Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
**Высшая школа прикладной математики и вычислительной физики**

# Лабораторная работа №4

**Тема: «Решение задач многомерной минимизации без ограничений»**

по дисциплине

«Методы оптимизаций»

Выполнили студенты группы 3630102/80401: Мамаева Анастасия Сергеевна, Веденичев Дмитрий Александрович, Тырыкин Ярослав Алексеевич

Руководитель Доцент, к.ф.-м.н.: Родионова Елена Александровна

Санкт-Петербург
2021

# СОДЕРЖАНИЕ

**1 Постановка задачи** . . . . . . . . . . **3**

**2 Применимость методов** . . . . . . . . . . **3**

**3 Описание алгоритмов** . . . . . . . . . . **8**

3.1 Градиентный метод наискорейшего шага . . . . . . . . . . 8

3.2 Метод золотого сечения . . . . . . . . . . 8

3.3 Метод Ньютона с постоянным шагом . . . . . . . . . . 8

3.4 Метод Бройдена — Флетчера — Гольдфарба — Шанно с постоянным шагом . . . . . 9

**4 Результаты решения задачи** . . . . . . . . . . **10**

4.1 Численное решение задачи . . . . . . . . . . 10

4.2 Линии уровня, рельефы . . . . . . . . . . 10

**5 Сравнительный анализ** . . . . . . . . . . **13**

**6 Проведение сравнительных экспериментов** . . . . . . . . . . **13**

**7 Оценка достоверности полученного результата** . . . . . . . . . . **15**

**8 Дополнительные исследования** . . . . . . . . . . **16**

8.1 Уточняющие детали градиентного спуска . . . . . . . . . . 16

8.2 Вариативность условия остановки методов . . . . . . . . . . 16

**9 Программная реализация** . . . . . . . . . . **17**

**Список литературы** . . . . . . . . . . **17**

# 1 Постановка задачи

Пусть имеется задача многомерной минимизации :

$$f(x) = 2x_1^2 + x_2^2 + cos(6x_1 + 5x_2) - x_1 + 2x_2 \longrightarrow \min_x$$

Необходимо:

1. Решить данную задачу градиентным методом наискорейшего спуска с шагом по методу золотого сечения.
2. Обосновать выбор условия окончания вычислений для метода одномерной минимизации по соответствующему условию градиентного метода. Показать справедливость своего вывода в ходе вычислительного эксперимента (для точности градиентного метода 0,01).
3. Решить данную задачу методом Ньютона с постоянным шагом.
4. Решить данную задачу методом Бройдена — Флетчера — Гольдфарба — Шанно с постоянным шагом.
5. Выполнить сравнительный анализ алгоритмов методов.
6. Нарисовать линии уровня функции цели и показать в ходе вычислительного эксперимента ортогональность звеньев градиентной ломаной для метода наискорейшего спуска.

# 2 Применимость методов

Алгоритмы безусловной минимизации делятся на классы, в зависимости от порядка привлекаемых производных:

- Нулевого порядка – без привлечения производных
- Первого порядка – с вычислением первой производной
  ...
- $p$–ого порядка - вычисление производных $p$–ого порядка

В рамках нашего курса остановимся на методах первого и второго порядков. Для того, чтобы градиентный метод наискорейшего спуска сходился, необходимо выполнение следующей теоремы:

**Теорема 1.** *Если $f(x)$– дифференцируемая функция, ограниченная снизу, её градиент удовлетворяет условию Липшица*

$$||\nabla f(x) - \nabla f(y)|| \leq R||x - y||, \quad R > 0$$

*то будет выполняться $||\nabla f(x_k)|| \longrightarrow 0$ при $k \to \infty$ при любой начальной точке $x_0$.*

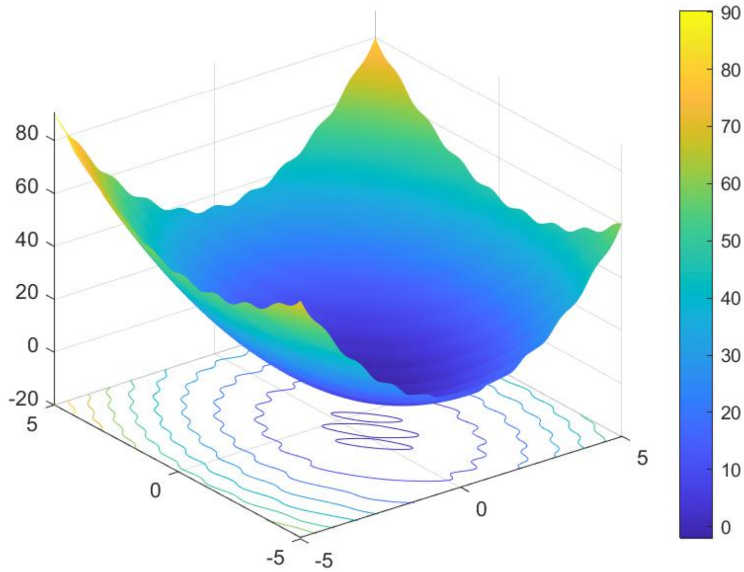
**Теорема 2.** *Если кроме того, функция $f(x)$ дважды непрерывно дифференцируема и существуют такие числа $0 < m \le M$, что*

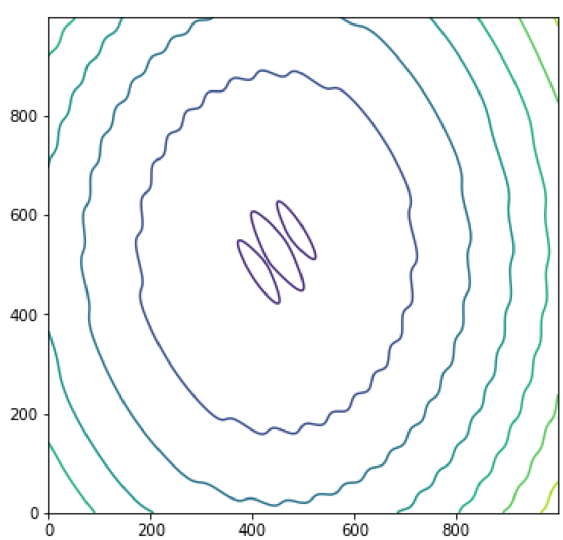
$$m||x||^2 \le x^T H(y) x \le M||x||^2, \quad \forall\ x, y$$

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

Существует аналогичная теорема о сходимости методов второго порядка. Единственное отличие в том, что сходиться алгоритм будет со сверхлинейной скоростью.

С помощью пакета MATLAB2020b построим график функции:




Видим, что у нас неупорядоченный тип рельефа, в котором имеется несколько минимумов.

Исследуем нашу функцию:

покажем, что она действительно ограничена снизу:

$$f(x_1, x_2) = 2x_1^2 + x_2^2 + \cos(6x_1+5x_2) - x_1 + 2x_2 = \underbrace{2x_1^2 - x_1 + \tfrac{1}{8}} - \tfrac{1}{8} + \underbrace{x_2^2 + 2x_2 + \tfrac{1}{4}} - \tfrac{1}{4} + \cos(6x_1 +$$
$$+ 5x_2) = \left(2x_1^2 - x_1 + \tfrac{1}{8}\right) - \tfrac{1}{8} + \left(x_2^2 + 2x_2 + \tfrac{1}{4}\right) - \tfrac{1}{4} + \cos(6x_1 + 5x_2) =$$
$$= \left(\sqrt{2}x_1 - \tfrac{1}{2\sqrt{2}}\right)^2 - \tfrac{1}{8} + \left(x_2 + \tfrac{1}{2}\right)^2 - \tfrac{1}{4} + \cos(6x_1 + 5x_2) =$$
$$= \underbrace{\left(\sqrt{2}x_1 - \tfrac{1}{2\sqrt{2}}\right)^2}_{\geq 0} + \underbrace{\left(x_2 + \tfrac{1}{2}\right)^2}_{\geq 0} + \cos(6x_1 + 5x_2) - \tfrac{3}{8} \geq -2\tfrac{3}{8}$$

видим, что каждое слагаемое ограничено, таким образом наша функция ограничена снизу. Её значение во всех точках не меньше $-2\frac{3}{8}$.

Теперь вычислим градиент функции и покажем, что он удовлетворяет условию Липшица:

$$\|\nabla f(x) - \nabla f(y)\| \leq R\|x - y\| \quad , \quad R > 0.$$

Считаем частные производные:

$$\frac{\partial f}{\partial x_1} = 4x_1 - 6\sin(6x_1 + 5x_2) - 1$$
$$\frac{\partial f}{\partial x_2} = 2x_2 - 5\sin(6x_1 + 5x_2) + 2 \qquad \Rightarrow \qquad \nabla f(x) = \begin{pmatrix} 4x_1 - 6\sin(6x_1 + 5x_2) - 1 \\ 2x_2 - 5\sin(6x_1 + 5x_2) + 2 \end{pmatrix}$$

Пусть $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ , $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}$ , тогда:

$$\|\nabla f(x) - \nabla f(y)\| = \left| 4x_1 + 2x_2 - 11\sin(6x_1+5x_2) - (4y_1 + 2y_2 - 11\sin(6y_1 + 5y_2)) \right| =$$
$$= \left| 4x_1 + 2x_2 - 11\sin(6x_1 + 5x_2) - 4y_1 - 2y_2 + 11\sin(6y_1 + 5y_2) \right| =$$
$$= \left| (4x_1 + 2x_2 - 4y_1 - 2y_2) + (11\sin(6y_1 + 5y_2) - 11\sin(6x_1 + 5x_2)) \right| \leq$$
$$\leq \left| 4x_1 + 2x_2 - 4y_1 - 2y_2 \right| + \left| 11\sin(6y_1 + 5y_2) - 11\sin(6x_1 + 5x_2) \right| =$$
$$= \left| 4(x_1 - y_1) + 2(x_2 - y_2) \right| + 11\left| \sin(6y_1 + 5y_2) - \sin(6x_1 + 5x_2) \right| =$$
$$= 2\left| 2(x_1 - y_1) + (x_2 - y_2) \right| + 11\left| \sin(6y_1 + 5y_2) - \sin(6x_1 + 5x_2) \right|$$

заменим $a = 6y_1 + 5y_2$ , $b = 6x_1 + 5x_2$

распишем разность синусов:

$$|\sin(a) - \sin(b)| = 2\left|\sin\frac{a-b}{2}\cos\frac{a+b}{2}\right| \leq 2\left|\sin\frac{a-b}{2}\right| \leq |a - b| =$$
$$= |6y_1 + 5y_2 - 6x_1 + 5x_2| = |6(y_1 - x_1) + 5(y_2 - x_2)|$$

Значит имеем:

$$\|\nabla f(x) - \nabla f(y)\| \leq 2|2(x_1 - y_1) + (x_2 - y_2)| + 11|6(y_1 - x_1) + 5(y_2 - x_2)| \leq 4|\Delta\vec{x}| + 121|\Delta\vec{x}| =$$
$$= 125|\Delta\vec{x}|$$

$$\Rightarrow \quad \exists R = 125, \quad 125 > 0.$$

Построим матрицу Гессе, и найдём область, где она положительна.

$$H(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix}$$

$$\frac{\partial^2 f}{\partial x_1^2} = 4 - 6\cdot 6\cos(6x_1 + 5x_2) = 4 - 36\cos(6x_1 + 5x_2)$$

$$\frac{\partial^2 f}{\partial x_2^2} = 2 - 5\cdot 5\cos(6x_1 + 5x_2) = 2 - 25\cos(6x_1 + 5x_2)$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = -6\cdot 5\cos(6x_1 + 5x_2) = -30\cos(6x_1 + 5x_2)$$

$$H(f) = \begin{pmatrix} 4 - 36\cos(6x_1 + 5x_2) & -30\cos(6x_1 + 5x_2) \\ -30\cos(6x_1 + 5x_2) & 2 - 25\cos(6x_1 + 5x_2) \end{pmatrix}$$

Для определения положительности $H(f)$ воспользуемся критерием Сильвестра:

$$\begin{cases} 4 - 36\cos(6x_1 + 5x_2) > 0 \\ \det(H) > 0 \end{cases}$$

$$\begin{cases} \det(H) = \left(4 - 36\cos(6x_1 + 5x_2)\right)\cdot\left(2 - 25\cos(6x_1 + 5x_2)\right) - 30^2\cdot\cos^2(6x_1 + 5x_2) > 0. \\ 4 - 36\cos(6x_1 + 5x_2) > 0 \end{cases}$$

$$\begin{cases} 4 - 36\cos(6x_1 + 5x_2) > 0 \\ 8 - 92\cos(6x_1 + 5x_2) > 0 \end{cases} \Leftrightarrow \begin{cases} -36\cos(6x_1 + 5x_2) > -4 \\ -92\cos(6x_1 + 5x_2) > -8 \end{cases} \Leftrightarrow \begin{cases} \cos(6x_1 + 5x_2) < \frac{4}{36} \\ \cos(6x_1 + 5x_2) < \frac{8}{92} \end{cases}$$

$$\Leftrightarrow \begin{cases} \cos(6x_1 + 5x_2) < \frac{1}{9} \\ \cos(6x_1 + 5x_2) < \frac{2}{23} \end{cases} \Rightarrow \boxed{\cos(6x_1 + 5x_2) < \frac{2}{23}}$$

Теперь нам необходимо найти числа $m$ и $M$, удовлетворяющие условиям:

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

Чтобы это сделать, необходимо вычислить наибольшее и наименьшее значение собственных чисел гессиана $H(y)$.

$M$ будет равняться наибольшему собственному числу, $m$ - наименьшему.

Найдём их, перед этим решив уравнение $|H(y) - \lambda E| = 0$

$$\begin{vmatrix} 4 - 36\cos(6x_1 + 5x_2) - \lambda & -30\cos(6x_1 + 5x_2) \\ -30\cos(6x_1 + 5x_2) & 2 - 25\cos(6x_1 + 5x_2) - \lambda \end{vmatrix} = 0$$

$$\left(4 - 36\cos(6x_1 + 5x_2) - \lambda\right)\left(2 - 25\cos(6x_1 + 5x_2) - \lambda\right) - 900\cos^2(6x_1 + 5x_2) = 0.$$

$$8 - 100\cos(6x_1 + 5x_2) - 4\lambda - 72\cos(6x_1 + 5x_2) + 900\cos^2(6x_1 + 5x_2) + 36\lambda\cos(6x_1 + 5x_2) - 2\lambda + 25\lambda\cos(6x_1 + 5x_2) + \lambda^2 - 900\cos^2(6x_1 + 5x_2) = 0.$$

$$\lambda^2 + 61\lambda\cos(6x_1 + 5x_2) - 6\lambda - 172\cos(6x_1 + 5x_2) + 8 = 0$$

получим квадратное уравнение относительно $\lambda$.

$$\lambda^2 + \lambda\left(61\cos(6x_1 + 5x_2) - 6\right) - \left(172\cos(6x_1 + 5x_2) - 8\right) = 0.$$

$D = b^2 - 4ac \quad ; \qquad x = \dfrac{-b \pm \sqrt{D}}{2a}$

$D = \left[61\cos(6x_1+5x_2) - 6\right]^2 + 4\cdot 1\cdot\left[172\cos(6x_1+5x_2) - 8\right] =$

$= 3721\cos^2(6x_1+5x_2) - 732\cos(6x_1+5x_2) + 36 + 688\cos(6x_1+5x_2) - 32 =$

$= 3721\cos^2(6x_1+5x_2) - 44\cos(6x_1+5x_2) + 4$

$$\lambda = \frac{6 - 61\cos(6x_1+5x_2) \pm \sqrt{3721\cos^2(6x_1+5x_2) - 44\cos(6x_1+5x_2) + 4}}{2}$$

Для того, чтобы найти наибольшее и наименьшее значение $\lambda$, продифференцируем его по $\cos(6x_1+5x_2)$

$$\frac{d\lambda}{d\cos(6x_1+5x_2)} = -\frac{61}{2} \pm \frac{7442\cos(6x_1+5x_2) - 44}{2\sqrt{3721\cos^2(6x_1+5x_2) - 44\cos(6x_1+5x_2) + 4}}$$

Для положительной определённости необходимо условие: $-1 \le \cos(6x_1+5x_2) < \frac{2}{23}$

Проводя исследование на участке $[-1; \frac{2}{23})$ заметим, что производная строго отрицательна в обоих случаях. Чтобы получить наибольшее и наименьшее значение $\lambda$ подставим значения на концах интервала.

$\cos(6x_1+5x_2) = -1$ и $\cos(6x_1+5x_2) = \frac{2}{23}$

- Если $\underbrace{\cos(6x_1+5x_2)}_{=-1}$, то $\lambda_{1,2} = \dfrac{6 - 61\cdot(-1) \pm \sqrt{3721\cdot(-1)^2 - 44\cdot(-1) + 4}}{2} = \dfrac{6+61 \pm \sqrt{3769}}{2} =$

$= \dfrac{67 \pm \sqrt{3769}}{2} = \dfrac{67 \pm 61{,}392}{2} \Rightarrow \left[\begin{array}{l}\lambda_1 = 64{,}196 \\ \lambda_2 = 2{,}804\end{array}\right.$

- Если $\cos(6x_1+5x_2) = \frac{2}{23}$, то

$\lambda_{3,4} = \dfrac{-61\cdot\frac{2}{23} + 6 \pm \sqrt{3721\cdot(\frac{2}{23})^2 - 44\cdot\frac{2}{23} + 4}}{2} = \dfrac{-\frac{122}{23} + \frac{138}{23} \pm \sqrt{3721\cdot\frac{4}{529} - \frac{88}{23} + \frac{92}{23}}}{2} =$

$= \dfrac{\frac{16}{23} \pm \sqrt{\frac{14884}{529} - \frac{88}{23} + \frac{92}{23}}}{2} = \dfrac{\frac{16}{23} \pm \sqrt{\frac{14884 - 2024 + 2116}{529}}}{2} = \dfrac{\frac{16}{23} \pm \sqrt{\frac{14976}{529}}}{2} =$

$= \dfrac{\frac{16}{23} \pm \frac{122{,}376}{23}}{2} \Rightarrow \left[\begin{array}{l}\lambda_3 = \left(\frac{16}{23} + \frac{122{,}376}{23}\right) : 2 \\ \lambda_4 = \left(\frac{16}{23} - \frac{122{,}376}{23}\right) : 2\end{array}\right. \Rightarrow \left[\begin{array}{l}\lambda_3 = 3{,}008 \\ \lambda_4 = -2{,}313\end{array}\right.$

Тогда имеем, $m = 2{,}804$, $M = 64{,}196$.

Данными выкладками показали справедливость теорем для методов первого и второго порядков.

Чем ближе $m$ и $M$ друг к другу, тем быстрее метод сходится к $x^*$.
Иными словами $q = \dfrac{M-m}{M+m} = \dfrac{64{,}196 - 2{,}804}{64{,}196 + 2{,}804} = \dfrac{61{,}392}{67} = 0{,}916$

У нас задача плохо обусловлена, значит метод будет сходиться медленно.

# 3 Описание алгоритмов

## 3.1 Градиентный метод наискорейшего шага

1. Начальный этап:

   I Выберем $\varepsilon > 0, \quad 0 < \alpha_0 < 1, \quad x_0$ – начальное приближение

   II Положим $k = 0$

2. Основной этап:

   I Вычисляем $\nabla f(x_k), \quad \alpha_k = \alpha_0$

   II Подбор шага осуществляем наилучшим способом, то есть так, чтобы минимизировать разность между значением функции в следующей точке и значением функции в предыдущей точке: $min f(x_k - \alpha_k \nabla f(x_k)), \quad 0 < \alpha_k < 1$
   Для решения этой задачи можем использовать любой метод одномерной минимизации.

   III Полагаем $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$, заменяем $k$ на $k + 1$, переходим к шагу 2.I

3. Условие окончания вычислений: $||\nabla f(x_k)|| < \varepsilon$

В данном алгоритме есть необходимость использования одномерной минимизации, для поиска подходящего $\alpha_k$. Для этого воспользуемся методом золотого сечения.

## 3.2 Метод золотого сечения

1. Найдём $x_1 = a + \frac{3-\sqrt{5}}{2}(b - a), \quad x_2 = a + \frac{\sqrt{5}-1}{2}(b - a)$. Вычислим $f(x_1)$ и $f(x_2)$. Положим $\tau = \frac{\sqrt{5}-1}{2}, \quad \varepsilon_n = b - a$.

2. Проверка на окончание поиска: если $\varepsilon_n > \varepsilon$, то перейдём к шагу 3, иначе – к шагу 4.

3. Переходим к новому отрезку и новым пробным точкам. Если $f(x_1) \leq f(x_2)$, то положим $b = x_2, \ x_2 = x_1, \ f(x_2) = f(x_1), \ x_1 = b - \tau(b - a)$ и вычислим $f(x_1)$, иначе – положим $a = x_1, \ x_1 = x_2, \ f(x_1) = f(x_2), \ x_2 = a + \tau(b - a)$ и вычислим $f(x_2)$.
   Положим $\varepsilon_n = \tau \varepsilon_n$ и перейдём к шагу 2.

4. Окончание поиска: $x^* = \frac{a+b}{2}, \ f^* = f(x^*)$

## 3.3 Метод Ньютона с постоянным шагом

1. Задаём начальное приближение $x_0$ и точность вычислений $\varepsilon > 0$

2. Вычисляем градиент рассматриваемой функции и квадратной матрицы Гессе:

   градиент рассматриваемой функции:

$$\nabla f(x_k) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x_1, \ldots, x_n) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x_1, \ldots, x_n) \end{bmatrix}$$

   квадратная матрица Гессе:

$$H(x_k) = \nabla^2 f(x_k) = \begin{bmatrix} \frac{\partial^2}{\partial^2 x_1} f(x_1, \ldots, x_n) & \ldots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x_1, \ldots, x_n) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_1 \partial x_n} f(x_1, \ldots, x_n) & \ldots & \frac{\partial^2}{\partial^2 x_n} f(x_1, \ldots, x_n) \end{bmatrix}$$

3. Определяем новые значения аргументов функции после выполнения $k-$ого шага расчёта методом по следующей формуле: $x_{k+1} = x_k - H^{-1}(x_k)\nabla f(x_k)$

4. Условие окончаний вычислений: $||\nabla f(x_k)|| < \varepsilon$. Если заданная точность не достигнута, то возвращаемся к шагу 3.

## 3.4 Метод Бройдена — Флетчера — Гольдфарба — Шанно с постоянным шагом

1. Задаём начальное приближение $\varepsilon$. Выбираем начальную точку $x_0$. Полагаем $A_1 = E$ и строим множество $I_0 = \{n, 2n, \ldots\}$ моментов обновления алгоритма. Полагаем $k = 1$ и вычисляем $\omega_1 = -\nabla f(x_0)$.

2. На $k$–ом шаге проверяем выполнение неравенства: $||\omega_k|| < \varepsilon$. Если оно выполняется, то итерации прекращаются, $x^* = x_{k-1},\ f(x^*) = f(x_{k-1})$, иначе переходим к третьему пунтку.

3. Вычисляем направление спуска $p_k = A_k\omega_k$, шаг $\alpha_k : min\psi_k(\alpha) = f(x_{k-1} + \alpha p_k)$ и точку $x_k = x_{k-1} + \alpha_k p_k$. Вычисляем $\omega_{k+1} = -\nabla f(x_k)$. Если $k \in I_0$, то $A_{k+1} = E,\ k = k+1$ и возвращаемся к пункту 2. Иначе – переходим к пункту 4.

4. Полагаем $\Delta x_k = x_k - x_{k-1}$, $\Delta\omega_k = \omega_{k+1} - \omega_k$, вычисляем матрицу $A_{k+1}$, $k = k+1$ и возвращаемся к пункту 2. Матрицу $A_{k+1}$ вычисляем по формуле:

$$A_{k+1} = A_k - \frac{\Delta x_k (\Delta x_k)^T}{\Delta\omega_k^T \Delta x_k} - \frac{A_k \Delta\omega_k (\Delta\omega_k)^T A_k^T}{\Delta\omega_k^T A_k \Delta\omega_k} + \Delta\omega_k^T A_k \Delta\omega_k r^k (r^k)^T$$

где

$$r^k = \frac{A_k \Delta\omega_k}{\Delta\omega_k^T A_k \Delta\omega_k} - \frac{\Delta x_k}{\Delta x_k^T \Delta\omega_k}, \quad k \in \mathbb{N}$$

# 4 Результаты решения задачи

## 4.1 Численное решение задачи

Решение задачи методом наискорейшего спуска с точностью $\varepsilon = 10^{-3}$ и начальной точкой $[-2,\ 2.7]$:

$$x_k \ = \ [0.2722,\ -0.9582]$$

$$f(x_k) \ = \ -2.1221$$

Решение задачи методом Ньютона с точностью $\varepsilon = 10^{-3}$ и начальной точкой $[0.4,\ 0.4]$:

$$x_k \ = \ [0.2729,\ -0.9588]$$

$$f(x_k) \ = \ -2.1220$$

Решение задачи методом БФГШ с точностью $\varepsilon = 10^{-3}$ и начальной точкой $[2,\ -2.7]$:

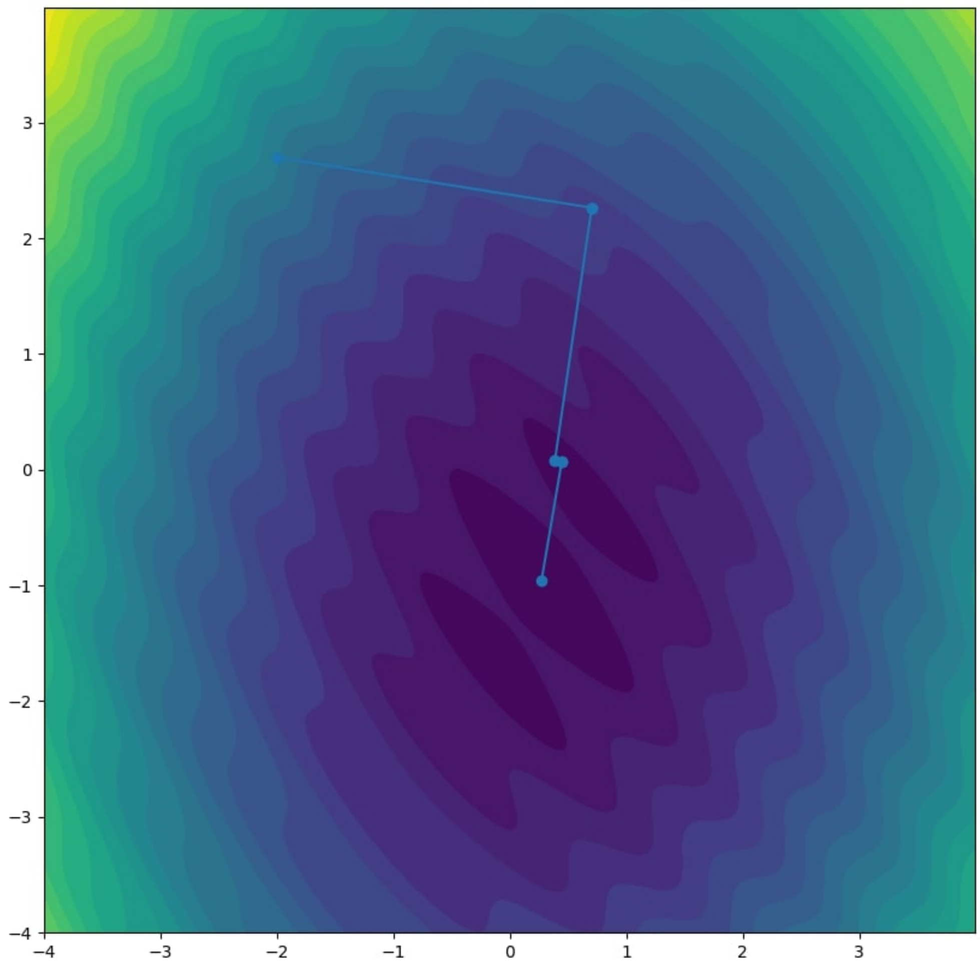
$$x_k \ = \ [0.2727,\ -0.9589]$$

$$f(x_k) \ = \ -2.1220$$

## 4.2 Линии уровня, рельефы

Все эффективные методы поиска минимума сводятся к построению траекторий, ыдоль которых функция убывает; разные методы отличаются способами построения таких траекторий. Метод, приспособленный к одному типу рельефа, может оказаться плохим на рельефе другого типа.

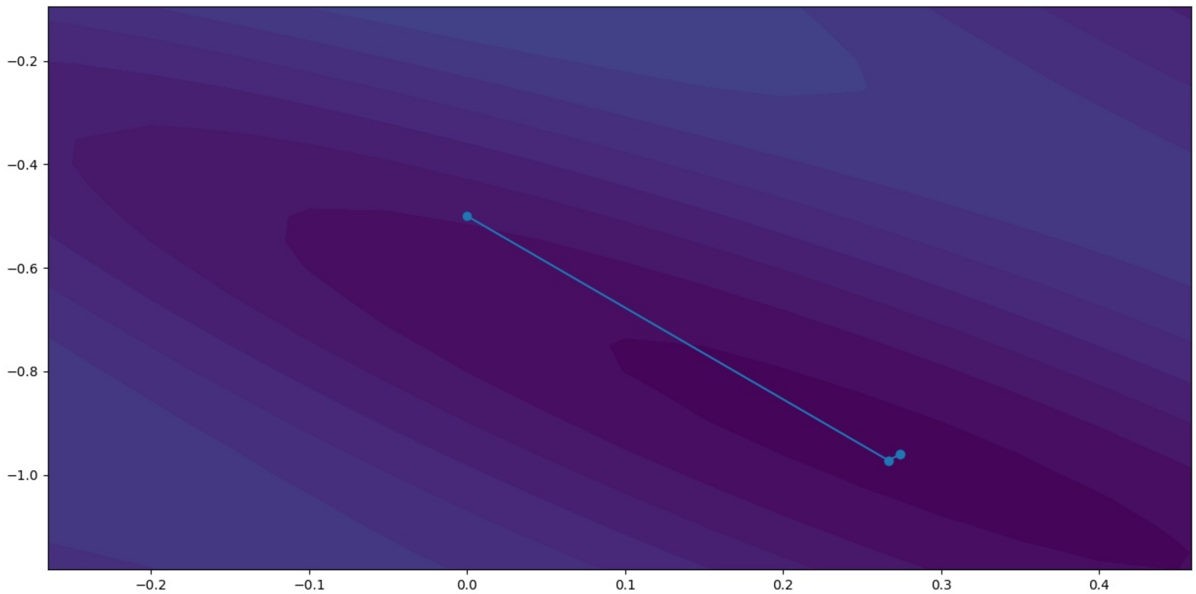
Линии уровня функции цели будут задаваться условием: $z = c$, то есть уравнениями вида $F(x,\ y) = c$. Совокупность этих линий уровня для разных значений $c$ показывает рельеф функции. Построим линии уровня и покажем звенья ломаных поверх рельефа.



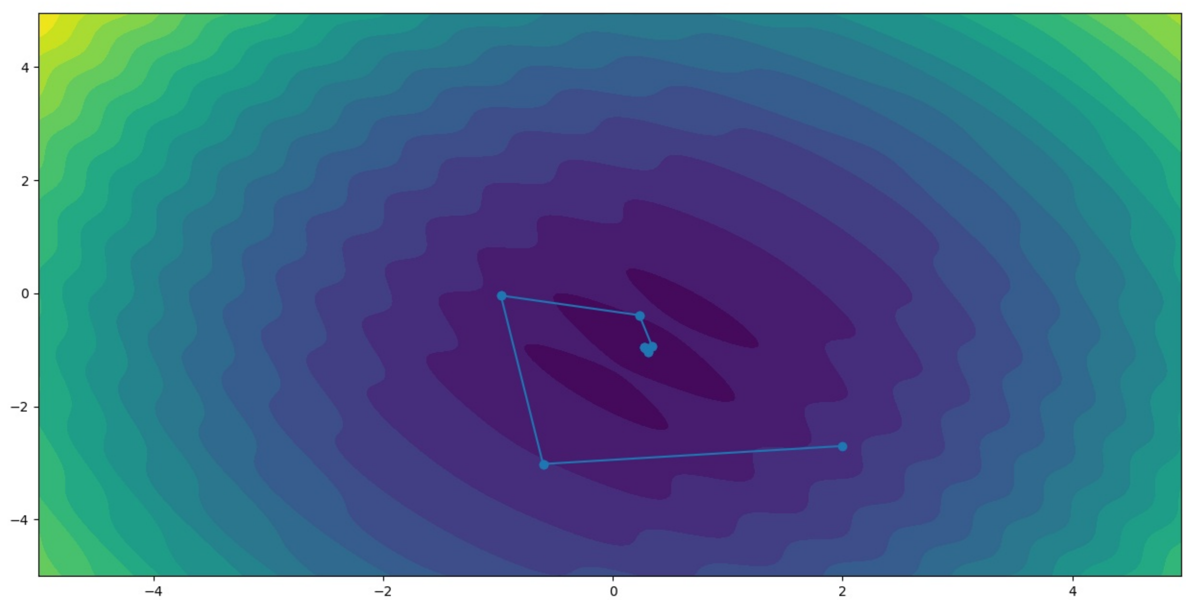

Видим, что выполняется условие ортогональности звеньев, которое аналитически записывается в виде:

$$-\nabla^T f(x_{k+1}) \nabla f(x_k) = 0$$

Далее построим ломанную для метода Ньютона:




И для метода Бройдена — Флетчера — Гольдфарба — Шанно:



# 5 Сравнительный анализ

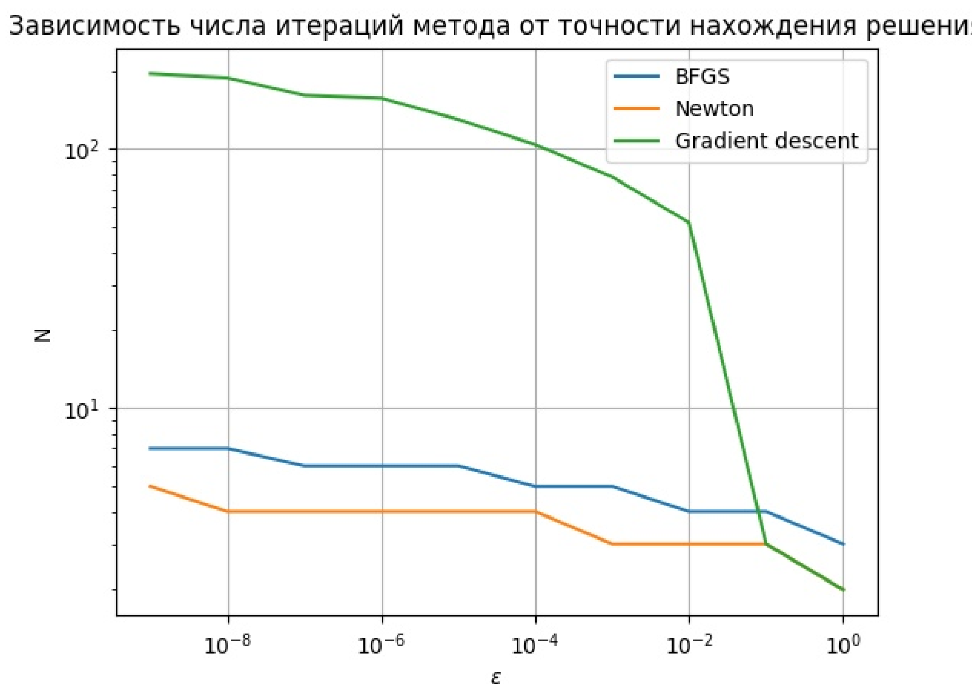
Метод наискорейшего спуска гарантирует сходимость лишь в смысле $\lim_{k\to\infty} ||\nabla f(x_k)|| = 0$, то есть сходимость по функции либо к точной нижней грани $inf f(x)$, либо к значению функции $f$ в некоторой стационарной точке $x^*$. При этом сама точка $x^*$ не обязательно является точкой локального минимума; она может быть точкой седлового типа. Однако на практике подобная ситуация маловероятна и применение градиентных методов, как правило, позволяет получить приближенное значение минимума целевой функции (вообще говоря, локально).

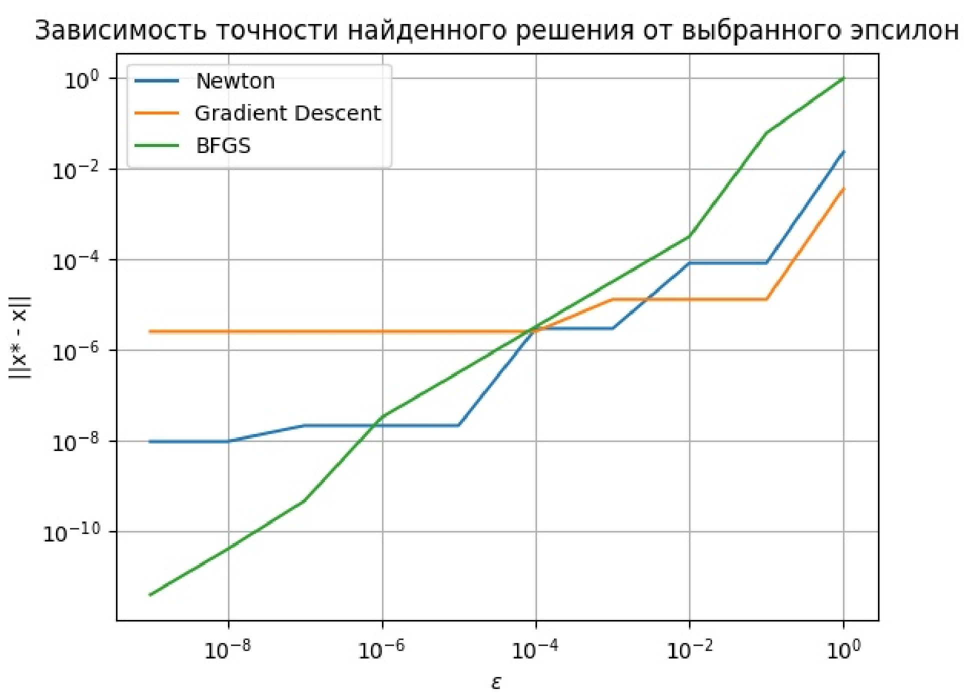
Метод Ньютона по сравнению с градиентным состоит в том, что он «не реагирует» на овражный характер минимизируемой функции. Градиентные методы, по существу, используют линейную аппроксимацию целевой функции и поэтому менее точно определяют направление на точку минимума. Таким образом, метод Ньютона позволяет достигнуть заданной точности за меньшее число итераций, чем градиентные методы. Однако каждая итерация метода Ньютона связана с вычислением матрицы Гессе и последующим обращением, что требует большего объёма вычислений по сравнению с одной итерацией градиентного метода. Если начальная точка выбрана не достаточно близко к оптимальной, то с большой вероятностью метод разойдётся.

Стремление уменьшить объем вычислений привело к созданию класса методов, близких по скорости сходимости к методу Ньютона, но не использующих вторые производные целевой функции $f(x)$ и процедуры обращения матрицы $f''(x)$. Недостатком этих методов является необходимость хранения в памяти ЭВМ матриц $A_k$, что при решении задач высокой размерности может создать определённые трудности.

# 6 Проведение сравнительных экспериментов

Ранее мы уже убедились, что наша фунция имеет несколько минимумов. Такое поведение характерно для квазивыпуклых функций. В связи с этим возникает желание провести эксперименты и понаблюдать, с какой скоростью будут сходиться методы, а также сравнить число итераций.




Видим, что для нашей функции оптимальным является использование метода БФГШ.

# 7 Оценка достоверности полученного результата

Для того, чтобы убедиться, что найденное оптимальное значение $x^*$ верно, необходимо и достаточно:

1. $\begin{cases} \dfrac{\partial f}{\partial x_1} = 0 \\ \dfrac{\partial f}{\partial x_2} = 0 \end{cases}$

2. $\begin{cases} \dfrac{\partial^2 f}{\partial x_1^2} > 0 \\ \dfrac{\partial^2 f}{\partial x_1^2} \cdot \dfrac{\partial^2 f}{\partial x_2^2} - \left(\dfrac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 > 0 \end{cases}$

$x^* = [\,0,2738936\,;\ -0,9601768\,]$ , $\varepsilon = 10^{-6}$.

$$\begin{cases} \dfrac{\partial f}{\partial x_1} = 4x_1 - 6\sin(6x_1 + 5x_2) - 1 \\ \dfrac{\partial f}{\partial x_2} = 2x_2 - 5\sin(6x_1 + 5x_2) + 2 \end{cases} \quad \text{подставим } x^*: \quad \Rightarrow$$

$$\Rightarrow \left.\frac{\partial f}{\partial x_1}\right|_{x^*} = 4 \cdot 0,2738936 - 6\sin(6 \cdot 0,2738936 + 5 \cdot (-0,9601768)) - 1 \approx -3.623 \cdot 10^{-8}$$

с необходимой точностью $10^{-6}$ $\dfrac{\partial f}{\partial x_1} = 0$.

$$\left.\frac{\partial f}{\partial x_2}\right|_{x^*} = 2 \cdot (-0,9601768) - 5\sin(6 \cdot 0,2738936 + 5 \cdot (-0,9601768)) + 2 \approx 1.036 \cdot 10^{-7}$$

с точностью до порядка $\varepsilon$ имеем $\dfrac{\partial f}{\partial x_2} = 0$.

$$\begin{cases} \dfrac{\partial^2 f}{\partial x_1^2} = 4 - 36\cos(6x_1 + 5x_2) > 0 \\ \dfrac{\partial^2 f}{\partial x_1^2} \cdot \dfrac{\partial^2 f}{\partial x_2^2} - \left(\dfrac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 = (4 - 36\cos(6x_1 + 5x_2)) \cdot (2 - 25\cos(6x_1 + 5x_2)) - 900\cos^2(6x_1 + 5x_2) > 0 \end{cases}$$

$$\left.\frac{\partial^2 f}{\partial x_1^2}\right|_{x^*} = 4 - 36\cos(6 \cdot 0,2738936 + 5 \cdot (-0,9601768)) = 39.9954324 > 0$$

$$\left.\frac{\partial^2 f}{\partial x_1^2} \cdot \frac{\partial^2 f}{\partial x_2^2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2\right|_{x^*} = (4 - 36\cos(0.2738936 \cdot 6 + 5 \cdot (-0,9601768))) \cdot$$

$$\cdot (2 - 25\cos(6 \cdot 0.2738936 + 5 \cdot (-0.9601768))) - 900\cos^2(6 \cdot 0.2738936 + 5 \cdot (-0,9601768)) =$$

$$= 179,978177 > 0$$

Таким образом, $x^*$ - точка минимума функции $f(x)$.

# 8 Дополнительные исследования

## 8.1 Уточняющие детали градиентного спуска

При поиске шага ставится задача: выйдя из точки $x_k$, найти в направлении антиградиента $-\nabla f(x_k)$ точку, дающую минимум функции $f(x_k - \alpha \nabla f(x_k))$. Пусть мы находимся в точке $x_k$ и выбираем шаг $\alpha_k$. Если $m$ – наименьшее собсвтенное число Гессиана, то ближайший локальный минимум не может быть ближе к $x_k$, чем $\frac{||\nabla f(x_k)||}{m}$. Будем искать шаг на промежутке $\alpha = [\frac{1}{m},\ 1]$.

Длина желаемого интервала неопределённости, до которого нужно сузить начальный, должна быть такой, чтобы в итоге достичь условия

$$||\nabla f(x_k)|| < \varepsilon, \quad k \to \infty$$

для этого достаточно локализовать точку минимума $x_k^*$ одномерной функции с точностью $\frac{\varepsilon}{R}$, где $R$ – константа Липшица. Такой выбор позволяет, двигаясь в направлении точки с $\nabla f = 0$, гарантированно попасть в $x_{k+1}$ такую, что условие малости нормы градиента будет выполняться.

## 8.2 Вариативность условия остановки методов

В наших алгоритмах мы использовали

$$||\nabla f|| < \varepsilon$$

как условие остановки итерационных процессов. Однако может быть использовано равносильное ему

$$||\nabla f||^2 < \bar{\varepsilon},\ \text{где}\ \bar{\varepsilon} = \varepsilon^2$$

Но как понять, когда лучше использовать то или иное условие?

Из курса функционального анализа известно, что в конечномерных пространствах любые две нормы эквивалентны. Так, если выбрана первая $||\overrightarrow{x}||_1 = \sum\limits_{k=1}^{n} x_k$ или бесконечная норма $||\overrightarrow{x}||_\infty = \max x_k$, то для упрощения вычислений легче воспользоваться критерием $||\nabla f|| < \varepsilon$. Если мы работаем со второй нормой $||\overrightarrow{x}||_2 = \sqrt{\sum\limits_{k=1}^{n} x_k^2}$ то с точки зрения вычислений проще воспользоваться критерием $||\nabla f||^2 < \bar{\varepsilon}$.

Евклидова норма является геометрическим расстоянием между двумя точками в многомерном пространстве, поэтому целесообразнее всего в работе было использовать её. В соответсвие с этим, во всех вышеописанных алгоритмах был подправлен критерий останова.

# 9 Программная реализация

В процессе реализации алгоритмов использовался язык программирования Python3.6. Для проверки полученных решений пользовались пакетом MATLAB2020b.

Исходный код находится в системе контроля версий GitHub
https://github.com/Brightest-Sunshine/Optimization-methods-2021

# Список литературы

[1] Лесин В. В., Лисовец Ю. П. Основы методов оптимизации: Учебное пособие. 3-е изд., испр, – СПб.: Издательство «Лань», 2011. – 352с.

[2] Сухарев А. Г, Тимохов А. В., Фёдоров В. В Курс методов оптимизации: Учебное пособие, 2-е издание, – М.:ФИЗМАТЛИБ, 2005, – 368с.

[3] Ногин В. Д., Протодьяконов И. О., Еврампиев И. И. Основы теории оптимизации : Учебное пособие для вузов