«НАЦИОНАЛЬНЫЙ ИССЛЕДВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»

Факультет информационных технологий и программирования

Кафедра компьютерных технологий

ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №2

Методы градиентной оптимизации

Выполнили студенты:

Ефимов Сергей Алексеевич
группа: M3237

Соколов Александр Андреевич
группа: M3234

Проверил:

Свинцов Михаил Викторович

г. Санкт-Петербург

Постановка задачи

Задача лабоpатpной работы – научиться реализовывать алгоритмы градиентной оптимизации каждым из следуших способов:

1. Метод градиентного спуска
2. Метод наискорейшего спуска
3. Метод сопряженных градиентов

Также необходимо оценить как меняется скорость сходимости, если для поиска величины шага использовать различные методы одномерного поиска.

Ход работы

Фомализем поставленную задачу: Пусть мы имеем отображение из $f : \chi \subset \mathbb{R}^n \longrightarrow \mathbb{R}$, тогда f - квадратичная функция. Необходимо реализовать алгоритмы минимизации данной функции, которые на выходе отдают $x^* = (x', y')$ - точка минимума, $f(x^*) = m$ - минимум функции Также мы будем испльзовать такие понятия как градиент $\nabla f(x)$ - вектор-столбец частных производных 1-ого порядка в этой точке, Матрица A - симметричная матрица коэфициентов при $x_i \cdot x_j$, b — вектор-столбец при x_i , $c \in \mathbb{R}$.

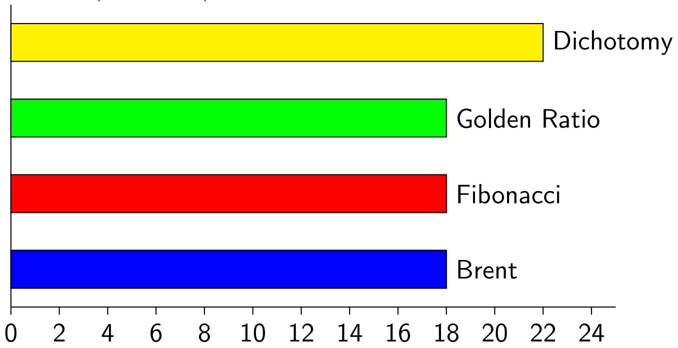
Оценка скорости сходимости

Давайте рассмотрим зависимость скорости сходмости алгоритма от метода поиска шага однмерной сходимости

Без ограничений общности возьмем функцию от двух переменных, у которой матрица A диагональная:

$$f(x, y) = 5x^2 + 7y^2 - 8x + 9y + 10$$

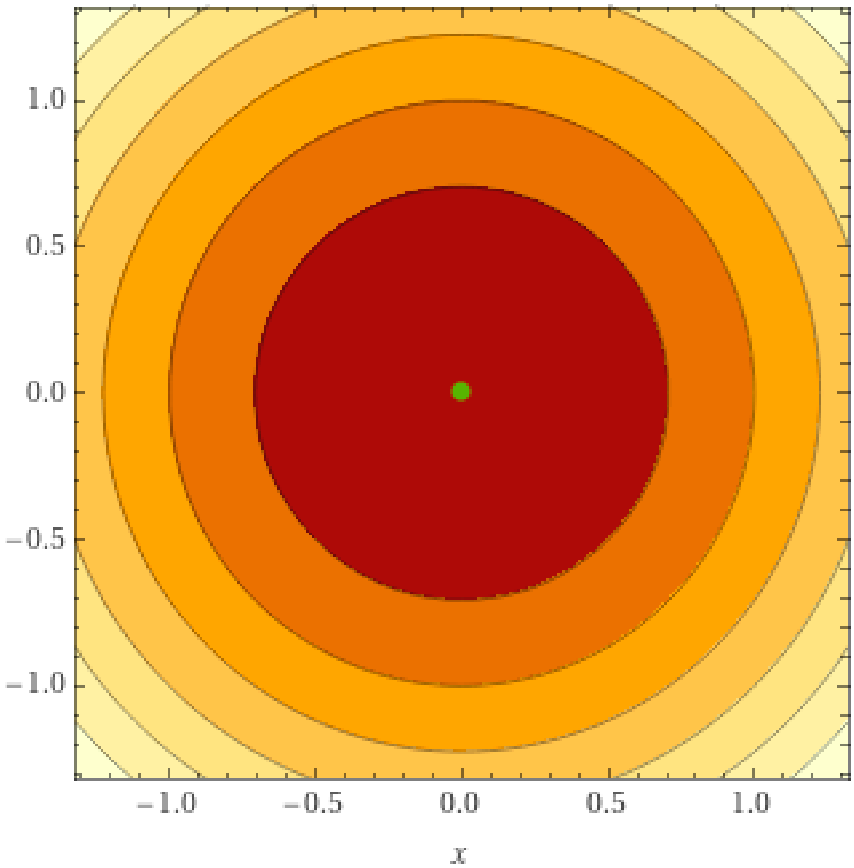
Минимум данной функции $\min f(x, y) = 3.908$ Минимум достигается в точке $X(0.8, -0.643)$ Точность для для каждого из методов была выбрана одинаковая($\epsilon = 10^{-5}$)



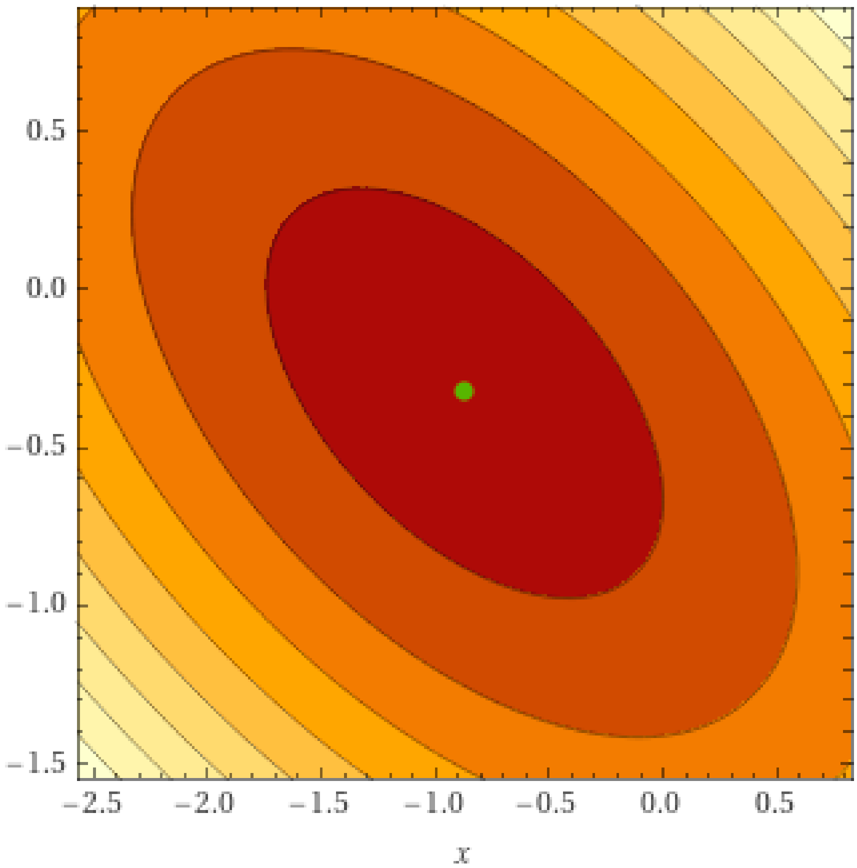
Анализ функций

Возьмем три функции для анализа и приведем для каждого линии уровня

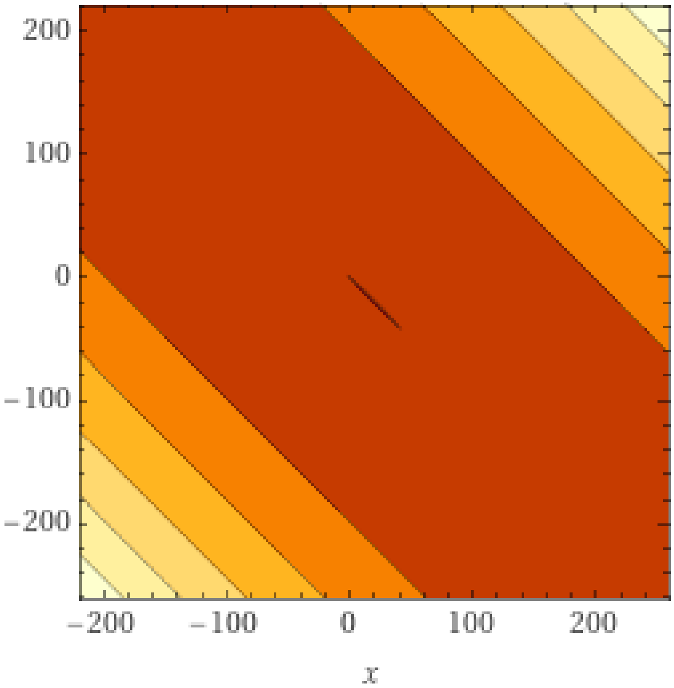
- $f(x, y) = x^2 + y^2$



- $f(x, y) = 5x^2 + 7xy + 9y^2 + 11x + 12y + 14$



- $f(x, y) = 254x^2 + 506xy + 254y^2 + 50x + 130y - 111$



Приведем аналитическое реешние для данных функций и найдем минимум:

$$1) f(x,y) = x^2 + y^2$$

$$\min x^* = (0,0), f(x^*) = 0$$

$$2) f(x,y) = 5x^2 + 7xy + 9y^2 + 11x + 12y + 14$$

$$f'_x = 10x + 7y + 11$$

$$f'_y = 7x + 18y + 12$$

$$\begin{cases} f'_x = 0 \\ f'_y = 0 \end{cases}$$

$$\begin{cases} 10x + 7y + 11 = 0 \\ 7x + 18y + 12 = 0 \end{cases}$$

$$\begin{cases} x^* = -\frac{114}{131} \\ y^* = -\frac{43}{131} \end{cases}$$

$$f''_{xx} \cdot f''_{yy} - f''_{xy} \cdot f''_{yx} = 10 \cdot 18 - 7 \cdot 7 > 0$$

Откуда $\min x^* = (-\frac{114}{131}, -\frac{43}{131})$

$$3) f(x,y) = 254x^2 + 506xy + 254y^2 + 50x + 130y - 111$$

$$f'_x = 508x + 506y + 50$$

$$f'_y = 506x + 508y + 130$$

$$\begin{cases} f'_x = 0 \\ f'_y = 0 \end{cases}$$

$$\begin{cases} 508x + 506y + 50 = 0 \\ 506x + 508y + 130 = 0 \end{cases}$$

$$\begin{cases} x^* = \frac{3365}{169} \\ y^* = -\frac{3393}{169} \end{cases}$$

$$f''_{xx} \cdot f''_{yy} - f''_{xy} \cdot f''_{yx} = 508 \cdot 508 - 506 \cdot 506 > 0$$

Откуда $\min x^* = (\frac{3365}{169}, -\frac{3393}{169})$

Общий принцип выполнения методов

Пусть есть f- минимизуемая функция. Задается некое первое приближения решения $x^{(0)}$. На каждом шаге метода мы вычисляем следущее приближение

$$x^k = \Phi(x^{k-1}, x^{k-2}, ..., x^0)$$

В нашем случае можно рассмотреть некое упрощение поиска следущего приближения как

$$x^k = x^{k-1} + \alpha_{k-1} + p^{k-1}$$

Где α_k - величина шага на k-той итерации, а p^k - направление поиска

Метод градиентного спуска

Задается

- ϵ - заданная точность
- α - величина шага на каждой итерации
- x^0- первое приближение

Изначально вычисляем $f(x^0)$. На каждой итерации будем вычислять $\nabla f(x^k)$, если найденное значение меньше ϵ, то метод завершается и ответом становится последнее найденное приблиение, пересчет на k-той итерации происходит следущм образом:

- $t = x^{k-1} - \alpha \nabla f(x^{k-1})$
- Пока $f(t) >= f(x^{k-1})$ повторять предыдущий пункт с $\alpha = \frac{\alpha}{2}$
- $x^k = t, f(x^k) = f(t)$

Метод наискорейшего спуска

Задается

- ϵ - заданная точность
- x^0- первое приближение

Метод похож на предыдущий с одним отличием, α не задается константная, анаходится одним из методов одномерной оптимизации(например, методом заолотого сечения)

Метод сопряженных градиентов

Задается

- ϵ - заданная точность
- x^0- первое приближение
- p^0- начальное направление поиска

p^0 находится как $-\nabla f(x^0)$. Пересчет происходит по следущим правилам:

$$x^k = x^{k-1} + \alpha_{k-1} p^{k-1}$$

$$p^k = -\nabla f(x^k) + \beta_{k-1} p^{k-1}$$

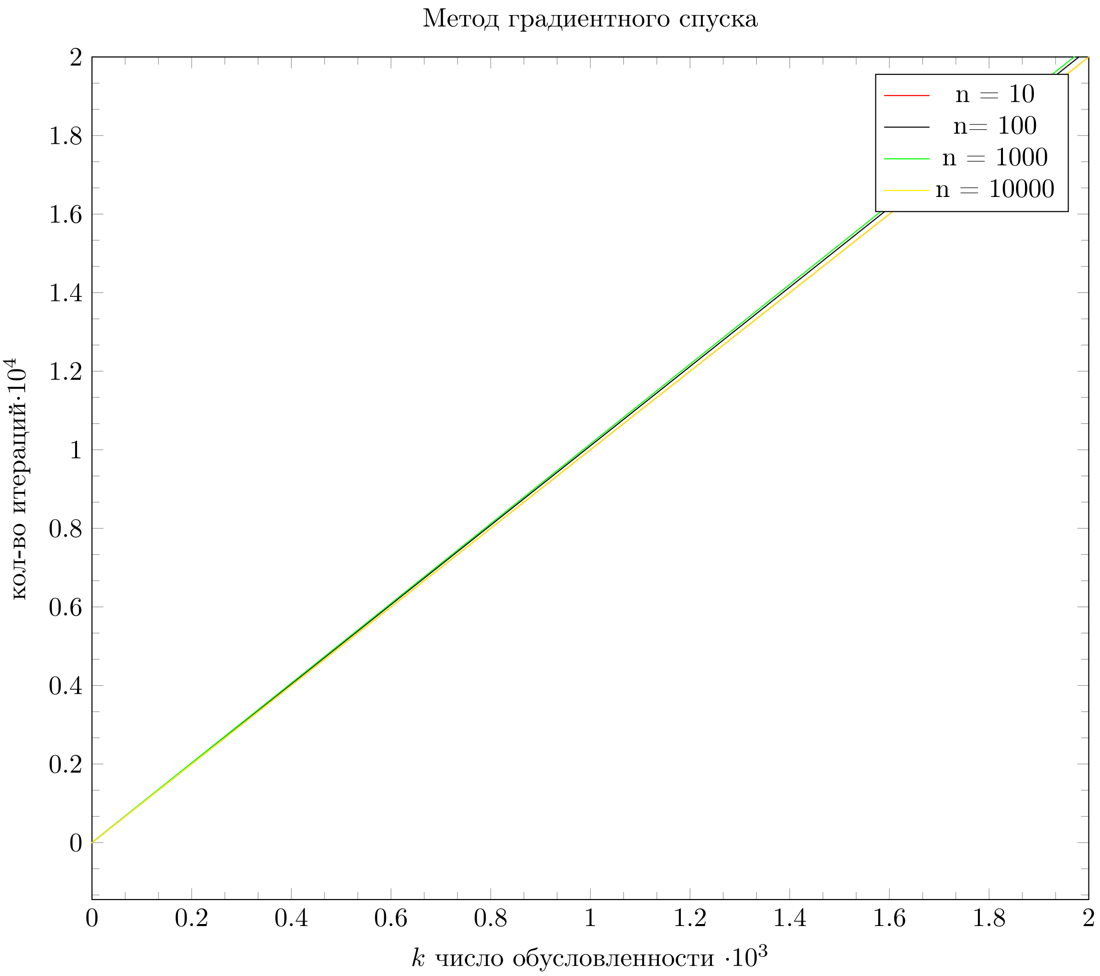
где

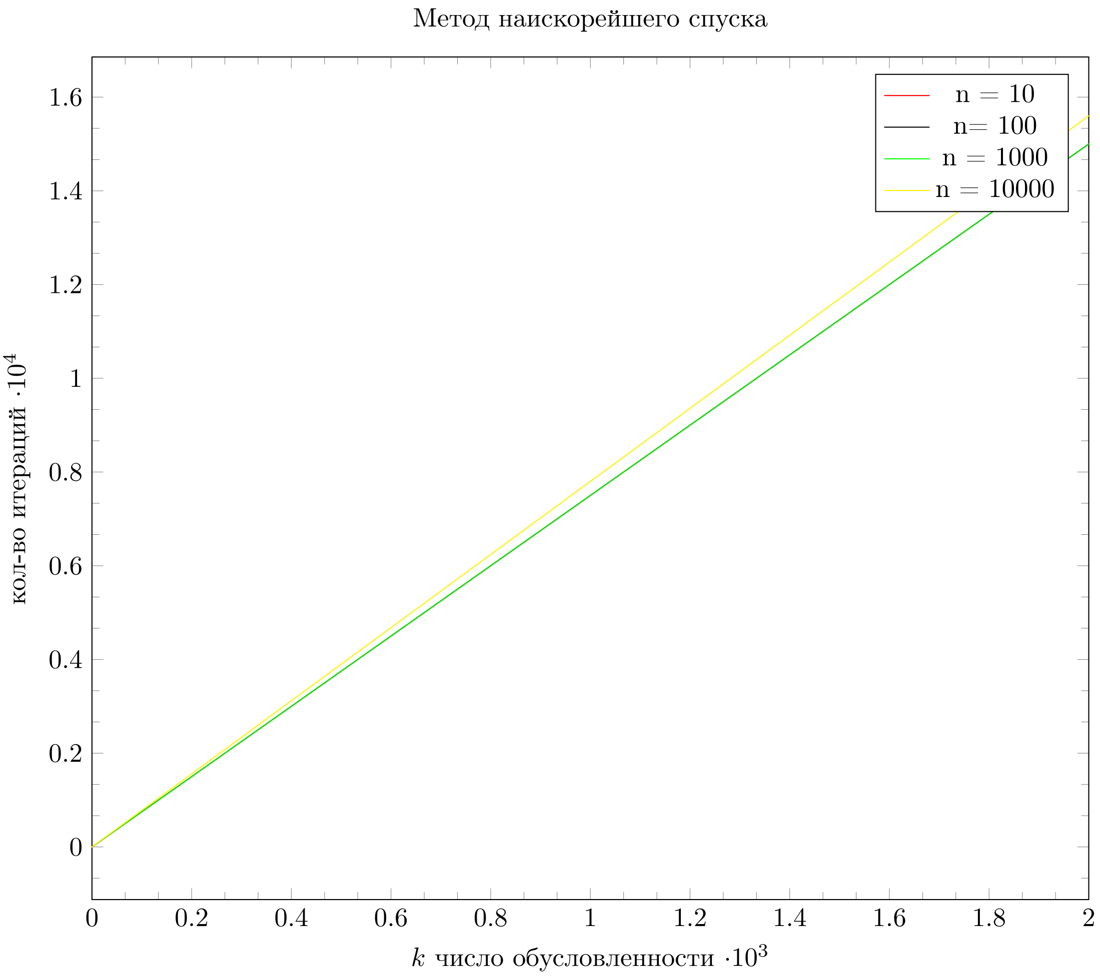
$$\alpha_k = \frac{||\nabla f(x^k)||^2}{\langle Ap^k, p^k \rangle}$$

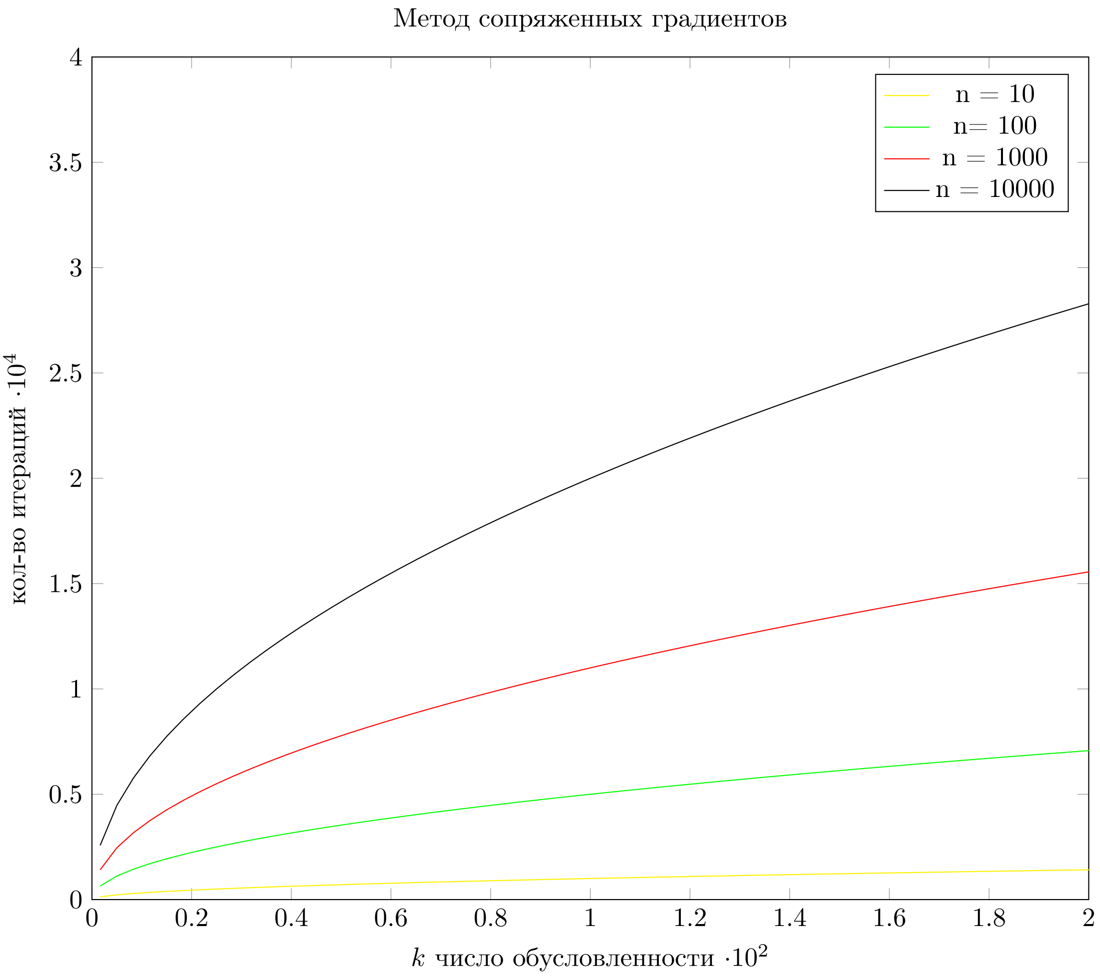
$$\beta_k = \frac{\|f(x^{k+1})\|^2}{\|f(x^k)\|^2}$$

Зависимость числа итераций

1) Автор условия лабораторной работы предлагает ввести отношение $T(n, k)$, где n – размерность пространства, а k – число обусловленности, давайте проведем несколько эксперементов


Метод градиентного спуска
n = 10
n= 100
n = 1000
n = 10000
кол-во итераций·10^4
k число обусловленности ·10^3
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2


Метод наискорейшего спуска
n = 10
n= 100
n = 1000
n = 10000
кол-во итераций ·10^4
k число обусловленности ·10^3
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2



Исходя из графиков можно сделать вывод о том, что число обусловленности k влияет на то, насколько может изменится значение фукнции при изменении аргумента

Вывод

Исходя из проделанной работы можно сделать вывод о том, что метод градиентного и наискорейшего спуска обладают линейной зависимостью числа итераций от числа обусловленности, так же стоит заметить, что они ведут себя примерно одинаково при одних и тех же входных данных. Размерность пространства не влияет на кол-во итераций

На методе сопряженных градиентов можем наюлдать другую картину, число итераций растет пропорцианально $\sqrt{k}$ от числа обусловленности, такую же тенденцию можно наблюдать в зависимсти от размерности