_1 + \alpha_2 x_2|| = \alpha_1^2 ||x_1||^2 + \alpha_2 ||x_2||^2$$

Отсюда получаем, что для конкретной матрицы Гессе выполнено:

$$\lambda_{min} ||x||^2 \leq x^T H(y) x \leq \lambda_{max} ||x||^2$$

Остается вопрос: возможно ли сделать так, чтобы констатны в левой и правой части неравентсва не зависили от конкретных x_1 и x_2?

Найдём наименьшие и наибольшие собственные числа гессиана, то есть решим:

$$|H - \lambda E| = 0$$

$$\begin{vmatrix} \frac{12(1+x_2^2)}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}} - \lambda & \frac{-12x_1x_2}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}} \\ \frac{-12x_1x_2}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}} & \frac{4(1+3x_1^2)}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}} - \lambda \end{vmatrix} = 0 \Leftrightarrow \frac{48}{\left(1+3x_1^2+x_2^2\right)^2} + \lambda^2 - 4\lambda \frac{3(1+x_2^2)+1+3x_1^2}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}} = 0 \Leftrightarrow$$

$$\Leftrightarrow (1+3x_1^2+x_2^2)^2\lambda^2 - 4\lambda(4+3x_1^2+3x_2^2)\cdot\sqrt{1+3x_1^2+x_2^2} + 48 = 0$$

$$D = 4(4+3x_1^2+3x_2^2)^2(1+3x_1^2+x_2^2) - 48(1+3x_1^2+x_2^2) =$$

$$= 4(1+3x_1^2+x_2^2)(16+9x_1^4+9x_2^4+24x_1^2+24x_2^2+18x_1^2x_2^2-12-36x_1^2-12x_2^2) =$$

$$= 4(1+3x_1^2+x_2^2)(4+9x_1^4+9x_2^4+18x_1^2x_2^2-12x_1^2+12x_2^2)$$

$$\lambda = \Big(2(4+3x_1^2+3x_2^2)\sqrt{1+3x_1^2+x_2^2}\pm$$

$$\pm 2\sqrt{1+3x_1^2+x_2^2}\cdot\sqrt{4+9x_1^4+9x_2^4+18x_1^2x_2^2-12x_1^2+12x_2^2}\Big)\cdot\left(1+3x_1^2+x_2^2\right)^{-2} =$$

$$= \frac{2}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}}\Big(4+3x_1^2+3x_2^2\pm\sqrt{4+9x_1^4+9x_2^4+18x_1^2x_2^2-12x_1^2+12x_2^2}\Big) =$$

$$= \frac{2}{(1+3x_1^2+x_2^2)^{\frac{3}{2}}}\Big(4+3x_1^2+3x_2^2\pm\sqrt{9x_2^4+18x_1^2x_2^2+12x_2^2+(3x_1^2-2)^2}\Big)$$

Что же мы видим? В ограниченной области наименьшее и наибольшее собственные значения сущетсвуют! То есть среди всех собственных чисел нет неограниченных. (знаменатель не обращается в ноль, под корнем всегда неотрицательные значения) Поэтому выберем

$$m = inf(\lambda_{min})$$
$$M = sup(\lambda_{max})$$

Итак, мы оказались в условиях теоремы. **Поэтому градиентный метод первого порядка можно применять!**

P.S Внимательный читатель вспомнит о предположении $x_1 \neq x_2$ для собственных векторов. Собственные вектора, соответствующие разным собственным числам, различны. А равенство собственных значений соответствует единственно возможной конфигурации: $x_1 = x_2 = 0$. В этом случае к тем значениям, по которым мы выбираем m и M стоит просто добавить ноль. Проблема оказывается решенной.

Theorem 2 *Если функция $f(x)$ дважды непрерывно дифференцируема и существуют такие числа $0 < m \leq M$, что*

$$m||x||^2 \leq x^T H(y) x \leq M||x||^2, \quad \forall\ x, y$$

тогда для классического метода Ньютона:$x_k \longrightarrow x^*$, $f(x_k) \longrightarrow f(x^*)$ при любой начальной точке x_0, где x^* - оптимальная.

Но константы m и M уже найдены (повторим, в ограниченной области!). Условие теоремы выполнено. **Градиентный метод второго порядка можно применять.**

3 Описание алгоритмов

3.1 Метод наискорейшего спуска

1. **Input**:
   ε - требуемая точность, x_0 - начальное приближение, $0 < \alpha_0 < 1$
   $k = 0$

2. Вычисляем $\nabla f(x_k)$

3. Подбираем шаг $(0 < \alpha_k < 1)$, решая задачу одним из методов одномерной минимизации для задачи $\min \left\{ f\Big(x_k - \alpha_k \nabla f(x_k)\Big) \right\}$

4. Вычисляем $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$
   k -> k + 1

5. Завершить процесс, если $||\nabla f(x_k)|| < \varepsilon$. Иначе перейти к шагу 2.

3.2 Метод Девидона - Флетчера - Пауэлла

1. **Input**:
   ε - требуемая точность, x_0 - начальное приближение, $A_1 = E$, E - соответствующая единичная матрица, $I_0 = \{n, 2n, ...\}$ - множество моментов обновления алгоритма,
   $\omega_1 = -\nabla f(x_0)$
   $k = 1$

2. Находим направление спуска $p_k = A_k \omega_k$

3. Подбираем шаг $\alpha_k : min\{\phi_k(\alpha)\} = f(x_{k-1} + \alpha_k p_k)$ одним из методов одномерной минимизации

4. Вычислим точку $x_k = x_{k-1} + \alpha_k p_k$ и $\omega_{k+1} = -\nabla f(x_k)$

5. Если $k \in I_0$, то положить $A_{k+1} = E$ и вернуться к шагу 2

6. Положим $\Delta x_k = x_k - x_{k-1}$ и $\Delta\omega_k = \omega_{k+1} - \omega_k$

7. Построим матрицу A_{k+1} по формуле:
   $$A_{k+1} = A_k - \frac{\Delta x_k (\Delta x_k)^T}{\Delta\omega_k^T \Delta x_k} - \frac{A_k \Delta\omega_k (\Delta\omega_k)^T A_k^T}{\Delta\omega_k^T A_k \Delta\omega_k} \tag{1}$$

8. Завершить процесс, если $||\omega_k|| < \varepsilon$. Иначе перейти к шагу 2.

4 Результаты

градиентный метод первого порядка		град. метод 2 порядка

Точность 0.1

GOLDEN SEARCH:	TPM:	DFP
[3, 4]	[3, 4]	[3, 4]
[-1.55126, 2.35271]	[-1.7136, 2.29395]	[-1.7136, 2.29395]
[-0.72663, -0.50919]	[-0.61944, -0.54405]	[-1.66665, -0.853]
[-0.43769, -0.4579]	[-0.40801, -0.43404]	[-0.84596, -1.05705]
[-0.46063, -0.33287]	[-0.45861, -0.36046]	[-0.39223, -0.75432]
[-0.42594, -0.33148]	[-0.4264, -0.34153]	[-0.36246, -0.37313]
		[-0.43449, -0.33944]

Точность 0.01

GOLDEN SEARCH:	TPM:	DFP
[3, 4]	[3, 4]	[3, 4]
[-1.66064, 2.31312]	[-1.67678, 2.30728]	[-1.67678, 2.30728]
[-0.6532, -0.53143]	[-0.64699, -0.52361]	[-1.67148, -0.78955]
[-0.42608, -0.44836]	[-0.4267, -0.44442]	[-0.823, -1.06991]
[-0.46117, -0.35202]	[-0.45968, -0.3511]	[-0.36546, -0.71473]
[-0.42776, -0.34001]	[-0.42818, -0.33949]	[-0.37128, -0.33107]
[-0.43328, -0.32602]	[-0.43305, -0.32559]	[-0.43004, -0.33732]
[-0.42866, -0.32423]	[-0.42875, -0.32405]	[-0.42892, -0.32163]

Точность 0.001

GOLDEN SEARCH:	TPM:	DFP
[3, 4]	[3, 4]	[3, 4]
[-1.67314, 2.3086]	[-1.67217, 2.30895]	[-1.67217, 2.30895]
[-0.64836, -0.52538]	[-0.64867, -0.526]	[-1.67201, -0.79149]
[-0.42613, -0.44515]	[-0.4261, -0.44542]	[-0.82225, -1.06806]
[-0.4602, -0.35135]	[-0.4603, -0.35139]	[-0.36585, -0.71511]
[-0.42808, -0.33966]	[-0.42806, -0.33969]	[-0.37139, -0.33068]
[-0.43313, -0.32568]	[-0.43314, -0.32569]	[-0.43003, -0.33716]
[-0.42873, -0.32409]	[-0.42873, -0.3241]	[-0.42892, -0.32165]
[-0.42942, -0.32218]	[-0.42943, -0.32219]	
[-0.42883, -0.32197]	[-0.42883, -0.32197]	
[-0.42892, -0.32171]	[-0.42892, -0.32171]	

Точность 0.0001

GOLDEN SEARCH:	TPM:	DFP
[3, 4]	[3, 4]	[3, 4]
[-1.67365, 2.30842]	[-1.67361, 2.30843]	[-1.67361, 2.30843]
[-0.64806, -0.52544]	[-0.64807, -0.52546]	[-1.67184, -0.79213]
[-0.42622, -0.44513]	[-0.4262, -0.44513]	[-0.82145, -1.0686]
[-0.46019, -0.35122]	[-0.46019, -0.35124]	[-0.36521, -0.7141]
[-0.42807, -0.3396]	[-0.42807, -0.33961]	[-0.37157, -0.32958]
[-0.43311, -0.32568]	[-0.43312, -0.32568]	[-0.42991, -0.33693]
[-0.42873, -0.32409]	[-0.42873, -0.32409]	[-0.42892, -0.32164]
[-0.42942, -0.32218]	[-0.42942, -0.32218]	[-0.42885, -0.32163]
[-0.42883, -0.32197]	[-0.42883, -0.32197]	
[-0.42892, -0.32171]	[-0.42892, -0.32171]	
[-0.42884, -0.32168]	[-0.42884, -0.32168]	
[-0.42886, -0.32164]	[-0.42886, -0.32164]	

5 Обоснование достоверности полученного решения

5.1 Теоретическая оценка алгоритмов

1. Градиентный метод первого порядка:
   Целевая функция удовлетворяет условиям теоремы из Исследований приметимости, и шаг выбирается оптимальным. Поэтому последовательность x_k будет сходиться к точке минимума со скоростью геометрической прогрессии. Или, что то же, скорость сходимости предполагается линейной.

2. Градиентный метод второго порядка (ДФП)
   Метод Ньютона с оптимальным выбором шага, удовлетворяющий условию теоремы, будет сходиться к точке минимума с квадратичной скоростью:

$$||x_{k+1} - x^*|| \leq C||x_k - x^*||^2$$

5.2 Сравнительный анализ

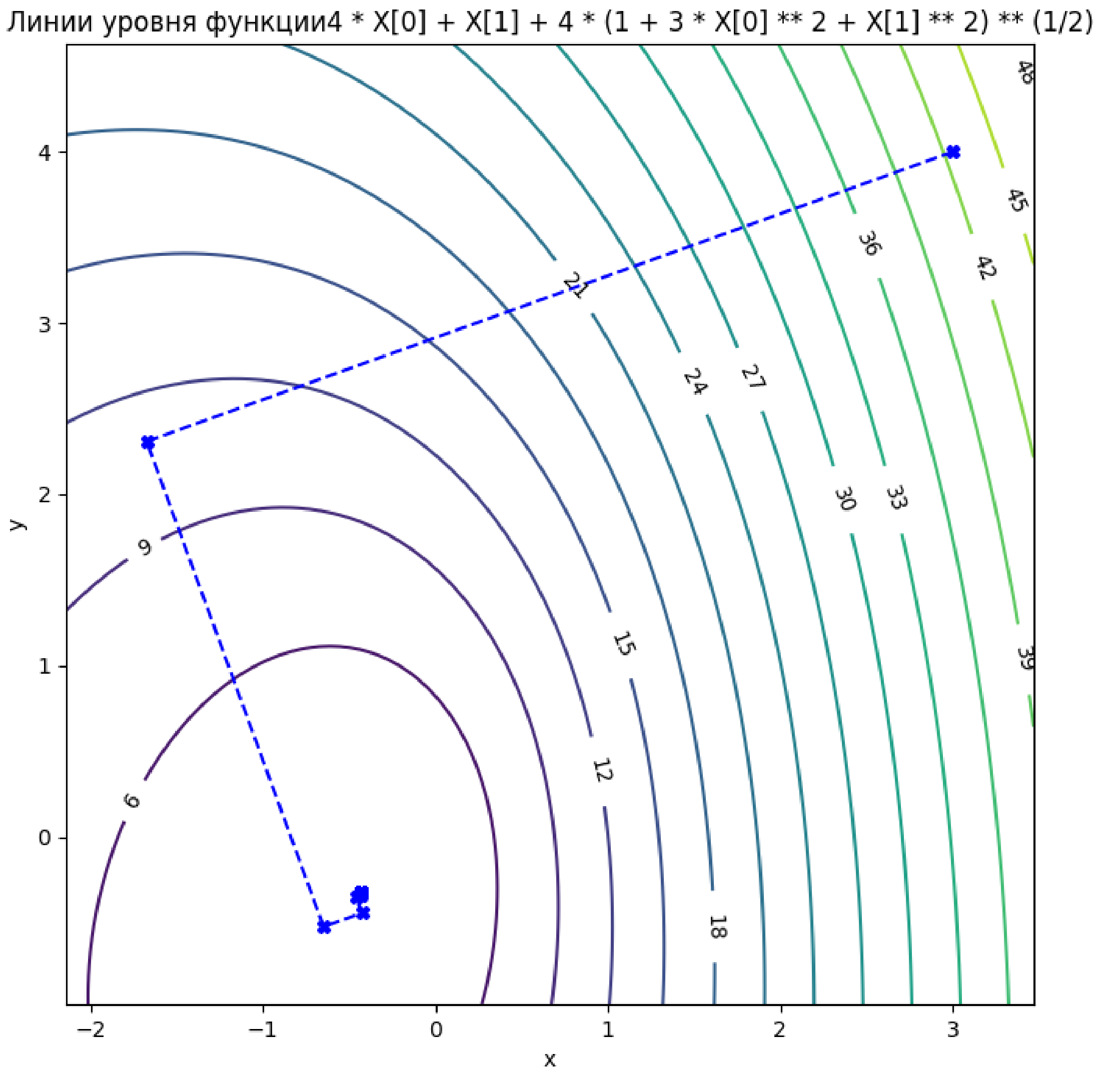

Рис. 1: Точки, соответствующие итерациям метода градиентного спуска первого порядка с оптимальным выбором шага (метод золотого сечения)

Особенно заметна на графике ортогональность соседних направлений метода. Действительно, если на k-м шаге мы двигались в направлении $\nabla f(x_{k-1})$ и достигли минимума по этому направлению (согласно принципу выбора шага), то на следующей итерации мы по направлению $\nabla f(x_{k-1})$ будем иметь нулевую производную, и следовательно, в следующий раз выберем направление, перпендикулярное текущему.

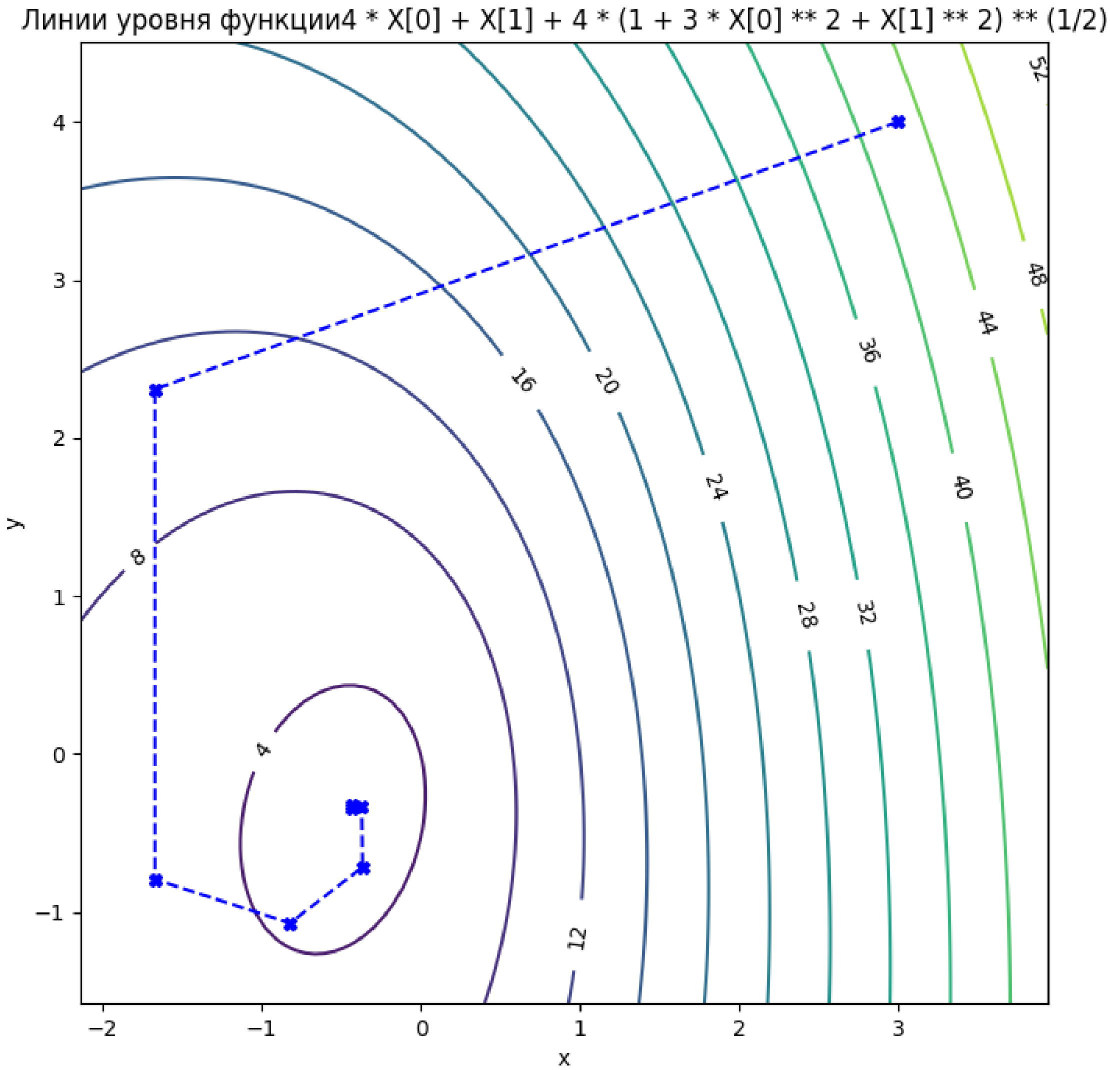

Рис. 2: Точки, соответствующие итерациям метода градиентного спуска второго порядка ДФП

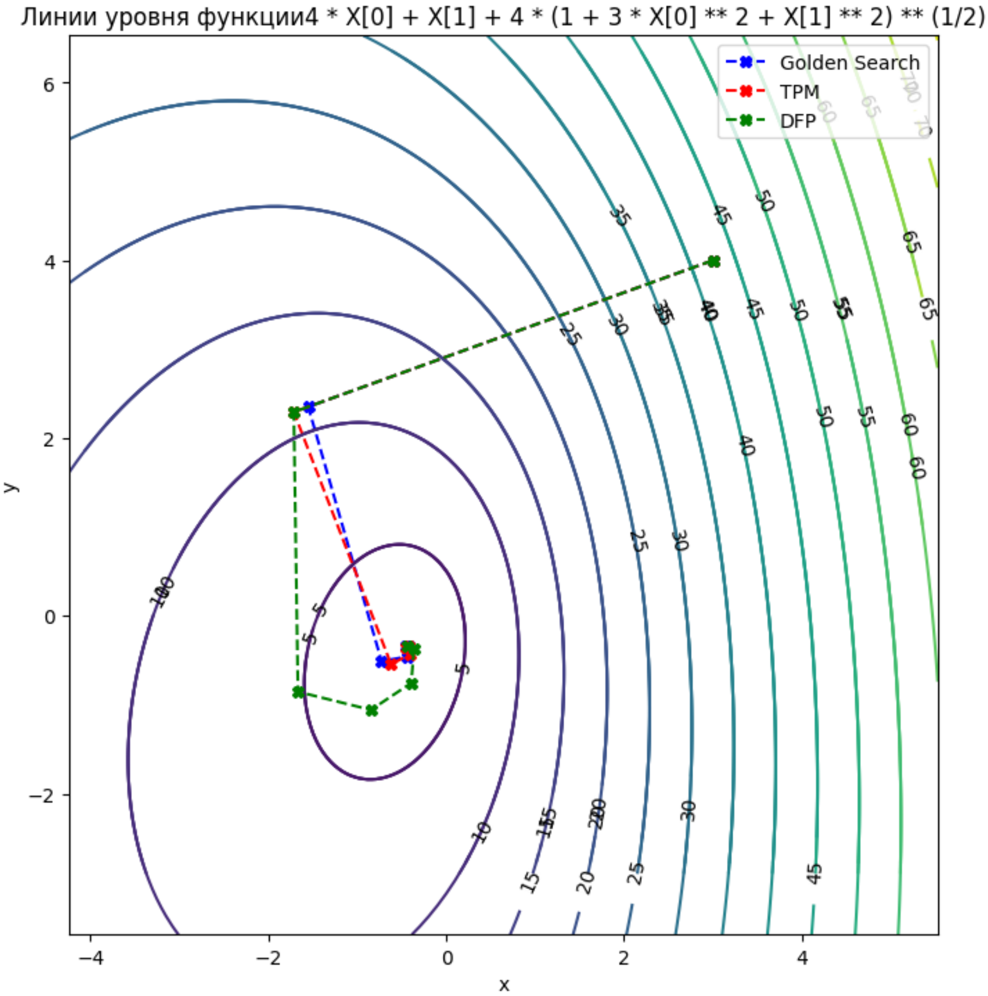

Рис. 3: Точки, соответствующие итерациям методов градиентного спуска первого порядка с оптимальным выбором шага (метод золотого сечения и пробных точек) и метода градиентного спуска второго порядка ДФП

Заметим, что метод DFP на первых шагах не так хорошо сходится к решению, как это делает метод первого порядка. Однако в блоке с численным решением число итераций по методу DFP было меньше. Отсюда следует предположение, что пусть вдали метод второго второго порядка сходится и не так быстро, в окрестности решения скорость сходимости становится ощутимо большой.

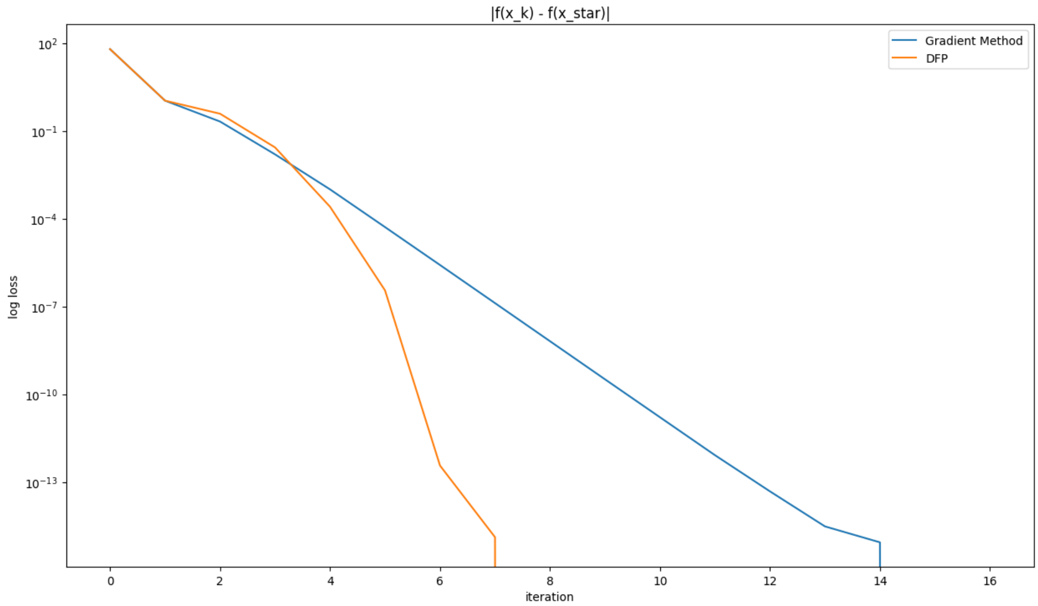

Рис. 4: График зависимости скорости сходимости градиентного метода и метода DFP от числа итераций.

Эксперимент ставится следующим образом: вычисляется решение задачи с точностью 10^{-14}. Далее на каждом шагу определяется разность между текущим значением функции и "точным"значением. Начальное приближение - $(x_1 = 6, x_2 = -2)$. В качестве критерия остановки использовалось неравенство: $\|\nabla f(x_k)\| < 1 * 10^{-8}$.
Хорошо видно, что градиентный метод наскоирейшего спуска обладает линейной скоростью сходимости. С методом DFP всё интереснее. Первые 4 итерации метод обладает линейной скоростью сходимости, но после того, как мы попадаем в некоторую окрестность оптимальной точки, метод начинает сходиться с квадратичной скоростью. Из этого можно сделать интересный вывод: Хорошим тоном будет комбинировать эти два метода. Т.е, первые несколько итераций делать градиентным методом, т.к затраты по времени и памяти для его реализации значительно меньше чем для DFP, а в скорости мы не проигрываем. После достижения "некоторой"близости к точке оптимума, можем использовать метод второго порядка, получая квадратичную скорость сходимости. (Машинный ноль на последних итерациях обусловлен тем, что "точное"значение функции вычисляется с помощью тех же методов, что мы рассматриваем на графиках)

6 Выводы

Бесхитростные идеи, лежащие в основе градиентных методов, нуждаются в весьма объемных исследованиях применимости. Тем не менее, если этот этап пройден, оба метода начинают показывать хорошие результаты. Стоит сделать важное замечание: метод ДФП вдали от точки оптимума скорее всего будет иметь линейную скорость сходимость, однако требовать более высоких вычислительных ресурсов, чем метод градиентного спуска первого порядка. В то же время вблизи точки оптимума метод ДФП имеет скорость сходимости порядка двух. Это вселяет надежду и дает право на высказывание некоторй рекомендации: на первых шагах использовать градиентный метод первого порядка, далее же переходить к методу ДФП.

7 Приложения

Реализация программы находится в репозитории GitHub по ссылке:
`https://github.com/IMZolin/multi-dimension-minimization`

8 Библиографический список

1. Кормен, Томас Х., Лейзерсон, Чарльз И., Ривест, Рональд Л., Штайн, Клиффорд. "Алгоритмы. Построение и анализ, 2-е издание"Издательский дом “Вильямс”, 2011. – 892–918 с.

   URL: `https://vk.com/doc191450968_561608466?hash=HUwStWS0yzrW9SaXn8POZtaz3gTyMTm` `dl=U9ivclLJBeeYQbs3MMhGtwYZ7Mx4nGJelTv0Hv56E4z` / [Электронный ресурс]. Режим доступа: (Дата обращения: 04.03.2023)

2. Родионова Е.А., Петухов Л.В., Серёгин Г.А. "Методы оптимизации. Задачи выпуклого программирования"Издательство Политехнического университета, Санкт-Петербург, 2014

   URL: `https://elib.spbstu.ru/dl/2/i17-98.pdf/info` / [Электронный ресурс]. Режим доступа: (Дата обращения: 10.03.2023)

3. Моисеев Н.Н. "Методы оптимизации"

   URL: `https://avidreaders.ru/book/metody-optimizacii-1.htmlPк`/ [Электронный ресурс]. Режим доступа: (Дата обращения: 07.03.2023)