МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное автономное образовательное учреждение высшего образования

Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики

Мегафакультет трансляционных информационных технологий

Факультет информационных технологий и программирования

Лабораторная работа № 2
Алгоритмы многомерной оптимизации

По дисциплине «Прикладная математика»

Выполнили студенты групп

M32011

Лунев Илья Андреевич

Семенов Георгий Витальевич

M32041

Смирнов Сергей Викторович

Преподаватель:

Москаленко Мария Александровна

САНКТ-ПЕТЕРБУРГ

2021

I. Теоретическая часть

В лабораторной работе рассматривается задача многомерной оптимизации:

$$\mathrm{f}(x): \mathbb{R}^n \to \mathbb{R} \text{ Найти } \min_x\{z = f(x)\} \text{ с точностью } \varepsilon,$$

Приведем описание ряда методов, на которых основываются алгоритмы в данной работе:

- Метод градиентного спуска
- Метод сопряженных направлений
- Метод Ньютона

Все рассматриваемые методы основываются на понятии **градиента многомерной функции** — вектора, указывающего направление наибольшего возрастания функции.

$$\nabla \mathrm{f}(x) = \left(\frac{\partial \mathrm{f(x)}}{\partial \mathrm{x}_1}, \frac{\partial \mathrm{f(x)}}{\partial \mathrm{x}_2}, \ldots, \frac{\partial \mathrm{f(x)}}{\partial \mathrm{x}_n}\right)$$

Антиградиент — вектор, противоположный градиенту функции и указывающий направление наибольшего убывания функции.

Для каждого из этих методов требуется дополнительное исследование, является ли та или иная точка искомой.

Метод градиентного спуска (первого порядка)

Градиентные методы основываются на продвижении в направлении антиградиента с выбранным шагом.

Пока $||\mathrm{x}^{k+1} - \mathrm{x}^k|| >= \varepsilon$ И $||\nabla \mathrm{f}(\overrightarrow{\mathrm{x}^k})|| >= \varepsilon$ И $|\mathrm{f}(\overrightarrow{\mathrm{x}^{k+1}}) - \mathrm{f}(\overrightarrow{\mathrm{x}^k})| >= \varepsilon$:

t_k = t()

$\overrightarrow{\mathrm{x}^{k+1}} = \overrightarrow{\mathrm{x}^k} - t_k \cdot \nabla \mathrm{f}(\overrightarrow{\mathrm{x}^k})$

Если $\mathrm{f}(\overrightarrow{\mathrm{x}^{k+1}}) - \mathrm{f}(\overrightarrow{\mathrm{x}^k}) >= 0$:

t /= 2

Метод сопряженных направлений (первого порядка)

Этот метод состоит в построении последовательности точек $\{\overrightarrow{\mathrm{x}^k}\} \mid \forall\, \mathrm{f}(\overrightarrow{\mathrm{x}^{k+1}}) < \mathrm{f}(\overrightarrow{\mathrm{x}^k})$. Для их подсчета вводится следующая процедура:

Пока $||\mathrm{x}^{k+1} - \mathrm{x}^k|| >= \varepsilon$ И $||\nabla \mathrm{f}(\overrightarrow{\mathrm{x}^k})|| >= \varepsilon$ И $|\mathrm{f}(\overrightarrow{\mathrm{x}^{k+1}}) - \mathrm{f}(\overrightarrow{\mathrm{x}^k})| >= \varepsilon$:

$\overrightarrow{\mathrm{x}^{k+1}} = \overrightarrow{\mathrm{x}^k} + t_k \cdot \overrightarrow{\mathrm{d}_k}$

$$\beta_{k-1} = \frac{\left\|\nabla f\left(\overrightarrow{x^k}\right)\right\|^2}{\left\|\nabla f\left(\overrightarrow{x^{k-1}}\right)\right\|^2} \text{(метод Флетчера – Ривса)}$$

$$\beta_{k-1} = \begin{cases} \dfrac{\nabla f\left(\overrightarrow{x^k}\right) \cdot \left(\nabla f\left(\overrightarrow{x^k}\right) - \nabla f\left(\overrightarrow{x^{k-1}}\right)\right)}{\left\|\nabla f\left(\overrightarrow{x^{k-1}}\right)\right\|^2} \\ 0, \quad \text{каждые } n \text{ шагов} \end{cases} \text{(метод Полака – Рибьера)}$$

$$\overrightarrow{d_k} = -\nabla f(\overrightarrow{x^k}) + \beta_{k-1}\overrightarrow{d_{k-1}} \qquad (\overrightarrow{d_0} = \vec{0})$$

Метод Флетчера-Ривса может использоваться только для квадратичных функций (гессиан H положительно определен), а метод Полака-Рибьера распространяется на неквадратичные функции.

Метод Флетчера-Ривса сходится для квадратичных функций за n шагов.

Метод Полака-Рибьера сходится для сильно выпуклых функций.

Метод Ньютона (второго порядка)

Этот метод состоит в построении последовательности точек $\{\overrightarrow{x^k}\} \mid \forall\, f(\overrightarrow{x^{k+1}}) < f(\overrightarrow{x^k})$. Для их подсчета вводится следующая процедура:

Пока $||x^{k+1} - x^k|| >= \varepsilon$ И $||\nabla f(\overrightarrow{x^k})|| >= \varepsilon$ И $|f(\overrightarrow{x^{k+1}}) - f(\overrightarrow{x^k})| >= \varepsilon$:

Если $H^{-1}(x^k) > 0$:

$$\overrightarrow{d_k} = -H^{-1}(x^k) \cdot \nabla f\left(\overrightarrow{x^k}\right)$$

$$\overrightarrow{x^{k+1}} = \overrightarrow{x^k} + \overrightarrow{d_k}$$

Иначе:

$$\overrightarrow{d_k} = -\nabla f\left(\overrightarrow{x^k}\right)$$

$$t_k = t()$$

$$\overrightarrow{x^{k+1}} = \overrightarrow{x^k} + t_k \cdot \overrightarrow{d_k}$$

Сходится с квадратичной скоростью: $\|x^k - x^*\| \le \dfrac{4lq^{2^k}}{L}$

Методы поиска величины шага

До этого момента мы не поясняли, какую роль играет величина t_k — шаг. Он может быть выбран различным образом:

Постоянная величина шага

$$t_k = const$$

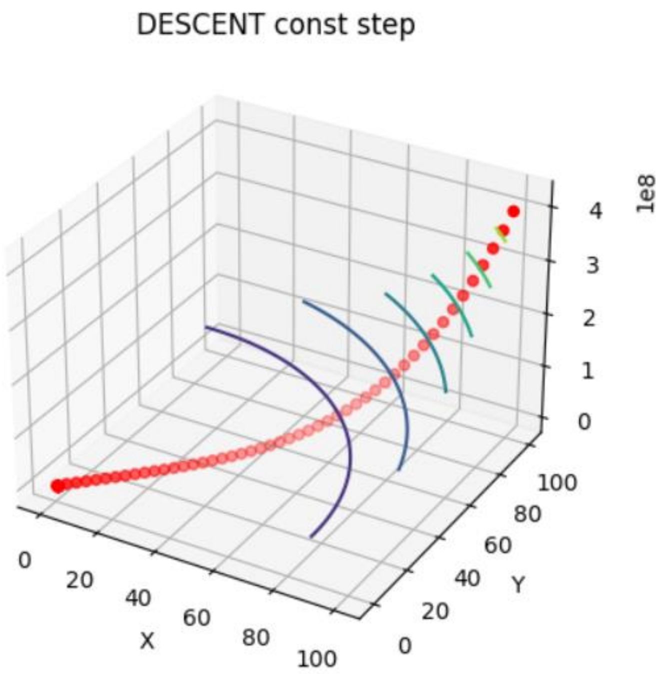
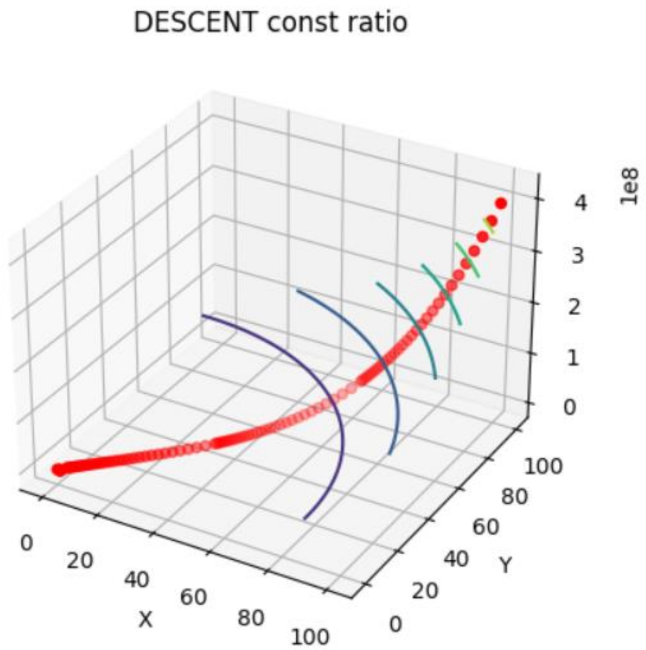
Метод дробления шага

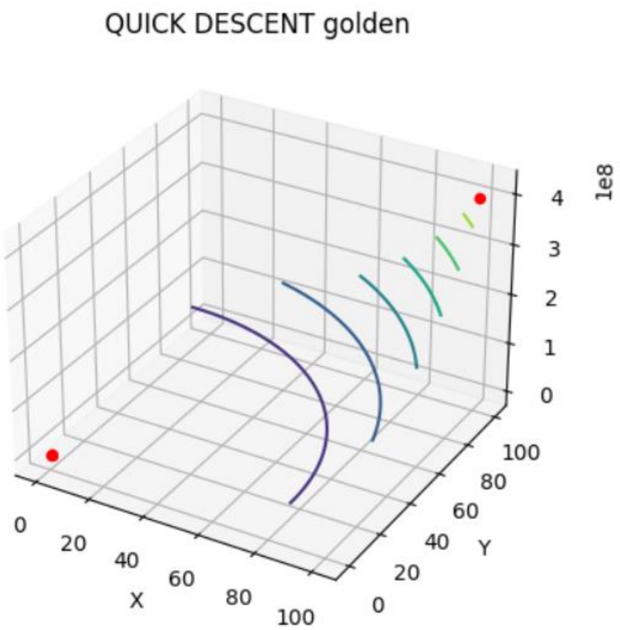
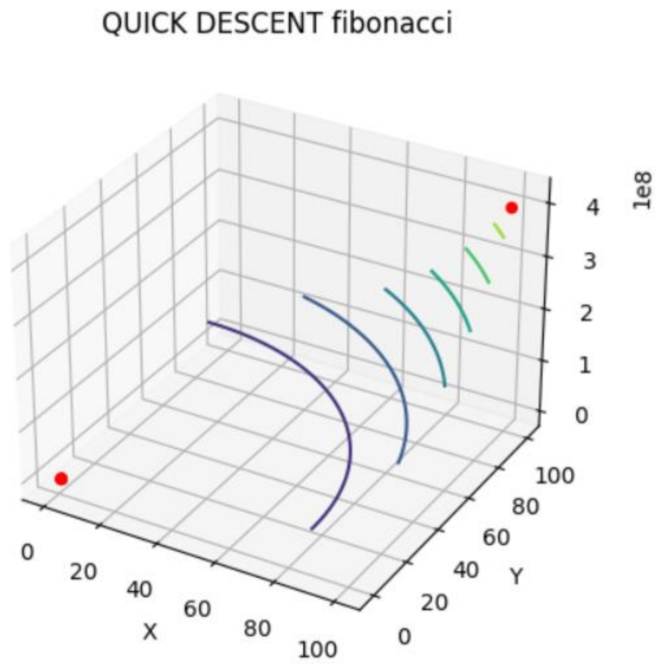
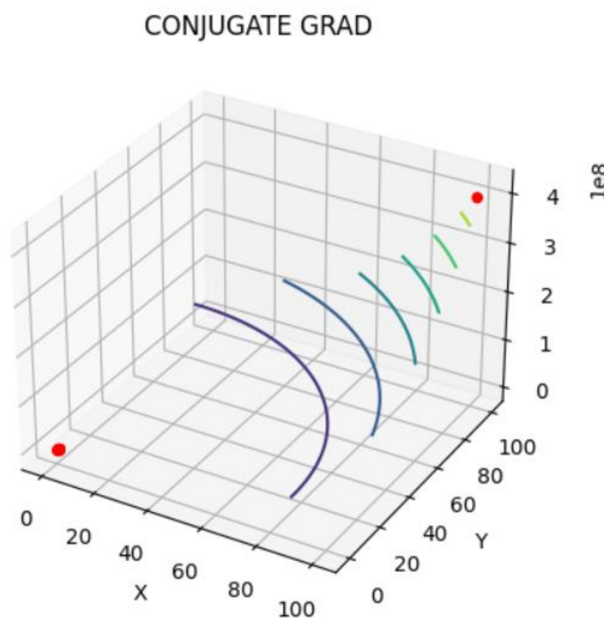
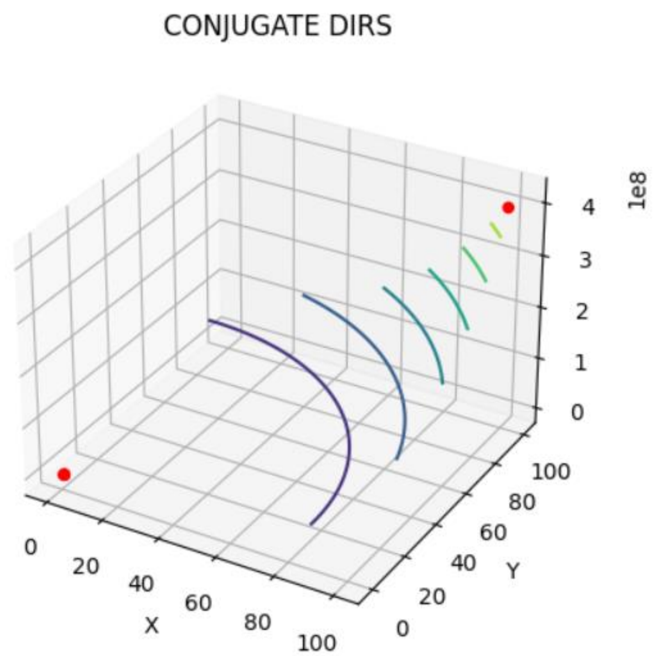
$$t_k /= const$$

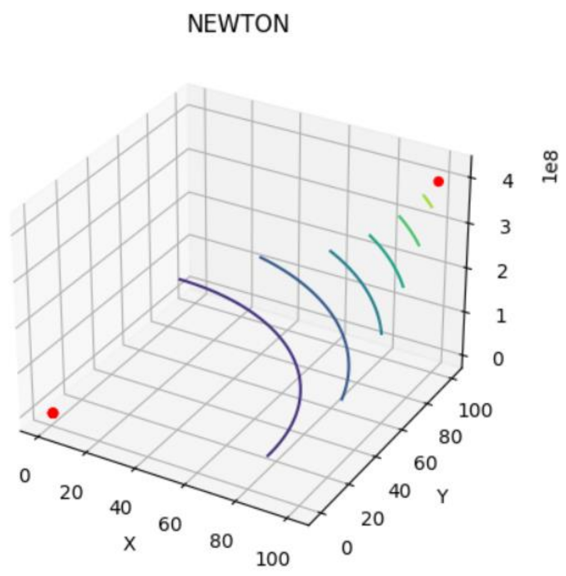
Метод золотого сечения / Фибоначчи

$$t_k = argmin_{t_k} f(\overrightarrow{x^k} — t_k \cdot \nabla f(\overrightarrow{x^k}))$$

Пример работы всех методов

$f(x, y) = x^4 + y^4 + 2x^2y^2 - 4x + 3$	
Метод градиентного спуска (постоянный шаг) [100.0, 100.0] 57 steps  	Метод градиентного спуска (дробный шаг) [100.0, 100.0] 97 steps 
Метод градиентного спуска (золотое сечение) [100.0, 100.0]	Метод градиентного спуска (Фибоначчи) [100.0, 100.0]

<table>
<tr>
<td>10 steps

</td>
<td>11 steps

</td>
</tr>
<tr>
<td>Метод сопряженных градиентов
[100.0, 100.0]
Расходится 471 steps

</td>
<td>Метод сопряженных направлений
[100.0, 100.0]
10 steps

</td>
</tr>
<tr>
<td>Метод Ньютона
[100.0, 100.0]</td>
<td></td>
</tr>
</table>

10 steps 	

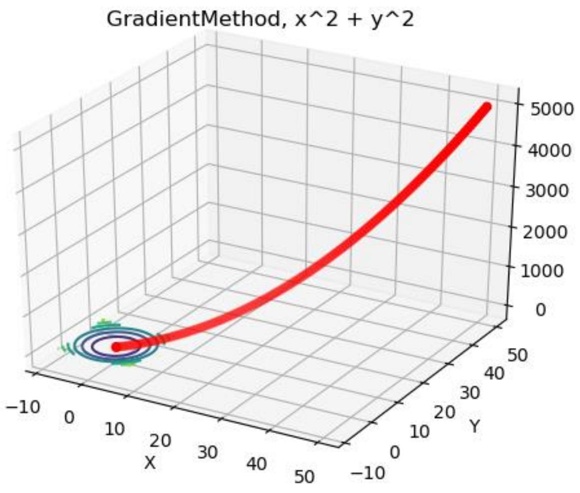
II. Экспериментальная часть

- Проанализировать траекторию реализованных методов для 2-3 квадратичных двумерных функций (чтобы работа функций отличалась)
- Рассмотреть различные начальные приближения
- Нарисовать графики с линиями уровня функций и траектории методов

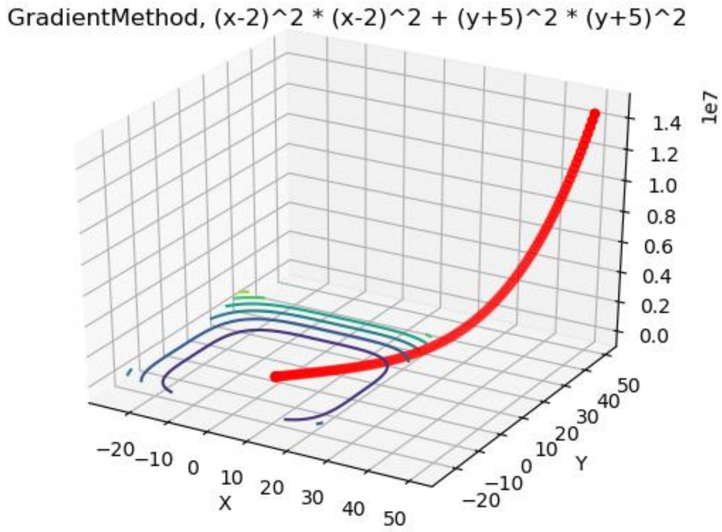
$f(x,y) = x^2 + y^2$	$f(x,y) = (x-2)^2(x-2)^2 + (y+5)^2(y+5)^2$

<table>
<tr>
<td>[50.0, 50.0]

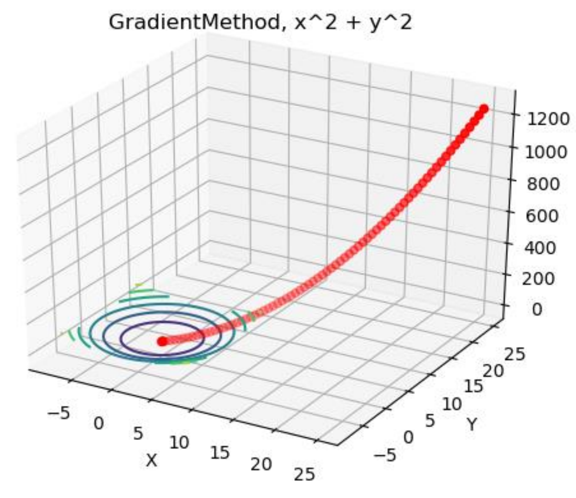
147 steps


</td>
<td>[50.0, 50.0]

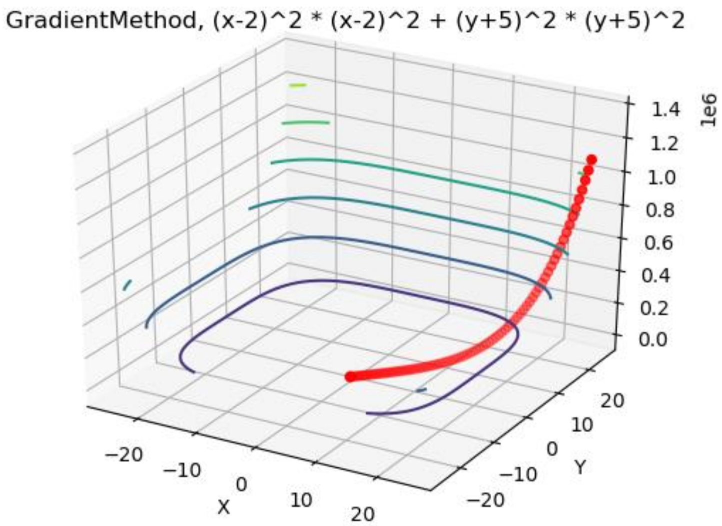
151 steps


</td>
</tr>
<tr>
<td>[25.0, 25.0]

78 steps

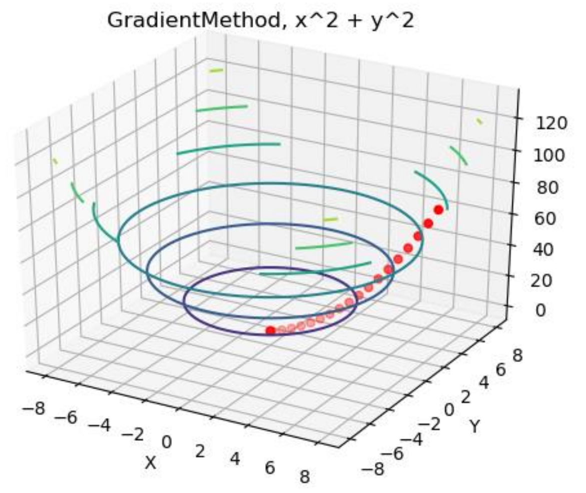
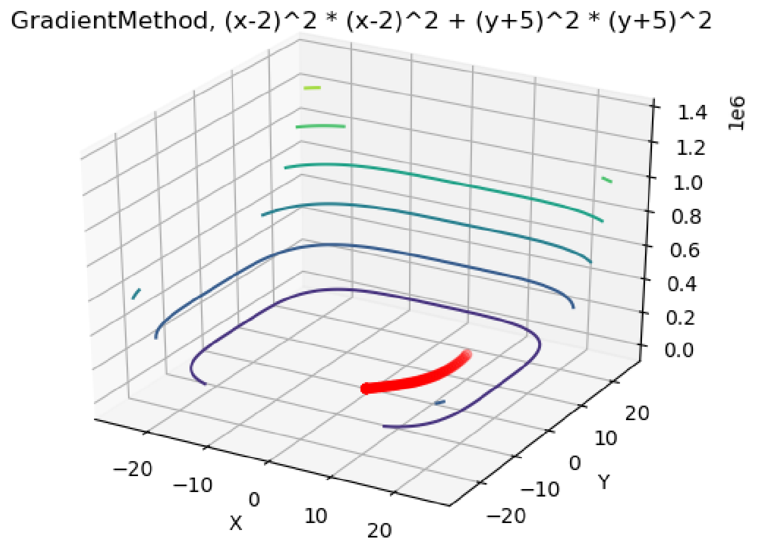
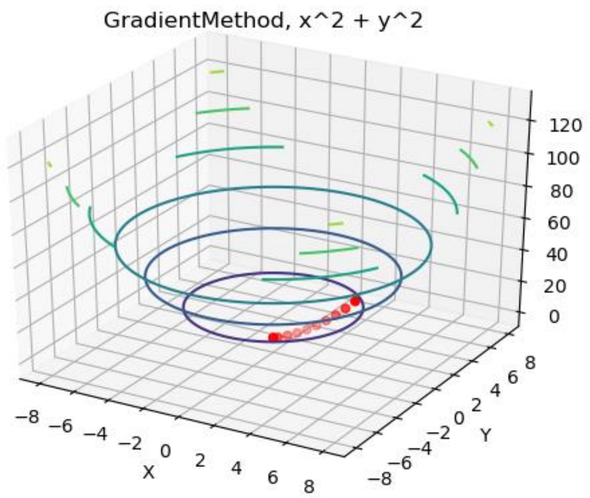
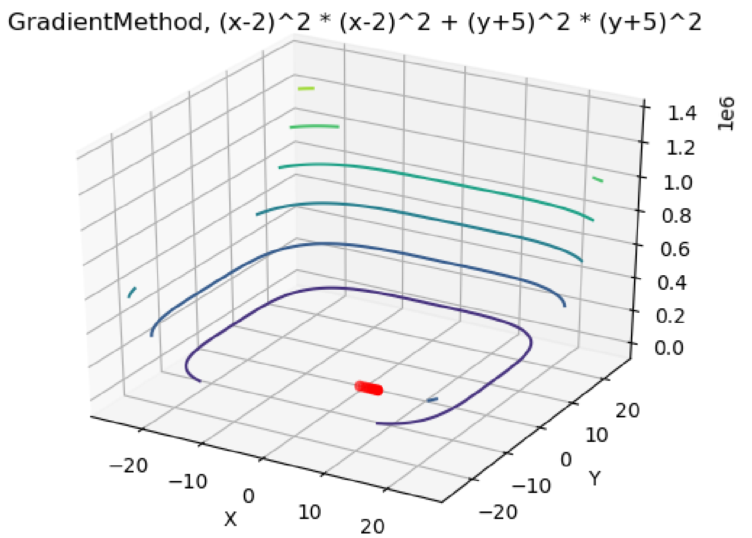

</td>
<td>[25.0, 25.0]

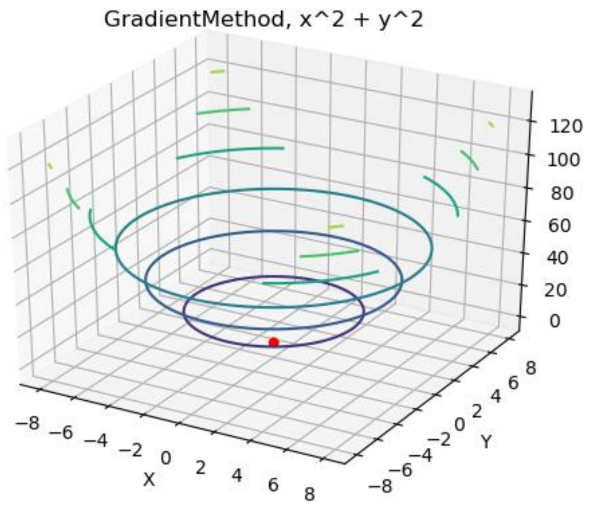
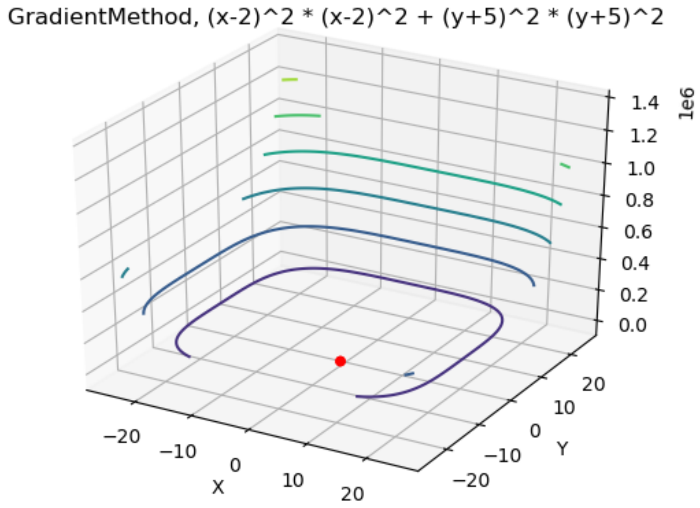
80 steps


</td>
</tr>
<tr>
<td>[7.0, 5.0]

22 steps</td>
<td>[10.0, 10.0]

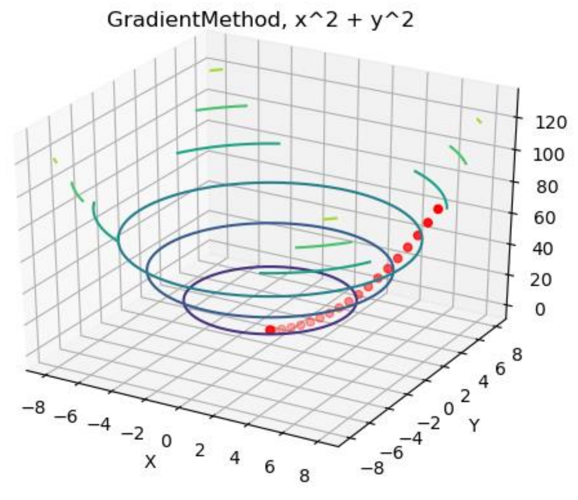
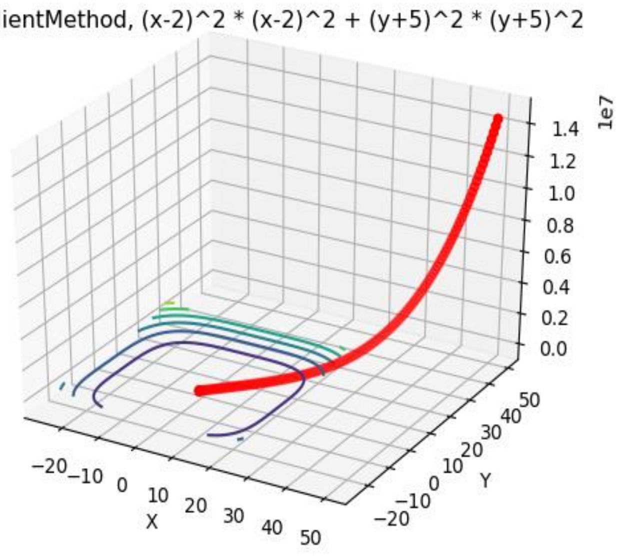
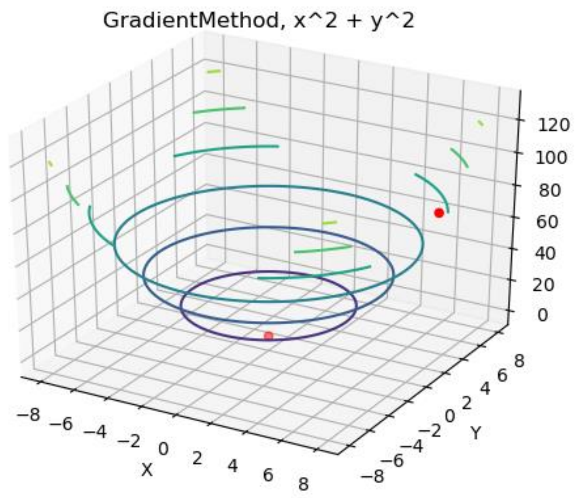
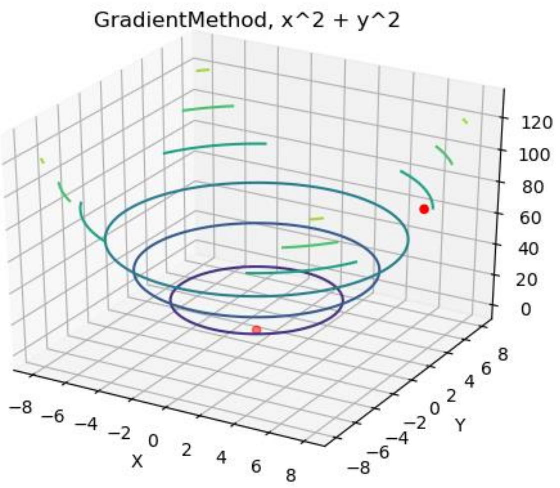
40 steps</td>
</tr>
</table>

	
[1.0, 1.0] 9 steps 	[5.0, -5.0] 7 steps 

[0.0, 0.0] 0 steps  	[2.0, -5.0] 0 steps 

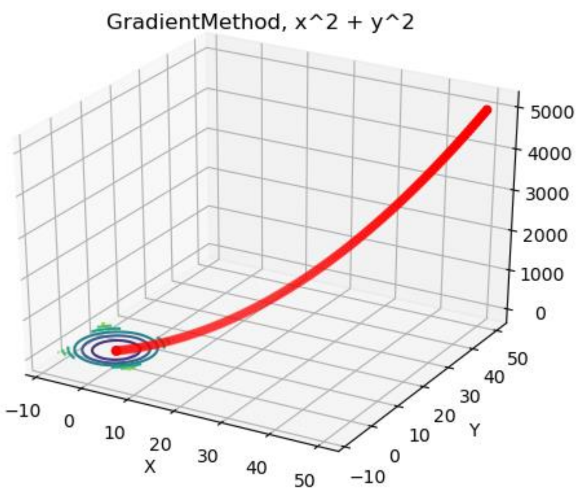
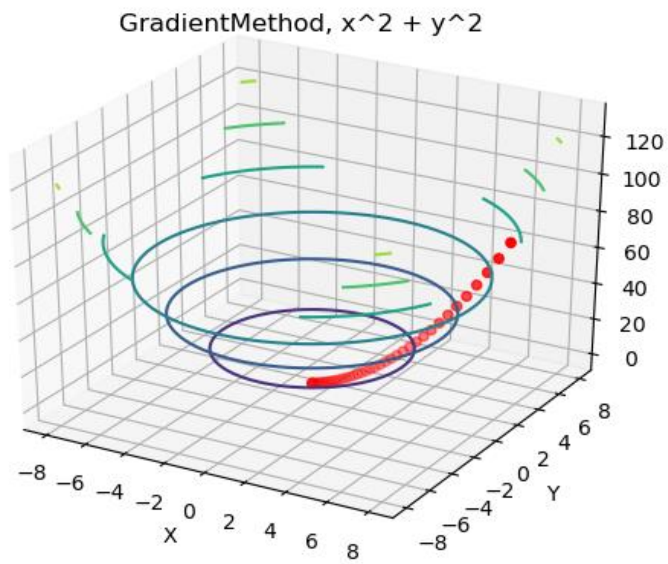
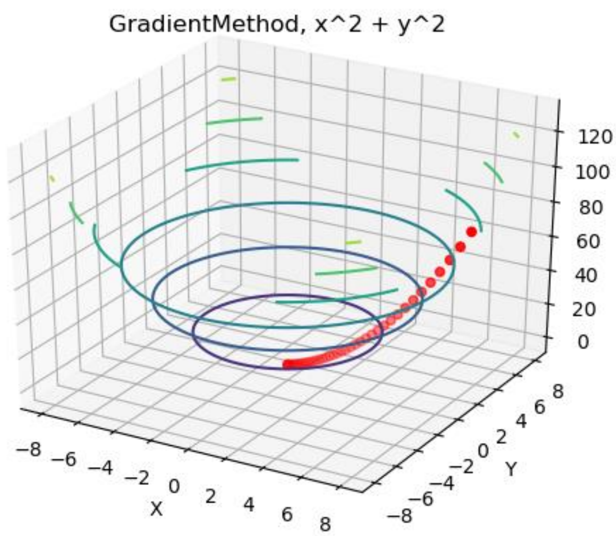
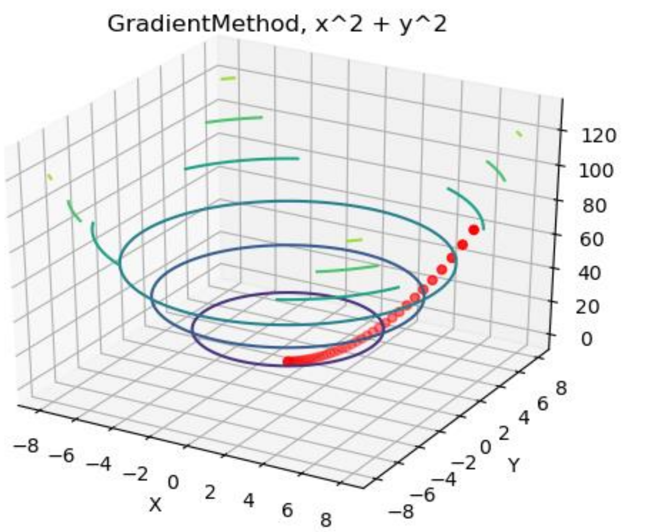
Проверена зависимость сходимости от выбора подсчёта шага

Постоянный шаг 0.5 [7.0, 5.0]	Дробление шага [7.0, 5.0]

22 steps  	42 steps 
Метод золотого сечения [7.0, 5.0] 2 steps  	Метод Фибоначчи [7.0, 5.0] 2 steps 

Вывод: метод постоянного шага работает быстрее метода дробления шага, но при неудачном выборе шага может расходиться, в отличии от других методов выбора шага. Однако метод дробления шага так же может разойтись при неудачном выборе коэффициента α, коэффициента уменьшения шага.

Постоянный шаг 0.5 [7.0, 5.0]	Постоянный шаг 0.05 [7.0, 5.0]

<table>
<tr><td>22 steps
</td><td>175 steps
</td></tr>
<tr><td>Дробление шага (step=0.5, alpha=0.95)
[7.0, 5.0]
42 steps
</td><td>Дробление шага (step=0.5, alpha=0.9)
[7.0, 5.0]
Не сходится 79 steps
</td></tr>
</table>

- Как отличается поведение методов в зависимости от числа обусловленности функции, выбора начальной точки и стратегии выбора шага

Исследуйте, как зависит число итераций, необходимое градиентному спуску для сходимости, от следующих двух параметров:

(a) числа обусловленности $k \leq 1$ оптимизируемой функции

(b) размерности пространства n оптимизируемых переменных.

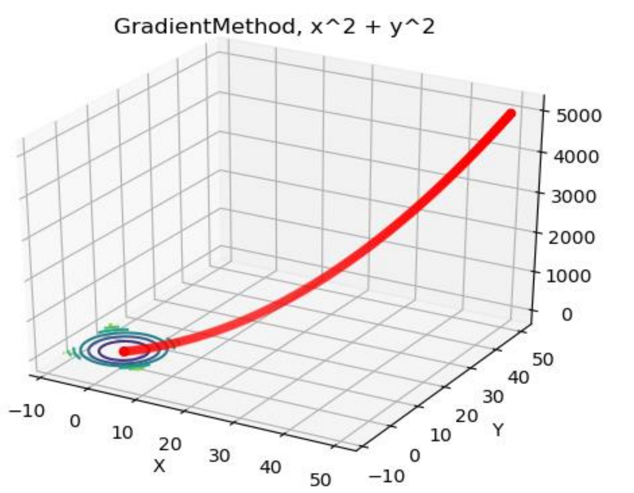
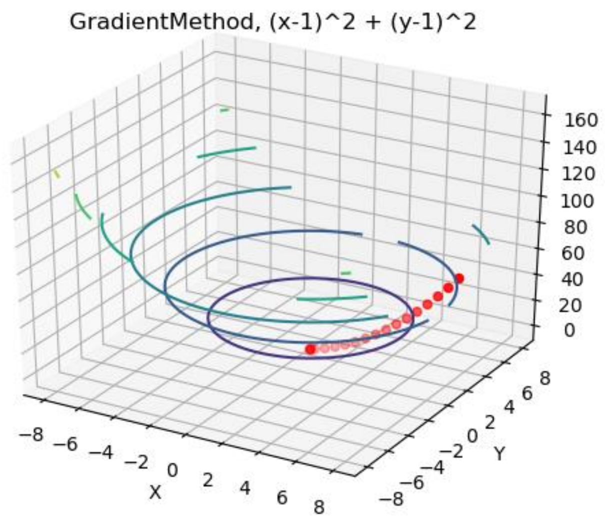
Для этого для заданных параметров n и k сгенерируйте случайным образом квадратичную задачу размера n с числом обусловленности k и запустите на ней градиентный спуск с некоторой фиксированной требуемой точностью. Замерьте число итераций $T(n, k)$, которое потребовалось сделать методу до сходимости (успешному выходу по критерию остановки).

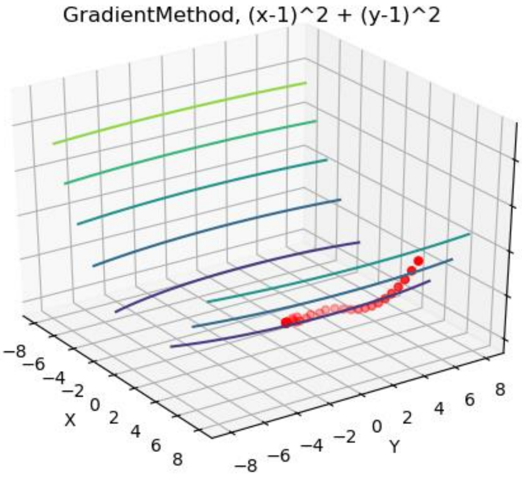
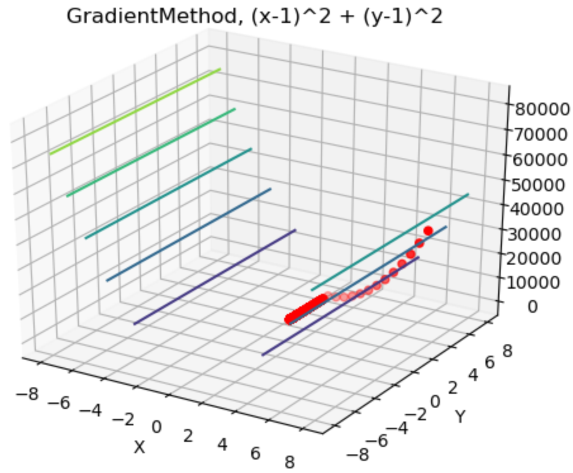
Размерность пространства n

Увеличение размерности пространства n не влияет на градиентный спуск. Более значим критерием является расстояние до точки искомого минимума.

Число обусловленности k

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

k = 0.1, точность = 0.1 29 steps 	k = 1, точность = 0.001 19 steps 

<table>
<tr>
<td>k = 10, точность = 0.001
37 steps

</td>
<td>k = 1000, точность = 0.001
2037 steps

</td>
</tr>
</table>

Вывод: при малых k уменьшается точность полученных значений, ввиду того, что одним из критериев остановки является малая длина градиента и метод слишком быстро останавливается не достигнув нужной точности.