

Национальный исследовательский университет ИТМО  
Факультет информационных технологий и программирования  
Прикладная математика и информатика

**Методы оптимизации**

**Отчет по лабораторной работе №1**

Работу выполнил:

Лаптев Александр М34371

Санкт-Петербург  
2023

# 1 Постановка задачи

1. Реализовать градиентный спуск с постоянным шагом.
2. Реализовать метод одномерного поиска (метод дихотомии, метод Фибоначчи, метод золотого сечения) и градиентный спуск на его основе.
3. Проанализировать траекторию градиентного спуска на примере двух-трёх квадратичных функций, на которых работа методов отличается.
4. Для каждой функции:
  - исследовать сходимость градиентного спуска с постоянным шагом, сравнить полученные результаты для выбранных функций
  - сравнить эффективность градиентного спуска с использованием одномерного поиска с точки зрения количества вычислений минимизируемой функции и ее градиентов
  - исследовать работу методов в зависимости от выбора начальной точки
  - исследовать влияние нормализации (scaling) на сходимость на примере масштабирования осей плохо обусловленной функции
  - в каждом случае нарисовать графики с линиями уровня и траекториями методов
5. Реализовать генератор случайных квадратичных функций  $n$  переменных с числом обусловленности  $k$ .
6. Исследовать зависимость числа итераций  $T(n, k)$ , необходимых градиентному спуску для сходимости в зависимости от размерности пространства  $2 \leq n \leq 10^3$  и числа обусловленности оптимизируемой функции  $1 \leq k \leq 10^3$ . Провести множественный эксперимент.
7. Реализовать одномерный поиск с учетом условий Вольфе и исследовать его эффективность. Сравнить полученные результаты с реализованными ранее методами.

## 2 Описание используемых методов

### 2.1 Градиентный спуск с постоянным шагом

#### Описание метода

Для нахождения локального минимума функции, точка итеративно сдвигается в направлении антиградиента с постоянным шагом. С помощью данной процедуры, значение оптимизируемой функции уменьшается и метод достигает локального минимума функции.

#### Алгоритм

Входные данные:

1. Функция  $f: \mathbf{R}^n \mapsto \mathbf{R}$  и начальная точка  $x \in \mathbf{R}^n$
2. Шаг градиентного спуска  $\alpha$
3. Максимальное значение нормы градиента функции  $\epsilon$

Алгоритм:

1.  $x = x - \alpha \nabla f(x)$
2. Если  $\|\nabla f(x)\| < \epsilon$  завершаем работу, иначе возвращаемся к первому шагу

### 2.2 Градиентный спуск с использованием метода золотого сечения

#### 2.2.1 Метод золотого сечения

#### Описание метода

Метод золотого сечения используется для нахождения точки минимума унимодальный функции от одной переменной на заданном отрезке. Разделив отрезок на 3 части двумя точками в отношении золотого сечения и вычислив значение функции в этих точках, можно понять в какой из частей лежит точка минимума функции. Итеративно уменьшая отрезок, метод достигает точки минимума.

## Алгоритм

Входные данные:

1. Функция  $f: \mathbf{R} \mapsto \mathbf{R}$  и отрезок  $[l, r] \subset \mathbf{R}$
2. Точность  $\epsilon$

Алгоритм:

Определим константу  $\phi = \frac{1}{\varphi+1} = \frac{3-\sqrt{5}}{2}$ , где  $\varphi$  - значение золотого сечения.

1. Делим отрезок  $[l, r]$  двумя точками в отношении золотого сечения:  $a = l + \phi(r - l)$  и  $b = r - \phi(r - l)$
2. Если  $f(a) > f(b)$  то  $l = a$ , иначе  $r = b$
3. Если  $r - l < \epsilon$  завершаем работу, ответом является точка  $\frac{r-l}{2}$ , иначе возвращаемся к первому шагу

## 2.2.2 Метод градиентного спуска с использованием метода золотого сечения

### Описание метода

Для нахождения локального минимума функции, точка итеративно сдвигается в точку минимума по направлению антиградиента с помощью метода золотого сечения. Аналогично градиентному спуску с постоянным шагом, значение оптимизируемой функции уменьшается и метод достигает локального минимума функции.

## Алгоритм

Входные данные:

1. Функция  $f: \mathbf{R}^n \mapsto \mathbf{R}$  и начальная точка  $x \in \mathbf{R}^n$
2. Максимальное значение шага градиентного спуска  $\alpha$
3. Максимальное значение нормы градиента функции  $\epsilon$

Алгоритм:

1. Находим точку  $\gamma = \underset{\beta \in [0, \alpha]}{\operatorname{argmin}} f(x - \beta \nabla f(x))$  с помощью метода золотого сечения
2.  $x = x - \gamma \nabla f(x)$
3. Если  $\|\nabla f(x)\| < \epsilon$  завершаем работу, иначе возвращаемся к первому шагу

## 2.3 Градиентный спуск с условием Вольфе

### Описание метода

Для нахождения локального минимума функции, точка итеративно сдвигается в точку минимума по направлению антиградиента. Шаг градиентного спуска  $\alpha$  подбирается на основании условия Вольфе.

Условие Вольфе:

1.  $f(x_k + \alpha \nabla f(x_k)) \leq f(x_k) - c_1 \alpha \nabla f(x_k) \nabla f(x_k)^T$  - с увеличением шага, максимальное значение в новой точке должно уменьшаться.
2.  $\nabla f(x_k + \alpha \nabla f(x_k)) \nabla f(x_k)^T \geq c_2 \alpha \nabla f(x_k) \nabla f(x_k)^T$  - значение градиента в новой точке, должно быть больше чем в старой.

## Алгоритм

Входные данные:

1. Функция  $f: \mathbf{R}^n \mapsto \mathbf{R}$  и начальная точка  $x \in \mathbf{R}^n$
2. Параметры  $c_1, c_2$ , такие что  $c_1 \leq c_2$
3. Максимальное значение нормы градиента функции  $\epsilon$

Алгоритм:

1. Находим значение шага  $\alpha$  удовлетворяющее условию Вольфе
2.  $x = x - \alpha \nabla f(x)$
3. Если  $\|\nabla f(x)\| < \epsilon$  завершаем работу, иначе возвращаемся к первому шагу

### 3 Анализ работы методов на квадратичных функциях от двух переменных

Проанализируем работу методов на примере трёх квадратичных функций от двух переменных:

1.  $f(x) = x^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x$
2.  $f(x) = x^T \begin{bmatrix} 0.09 & 0 \\ 0 & 1 \end{bmatrix} x$
3.  $f(x) = x^T \begin{bmatrix} 0.2 & 0.4 \\ 0.4 & 1 \end{bmatrix} x$

с числами обусловленности 1, 11.1111, 33.9705 соответственно.

Начальная точка:  $x_0 = (-0.9, -0.1)$

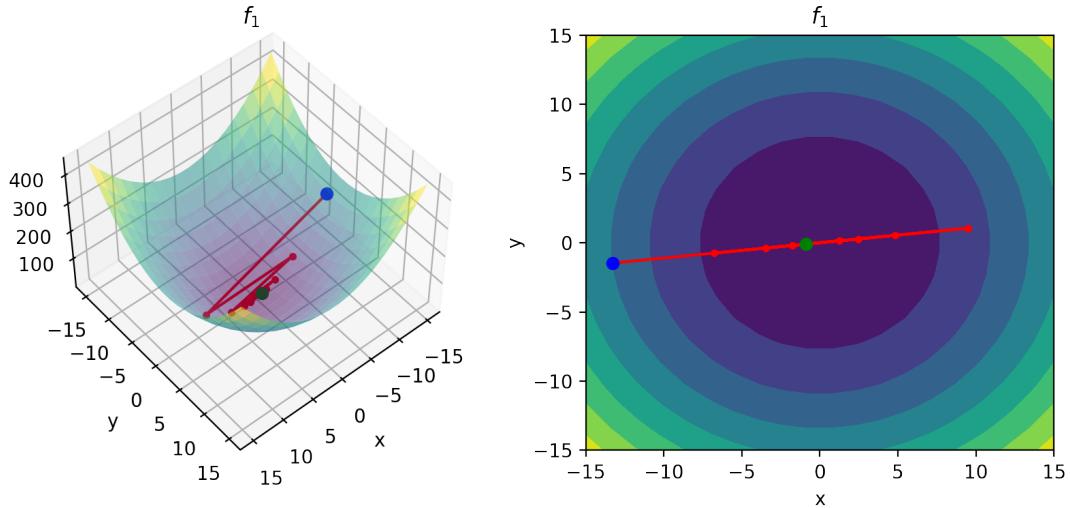
Далее будем полагать, что максимальное значение нормы градиента  $\epsilon = 0.0001$  для всех методов.

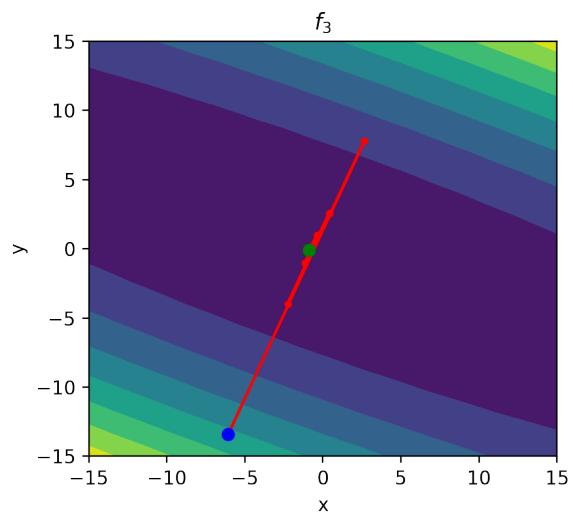
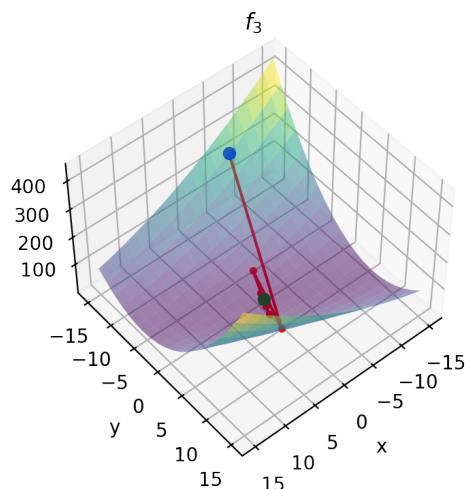
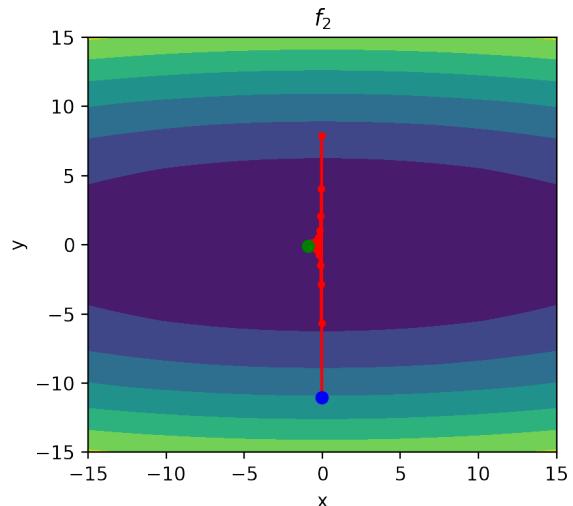
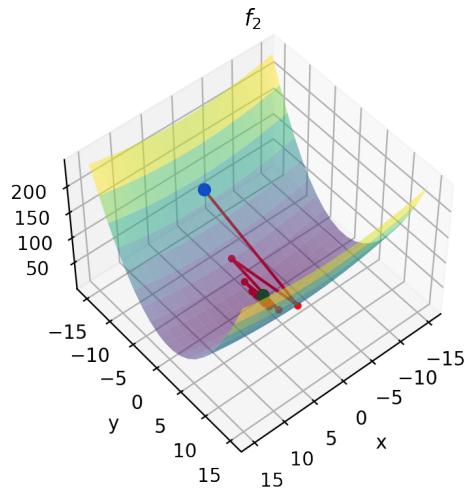
Также, на всех графиках траектория градиентного спуска будет нарисована красным цветом, начальная точка - зеленым, конечная - синим.

#### 3.1 Анализ сходимости градиентного спуска с постоянным шагом

Рассмотрим траекторию градиентного спуска и количество итераций необходимых для сходимости для различных значений шагов градиентного спуска:  $\alpha = 1.2, 0.85, 0.4, 0.1$ .

##### 3.1.1 $\alpha = 1.2$

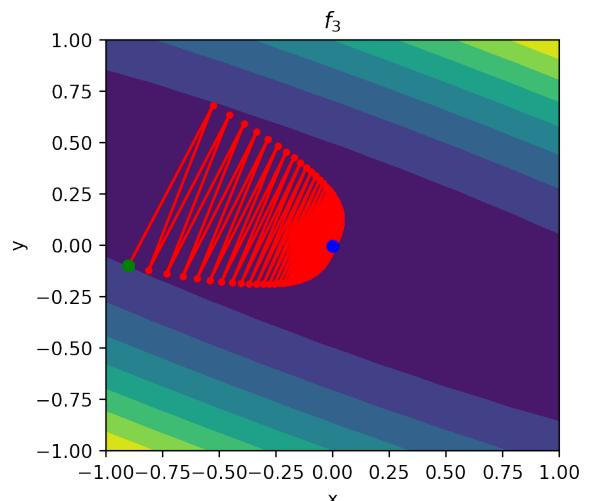
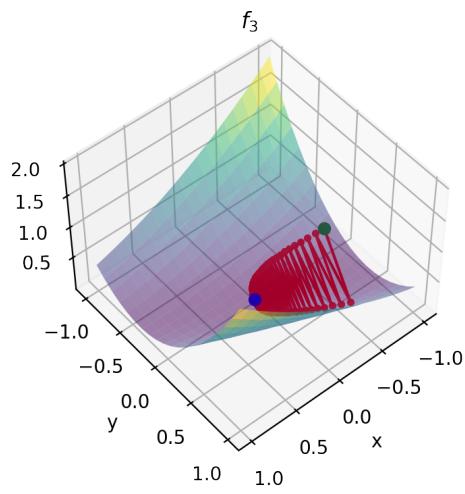
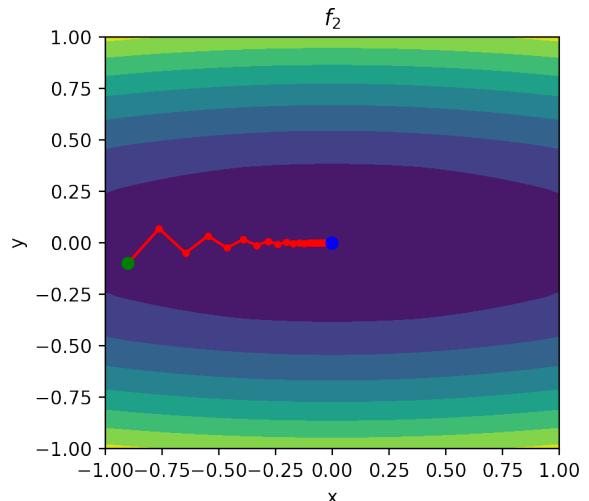
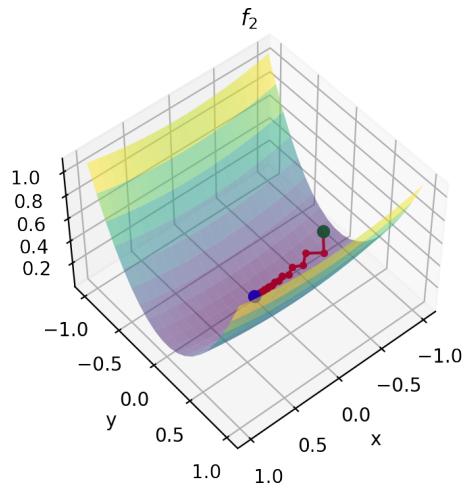
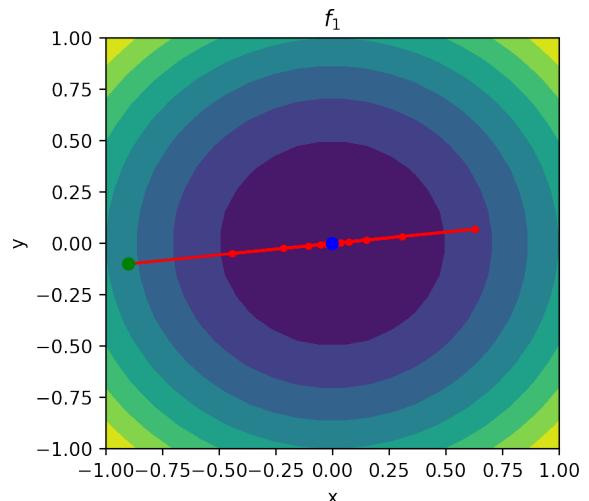
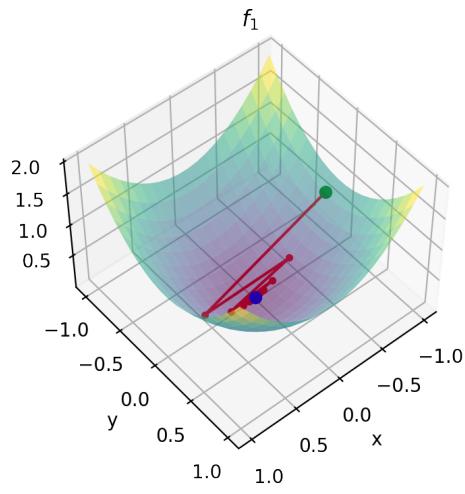




Функция	Количество итераций	Точка минимума	Значение в точке минимума	Количество вычислений функции	Количество вычислений градиента
$f_1$	$\infty$	-	-	$\infty$	$\infty$
$f_2$	$\infty$	-	-	$\infty$	$\infty$
$f_3$	$\infty$	-	-	$\infty$	$\infty$

В данном примере, градиентный спуск расходится на всех функциях, за счёт слишком большого значения шага градиентного спуска.

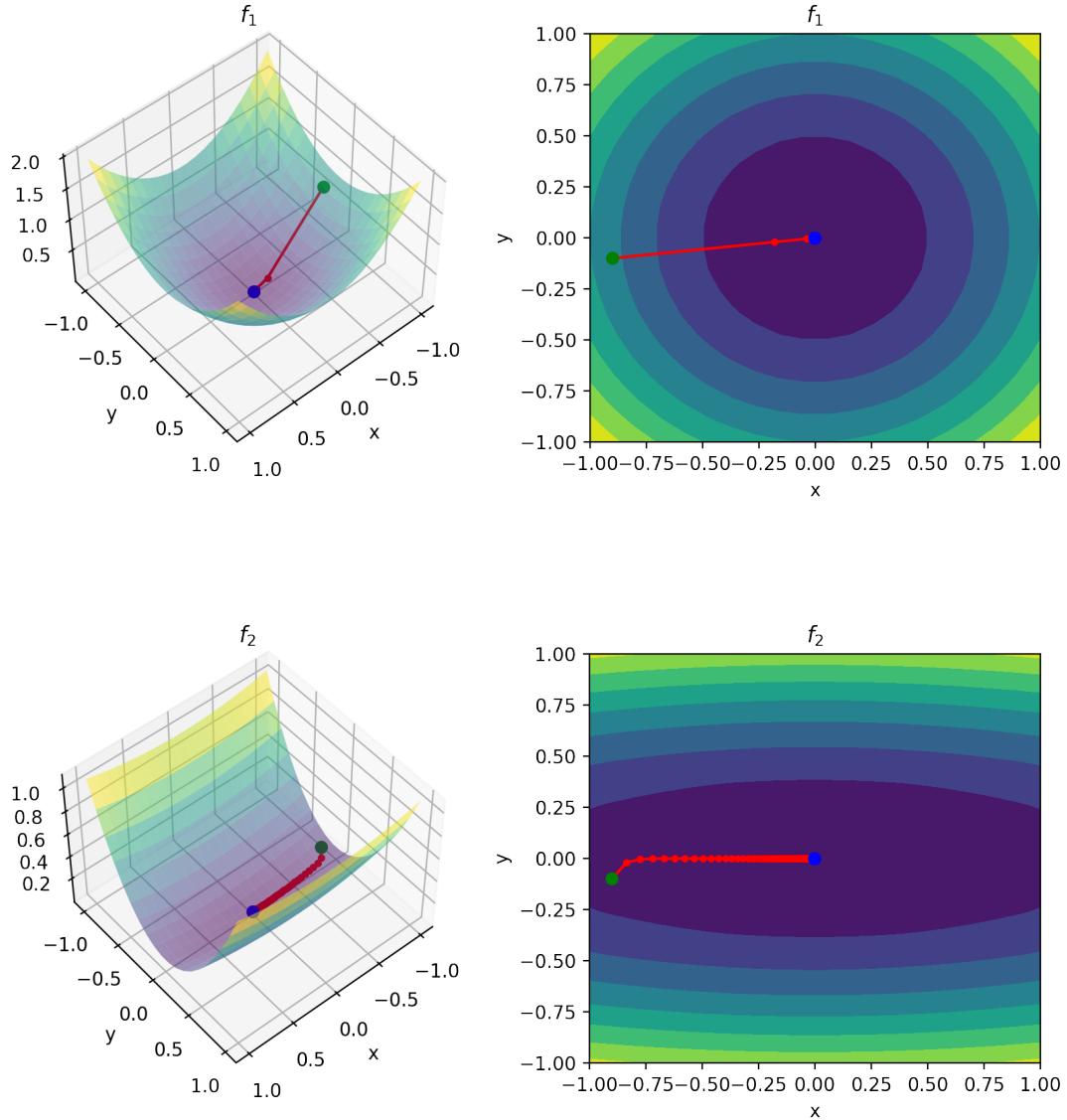
### 3.1.2 $\alpha = 0.85$

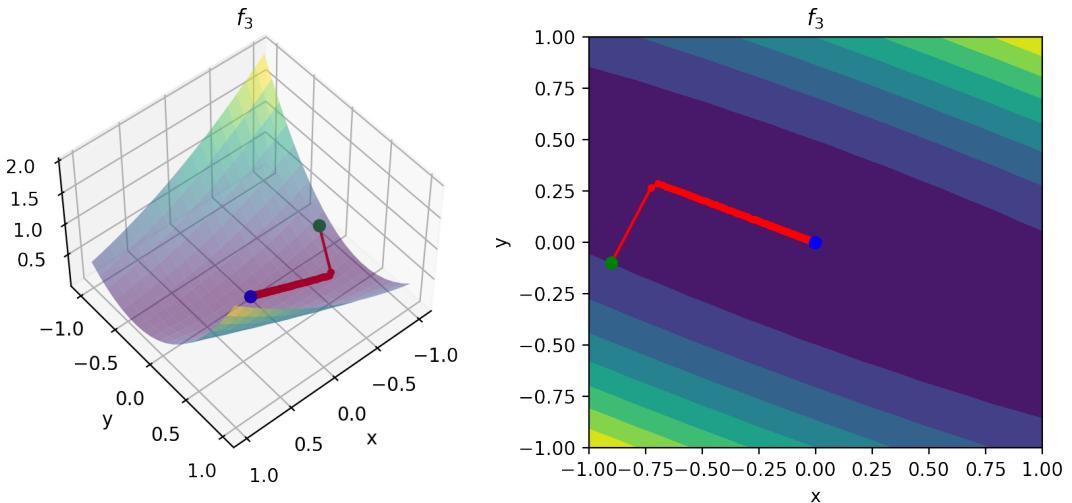


Функция	Количество итераций	Точка минимума	Значение в точке минимума	Количество вычислений функции	Количество вычислений градиента
$f_1$	28	(-0.001, -0.001)	0	0	28
$f_2$	44	(-0.001, -0.001)	0	0	44
$f_3$	490	(0.002, -0.002)	0	0	490

В данном примере, градиентный спуск сходится к точке минимума для всех функций. Для первой функции траектория градиентного спуска постоянно “перепрыгивает” точку минимума, за счёт большого шага градиентного спуска. Для второй и третьей функций, наблюдается зигзагообразная траектория и увеличение количества итераций с увеличением числа обусловленности, за счёт медленной сходимости по оси  $x$ . Также заметим, что для ускорения сходимости по оси  $x$  значение шага не может быть существенно увеличено для данных функций, так как в этом случае градиентный спуск будет расходиться, аналогично первому примеру.

### 3.1.3 $\alpha = 0.4$

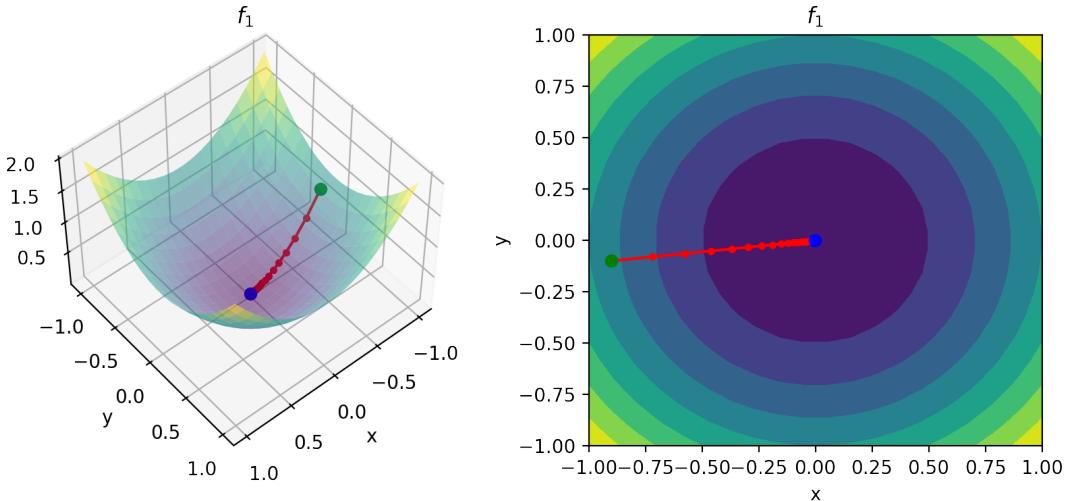


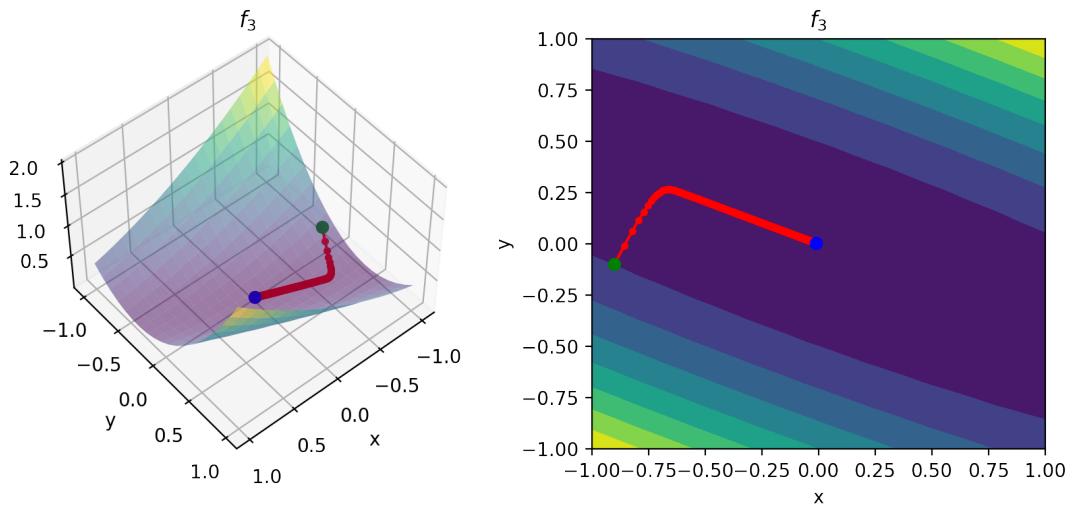
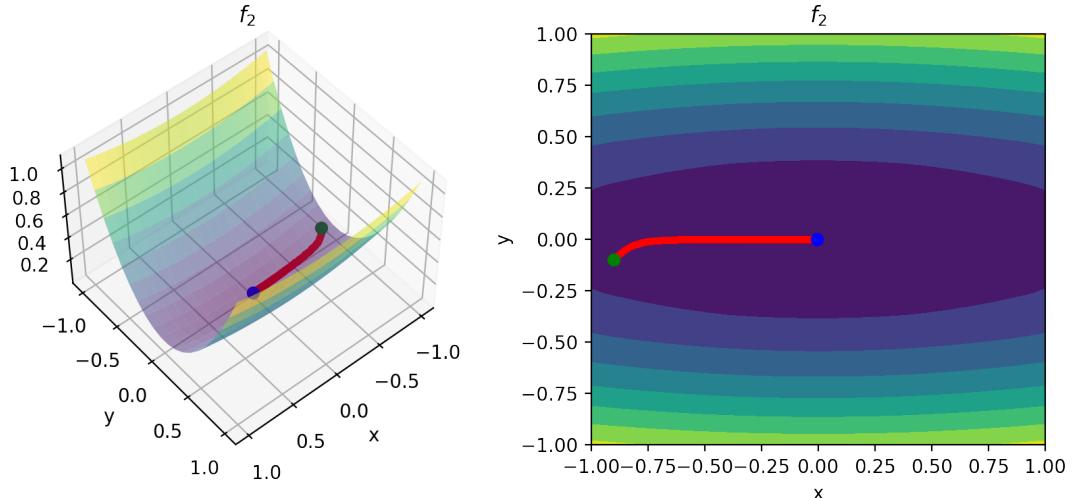


Функция	Количество итераций	Точка минимума	Значение в точке минимума	Количество вычислений функции	Количество вычислений градиента
$f_1$	6	(-0.001, -0.001)	0	0	6
$f_2$	87	(-0.002, 0)	0	0	87
$f_3$	194	(-0.001, 0)	0	0	194

В данном примере, градиентный спуск сходится к точке минимума для всех функций. Для первой и третьей функции уменьшилось количество итераций необходимых для сходимости градиентного спуска. Также, для второй и третьей функции, траектория градиентного спуска не является зигзагообразной, по сравнению с предыдущим примером. Аналогично предыдущему примеру, количество итераций возрастает с увеличением числа обусловленности.

### 3.1.4 $\alpha = 0.1$





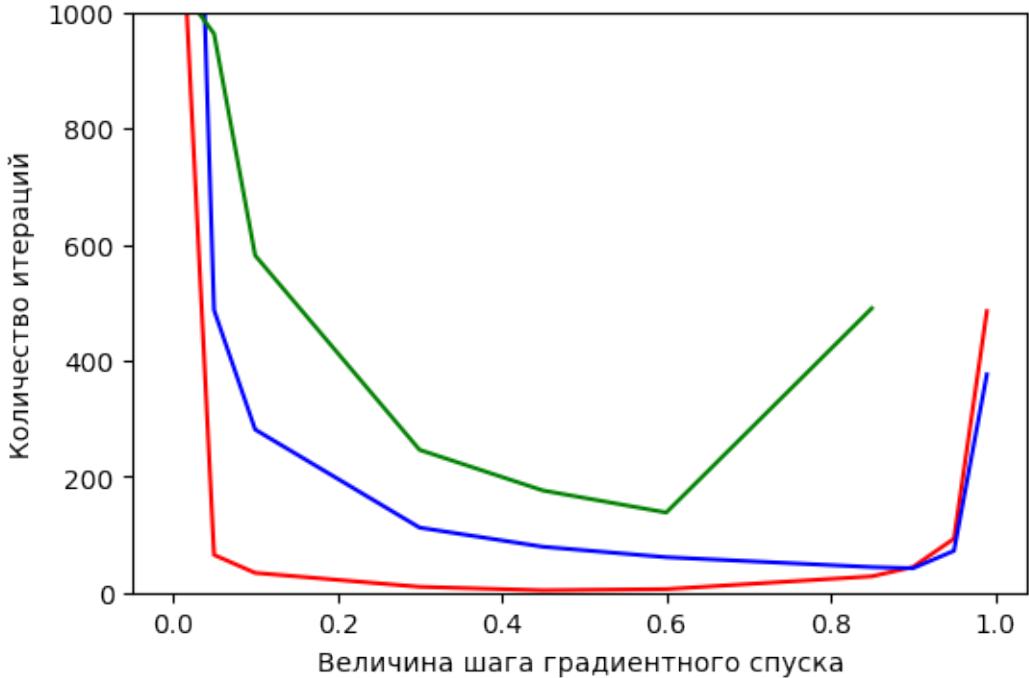
Функция	Количество итераций	Точка минимума	Значение в точке минимума	Количество вычислений функции	Количество вычислений градиента
$f_1$	34	(-0.001, -0.001)	0	0	34
$f_2$	281	(-0.006, 0)	0	0	281
$f_3$	581	(-0.011, 0.004)	0	0	581

В данном примере, градиентный спуск сходится к точке минимума для всех функций. За счёт маленького значения шага градиентного спуска, траектория равномерно движется к точке минимума. Количество итераций необходимых для сходимости градиентного спуска возросло по сравнению с предыдущим примером из-за маленького значения шага.

### 3.1.5 Анализ на других значениях шага градиентного спуска

Проанализируем количество итераций необходимых для сходимости в зависимости от размера шага. Для этого рассмотрим шаги равные 1.3, 0.99, 0.95, 0.9, 0.85, 0.6, 0.45, 0.3, 0.1, 0.05, 0.001.

Функция	1.3	0.99	0.95	0.9	0.85	0.6	0.45	0.3	0.1	0.05	0.001
$f_1$	$\infty$	485	93	44	28	6	4	10	34	65	1447
$f_2$	$\infty$	376	72	42	44	61	79	112	281	487	2678
$f_3$	$\infty$	$\infty$	$\infty$	$\infty$	490	138	176	246	581	963	1058



На основе данных из таблицы и графика зависимости количества итераций от значения шага градиентного спуска, можно сделать вывод, что количество итераций изначально убывает с возрастанием градиентного спуска, а после возрастает вплоть до того, что градиентный спуск перестает сходиться. Также, с увеличением числа обусловленности возрастает количество итераций, необходимых для сходимости.

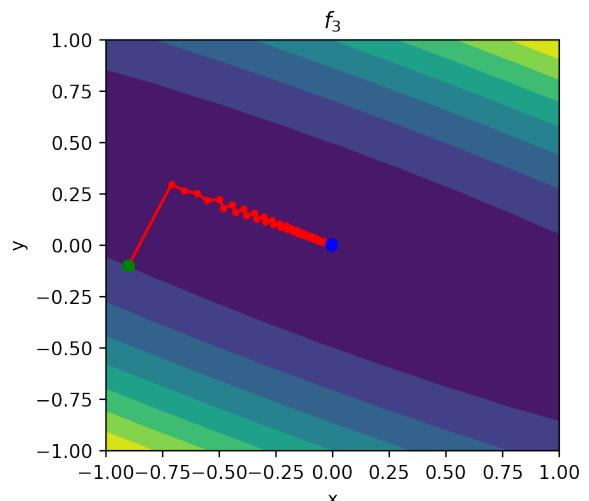
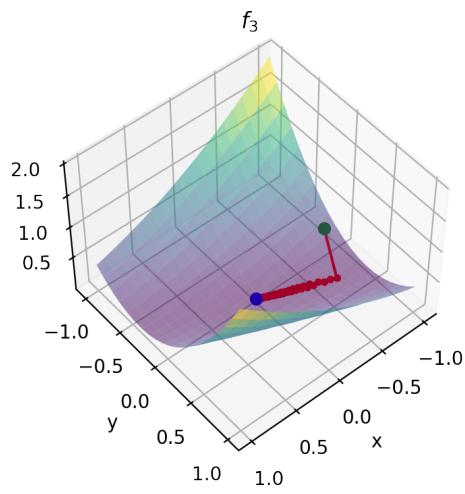
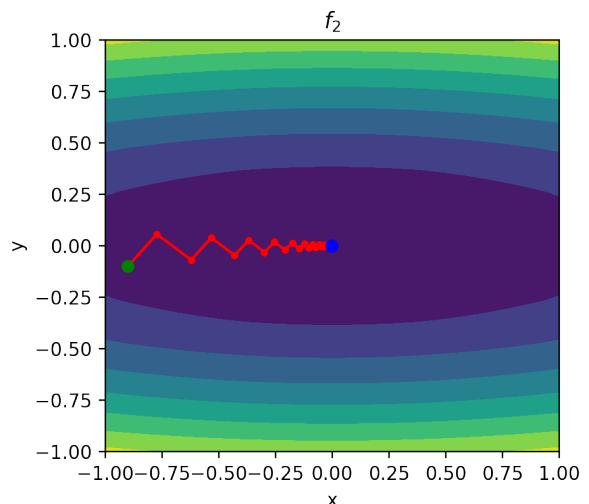
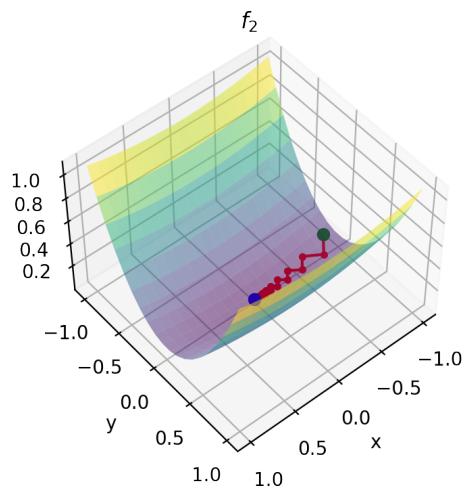
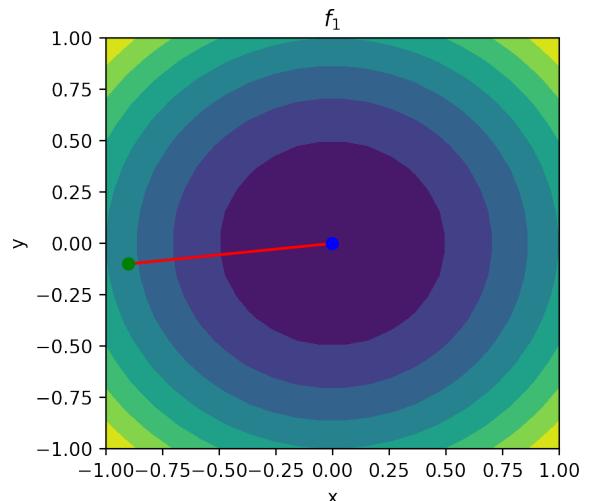
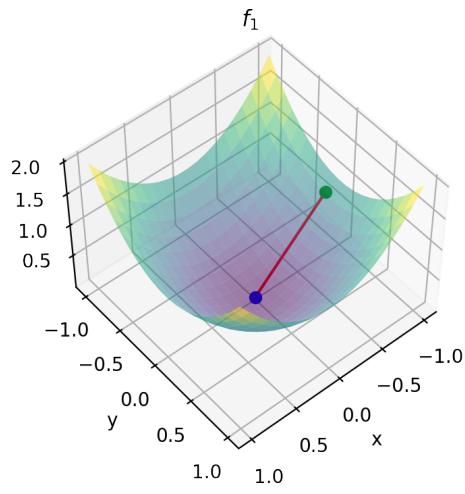
Таким образом, можно сформулировать общие выводы:

- Значение шага влияет на сходимость градиентного спуска.
- При уменьшении значения шага, с некоторого момента увеличивается количество итераций.
- При увеличении значения шага, с некоторого момента увеличивается количество итераций, вплоть до расходимости.
- При увеличении числа обусловленности функции, увеличивается количество итераций необходимых для сходимости.

### 3.2 Анализ сходимости градиентного спуска с использованием метода золотого сечения

Рассмотрим траекторию градиентного спуска и количество итераций необходимых для сходимости градиентному спуску с использованием метода золотого сечения, с максимальным значением шага  $\alpha = 1.2, 0.4, 0.1$

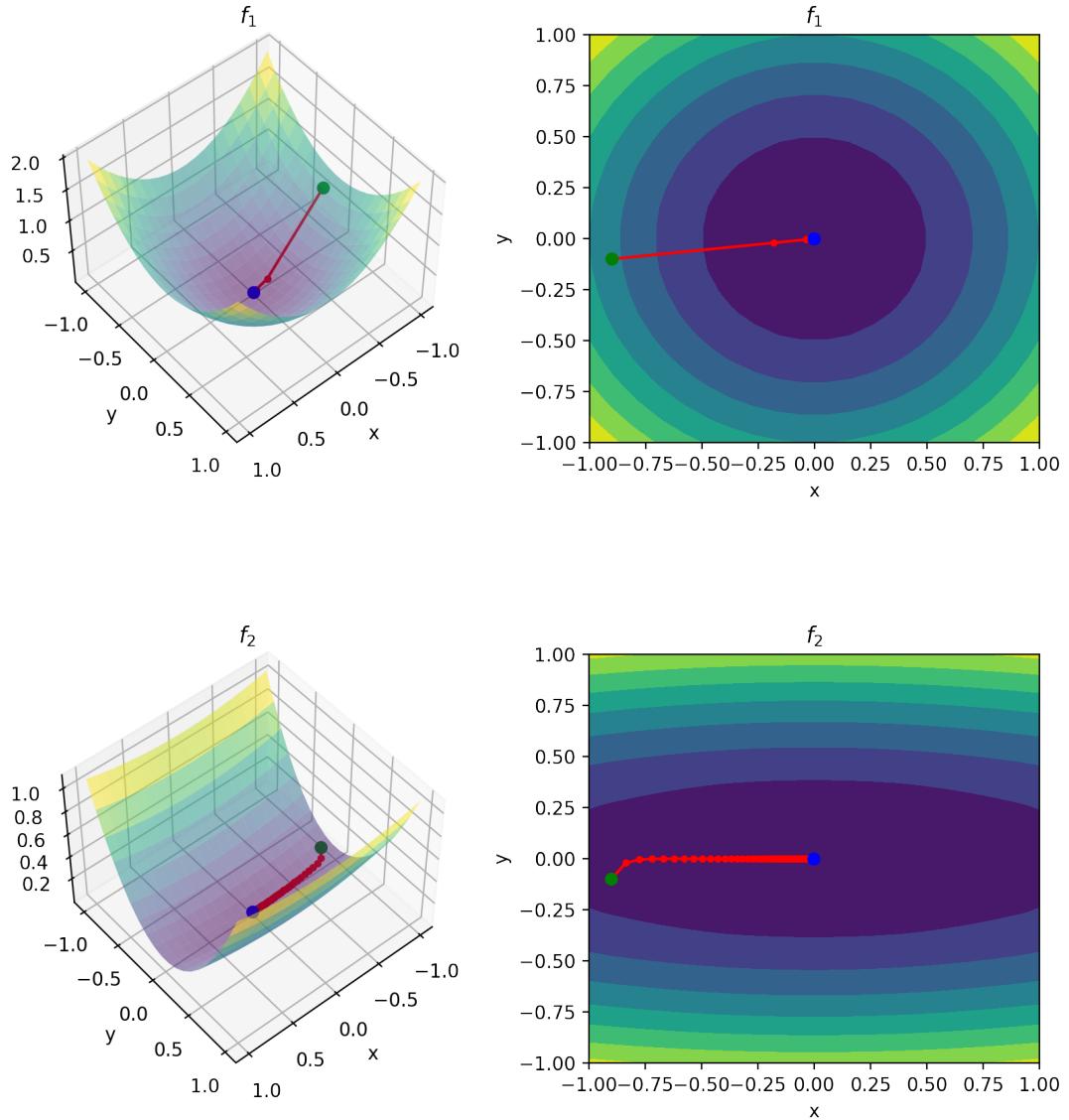
### 3.2.1 $\alpha = 1.2$

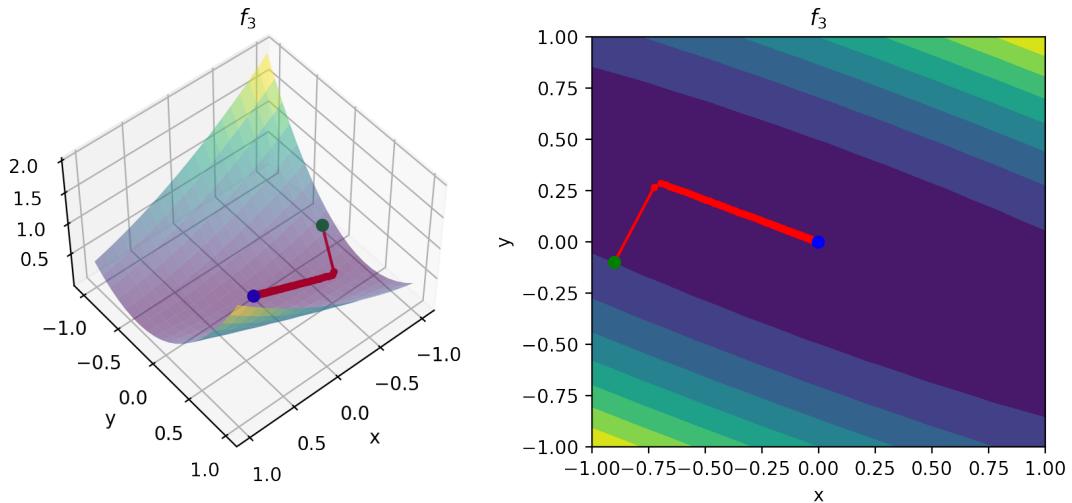


Функция	Количество итераций	Точка минимума	Значение в точке минимума	Количество вычислений функции	Количество вычислений градиента
$f_1$	1	(0, 0)	0	17	1
$f_2$	31	(-0.003, 0)	0	527	31
$f_3$	68	(-0.006, 0.002)	0	1156	68

В приведённом примере, градиентный спуск с методом золотого сечения сходится к точке минимума для всех функций. Также с увеличением числа обусловленности возрастает количество итераций и количество вычислений функции необходимых для сходимости и траектория принимает зигзагообразный вид.

### 3.2.2 $\alpha = 0.4$

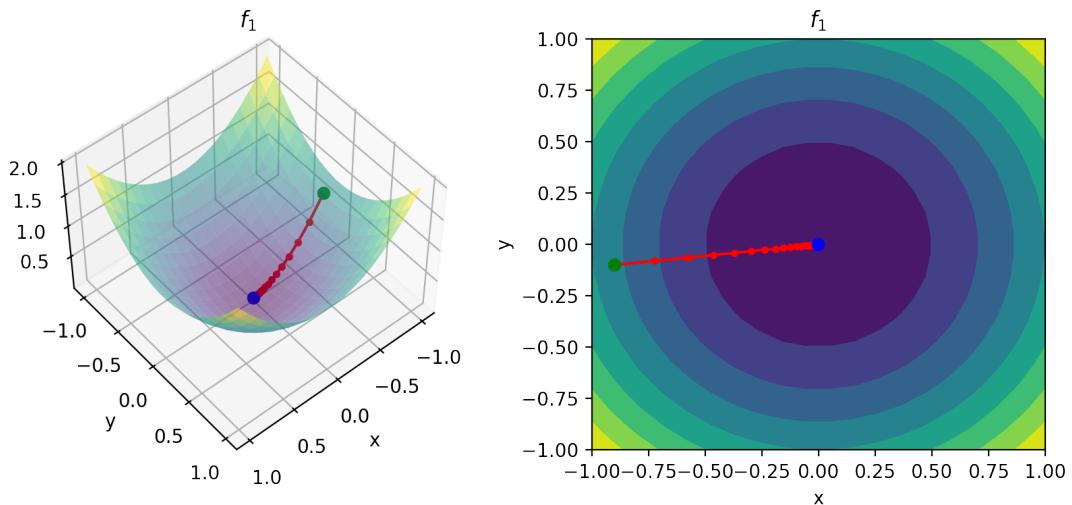


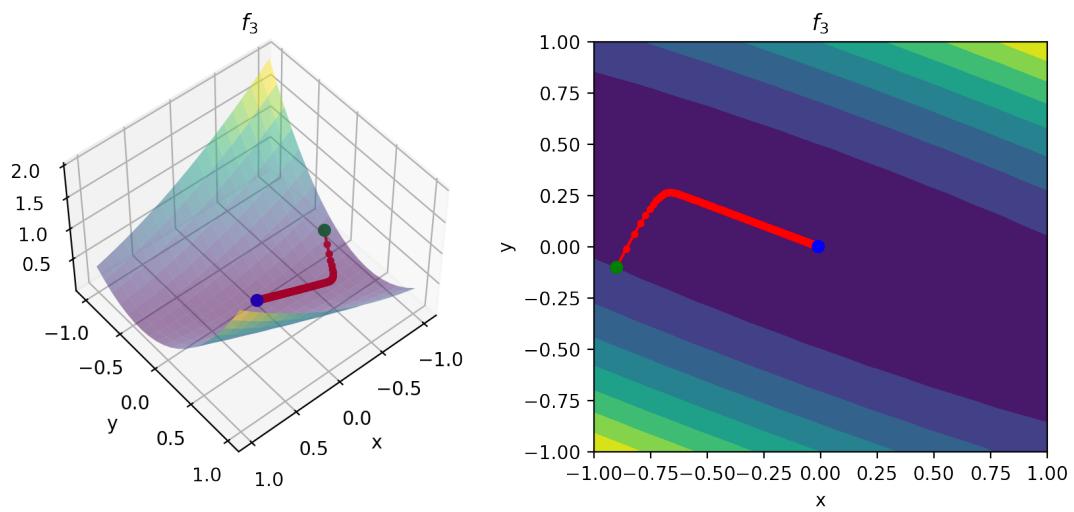
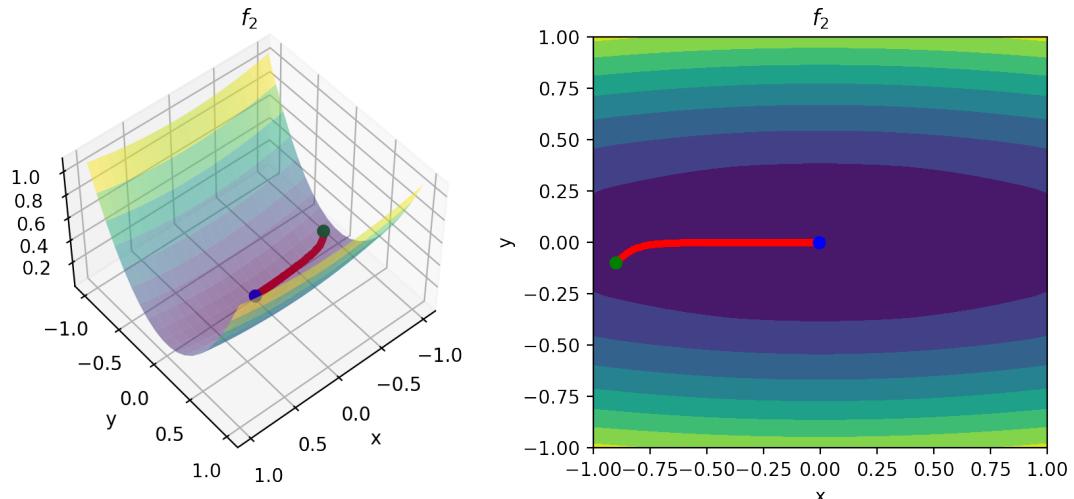


Функция	Количество итераций	Точка минимума	Значение в точке минимума	Количество вычислений функции	Количество вычислений градиента
$f_1$	6	(-0.001, -0.001)	0	90	6
$f_2$	87	(-0.002, 0)	0	1305	87
$f_3$	194	(-0.001, 0)	0	2910	194

В приведённом примере, градиентный спуск с методом золотого сечения сходится к точке минимума для всех функций. За счёт уменьшения значения максимального шага, возросло количество итераций и вычислений функции. Также, для плохо обусловленных функций, траектория перестала быть зигзагообразной.

### 3.2.3 $\alpha = 0.1$





Функция	Количество итераций	Точка минимума	Значение в точке минимума	Количество вычислений функции	Количество вычислений градиента
$f_1$	34	(-0.001, -0.001)	0	408	34
$f_2$	281	(-0.006, 0)	0	3372	281
$f_3$	583	(-0.011, 0.004)	0	6996	583

В приведённом примере, градиентный спуск с методом золотого сечения сходится к точке минимума для всех функций. Траектория эквивалентна примеру с  $\alpha = 0.4$ , но количество итераций и количество вычислений функции увеличилось.

### 3.2.4 Анализ на разных значениях максимального шага

#### Количество итераций/вычислений градиентов

Функция	50.0	10.0	3.0	2.5	1.2	0.85	0.6	0.4	0.1	0.05
$f_1$	2	2	1	1	1	2	2	6	34	66
$f_2$	22	22	24	24	31	44	61	87	281	489
$f_3$	23	27	39	43	68	93	134	194	583	967

#### Количество вычислений функции

Функция	50.0	10.0	3.0	2.5	1.2	0.85	0.6	0.4	0.1	0.05
$f_1$	50	44	19	19	17	34	32	90	408	726
$f_2$	550	484	456	456	527	748	976	1305	3372	5379
$f_3$	575	594	741	817	1156	1581	2144	2910	6996	10637

На основе данных из таблицы, можно сделать вывод, что при увеличении максимального значения шага, возрастает количество вычислений функции необходимых для сходимости, за счёт того что интервал поиска оптимального значения шага увеличивается, количество итераций же уменьшается. При уменьшении значения максимального шага, увеличивается количество итераций необходимых для сходимости, а следовательно и количество вычислений функции. Также с увеличением числа обусловленности функции возрастает количество итераций и количество вычислений функции.

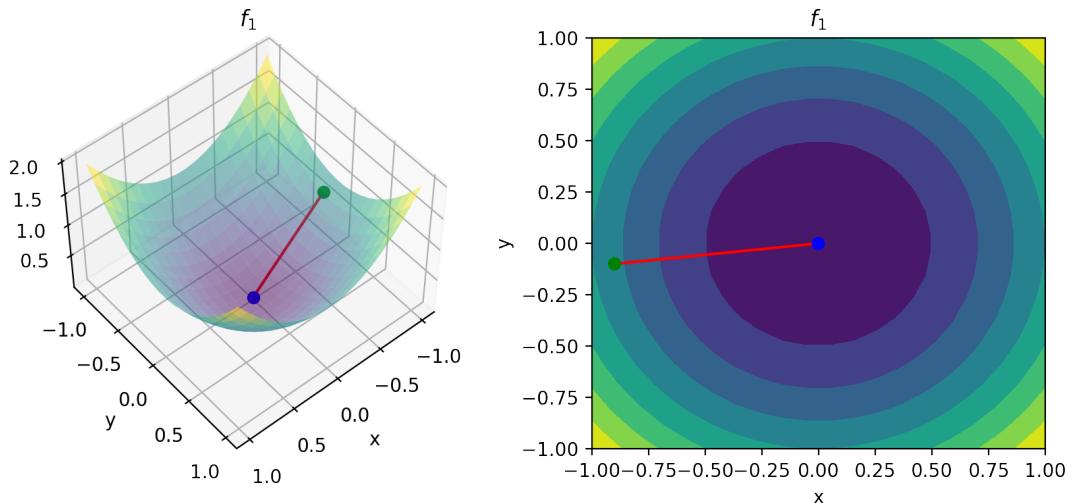
Таким образом, можно сформулировать общие выводы:

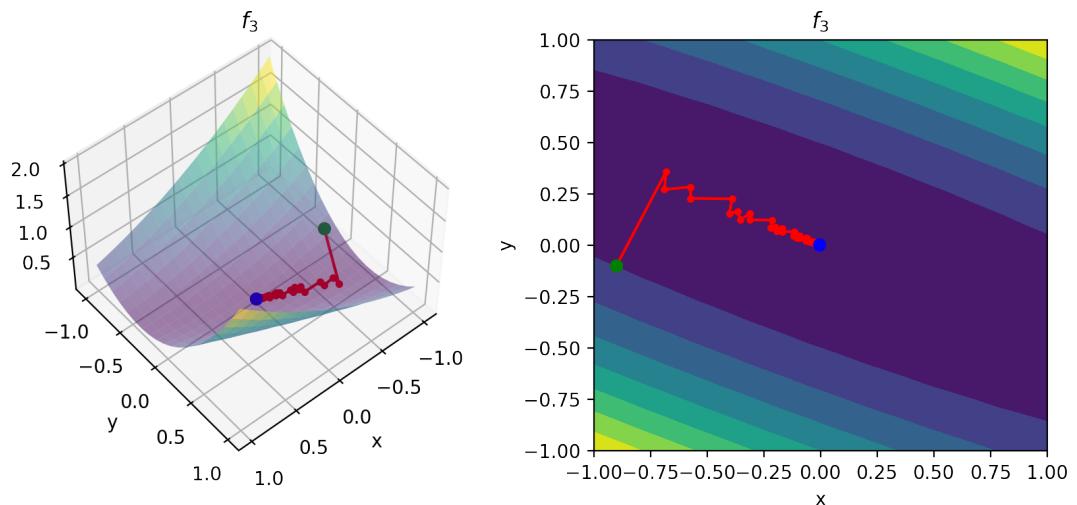
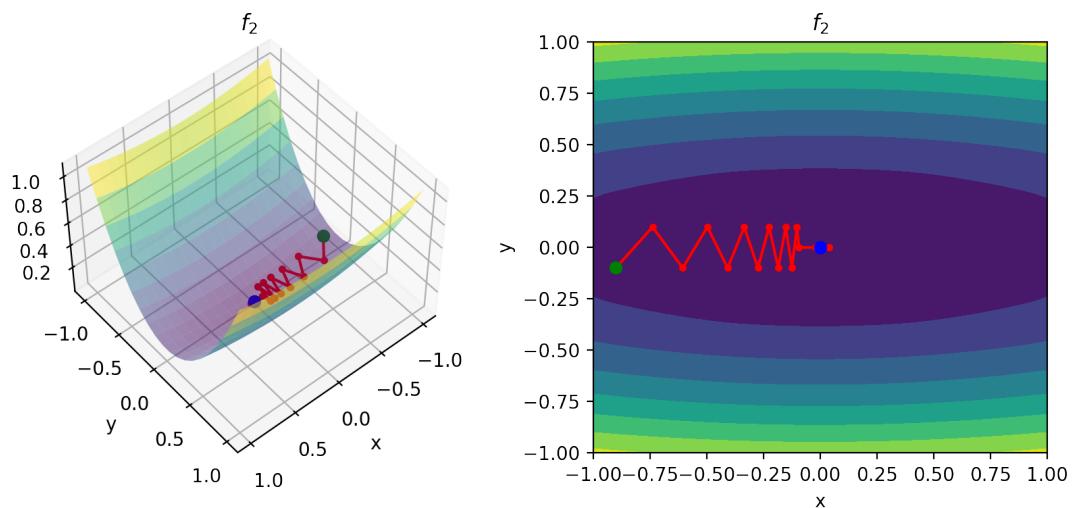
- При выборе достаточно большого значения максимального шага, можно сократить количество итераций необходимых для сходимости, за счёт увеличения количества вычислений функции.
- При увеличении числа обусловленности функции, возрастает количество итераций и количество вычислений функции необходимых для сходимости.

### 3.3 Анализ градиентного спуска с условием Вольфе

Рассмотрим сходимость градиентного спуска с условием Вольфе с параметрами  $c_1 = 0.001$ ,  $c_2 = 0.1$  и  $c_1 = 0.3$ ,  $c_2 = 0.9$

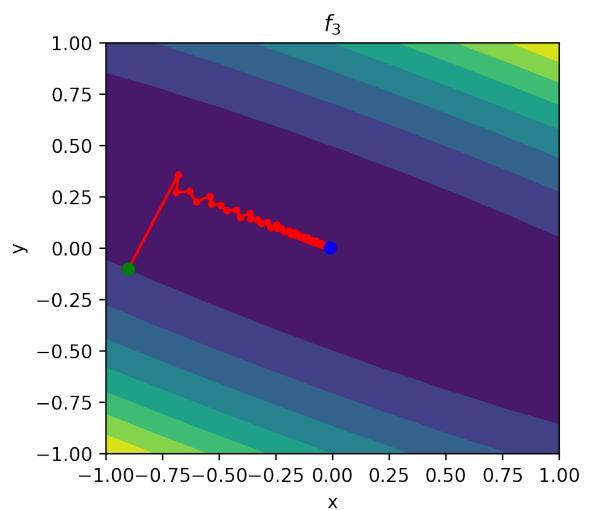
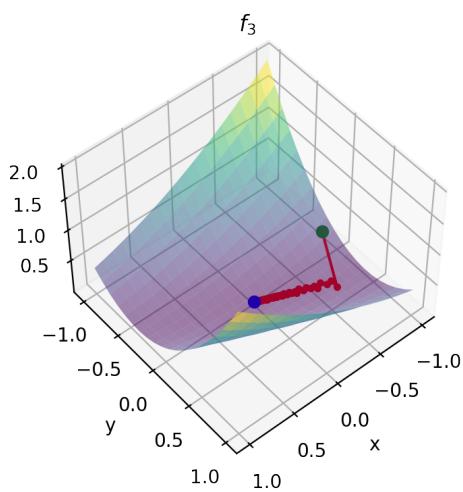
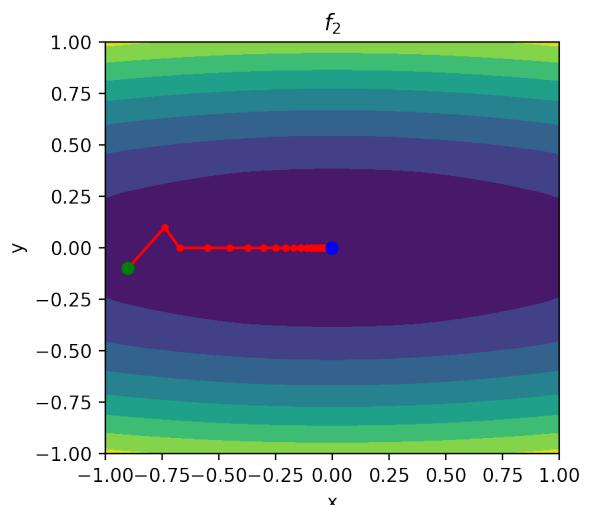
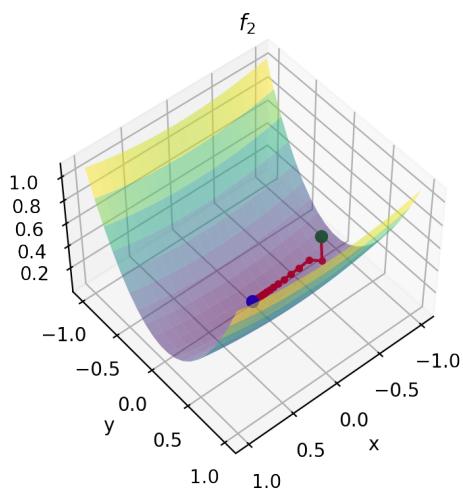
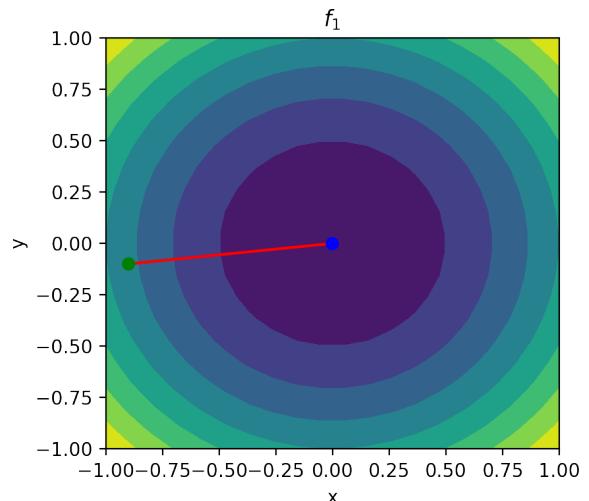
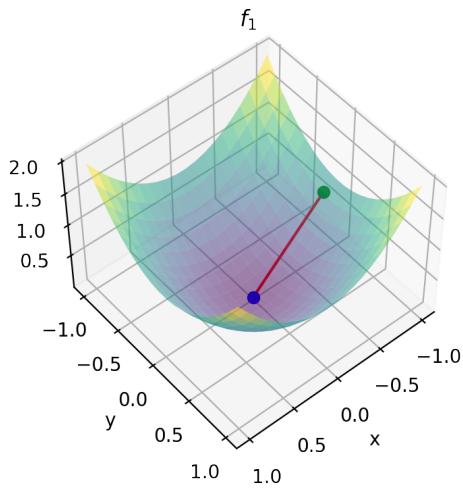
#### 3.3.1 $c_1 = 0.001$ , $c_2 = 0.1$





Функция	Количество итераций	Точка минимума	Значение в точке минимума	Количество вычислений функции	Количество вычислений градиента
$f_1$	1	(-0.001, 0)	0	3	3
$f_2$	19	(0, 0)	0	61	78
$f_3$	45	(-0.005, 0.002)	0	130	155

**3.3.2**  $c_1 = 0.3, c_2 = 0.9$



Функция	Количество итераций	Точка минимума	Значение в точке минимума	Количество вычислений функции	Количество вычислений градиента
$f_1$	1	(-0.001, 0)	0	3	3
$f_2$	38	(-0.001, 0)	0	77	114
$f_3$	66	(-0.009, 0.004)	0	149	198

В приведённых примерах, градиентный спуск с условием Вольфе сходится к точке минимума для всех функций. За счёт дополнительных вычислений функции и градиентов, уменьшилось количество итераций необходимых для сходимости по сравнению с градиентным спуском с постоянным шагом, но при этом количество вычислений функции уменьшилось по сравнению с градиентным спуском с методом золотого сечения.

### 3.4 Сравнение сходимости методов

Сравним количество количества итераций, вычислений градиентов и вычислений функций для трёх методов

- Градиентный спуск с постоянным шагом  $\alpha = 0.3$
- Градиентный спуск с методом золотого сечения с максимальным шагом  $\alpha = 2$
- Градиентный спуск с условием Вольфе  $c_1 = 0.3$  и  $c_2 = 0.9$

#### Количество итераций

Функция	Постоянный шаг	Условие Вольфе	Метод золотого сечения
$f_1$	10	1	2
$f_2$	112	38	27
$f_3$	246	66	53

#### Количество вычислений функции

Функция	Постоянный шаг	Условие Вольфе	Метод золотого сечения
$f_1$	0	3	36
$f_2$	0	77	486
$f_3$	0	149	954

#### Количество вычислений градиента

Функция	Постоянный шаг	Условие Вольфе	Метод золотого сечения
$f_1$	10	3	2
$f_2$	112	114	27
$f_3$	246	198	53

По представленным данным, можно сделать вывод, что градиентный спуск с постоянным шагом требует наибольшего количества итераций для сходимости. Количество итераций для сходимости можно уменьшить применим градиентный спуск с методом золотого сечения, при этом возрастёт количество вычислений функции. Уменьшить количество вычислений функции можно с помощью градиентного спуска с условием Вольфе, при это количество итераций для сходимости будет немногого больше, а количество вычислений градиента не больше чем при вычислении методом градиент-

ногого спуска с постоянным шагом.

Таким образом можно сделать общие выводы по сходимости методов:

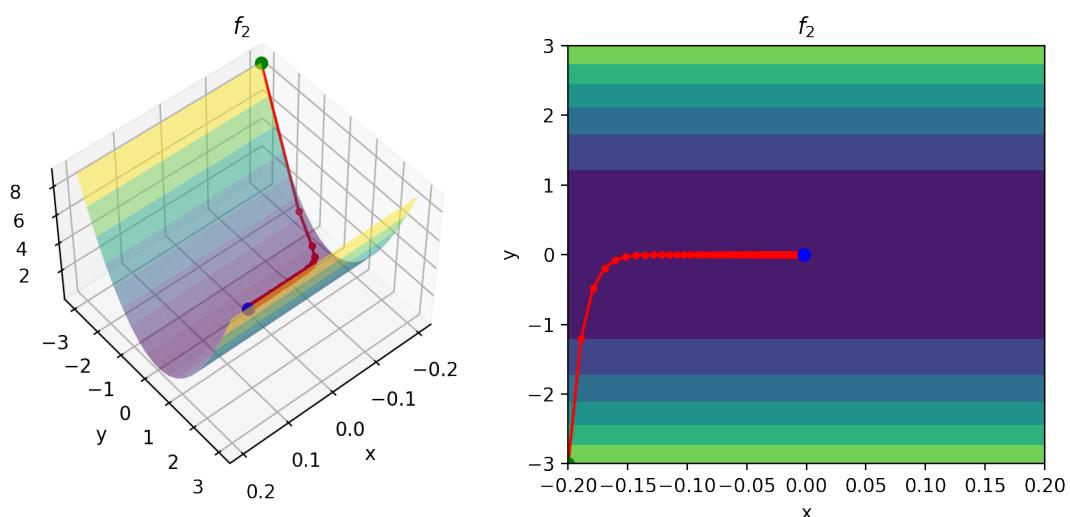
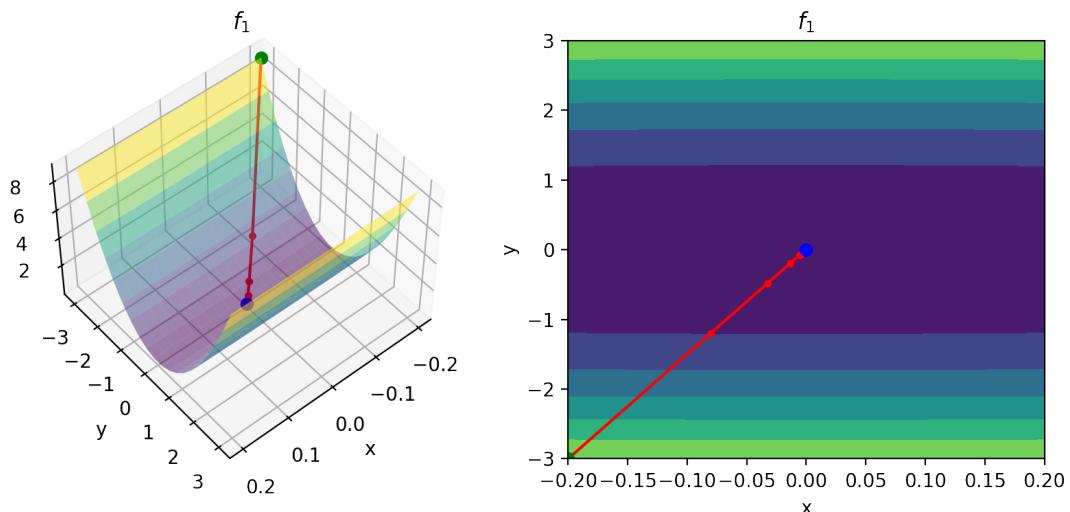
- Сходимость градиентного спуска с постоянным шагом сильно зависит от выбора шага и требует наибольшего количества итераций для сходимости по сравнению с остальными методами.
- Использование градиентного спуска с методом золотого сечения уменьшает количество итераций необходимых для сходимости, при этом возрастает количество вычислений функции.
- Использование градиентного спуска с условием Вольфе позволяет уменьшить количество итераций, необходимых для сходимости, при этом количество вычислений градиента не больше чем в градиентном спуске с постоянным шагом, а количество вычислений функции меньше чем в градиентном спуске с методом золотого сечения.

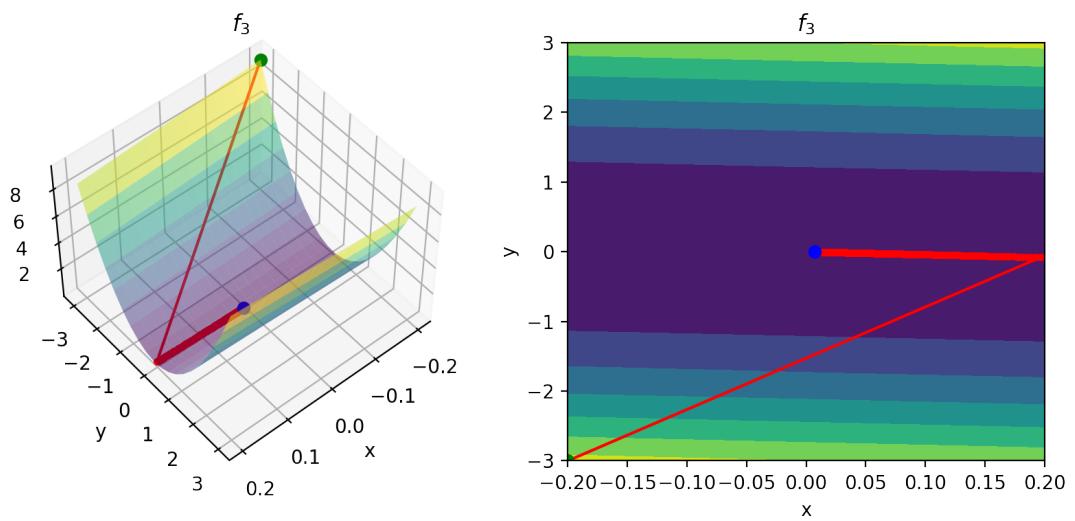
### 3.5 Анализ работы методов в зависимости от выбора начальной точки

Рассмотрим работу методов в зависимости от выбора начальной точки на примере точек  $(-1, -0.2)$ ,  $(-2, -10)$ ,  $(-7.2, -7.2)$

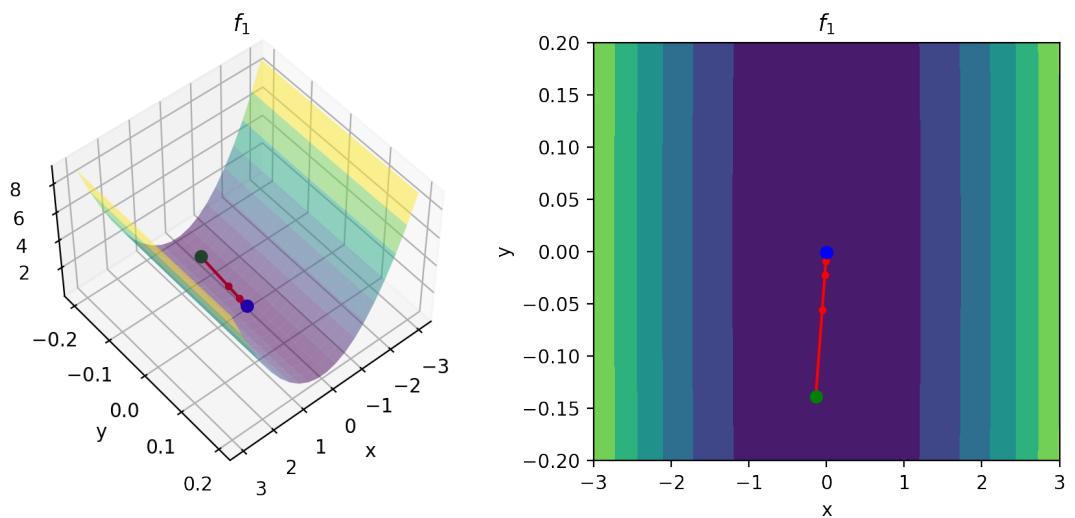
#### 3.5.1 Градиентный спуск с постоянным шагом

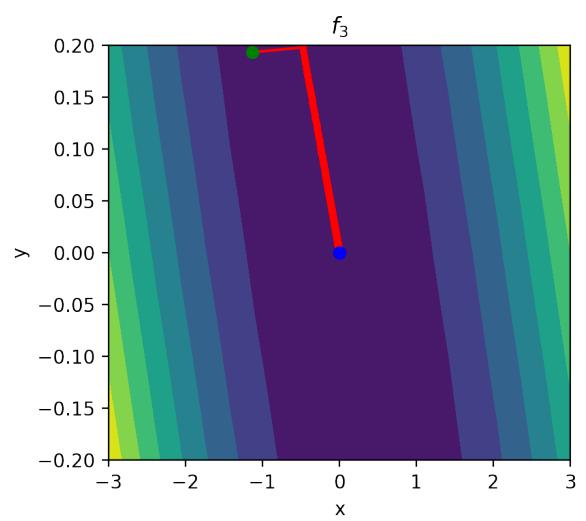
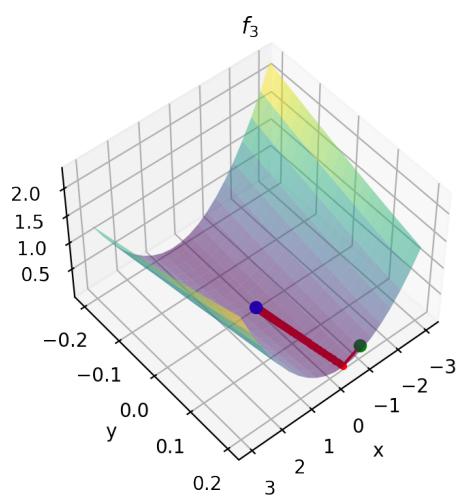
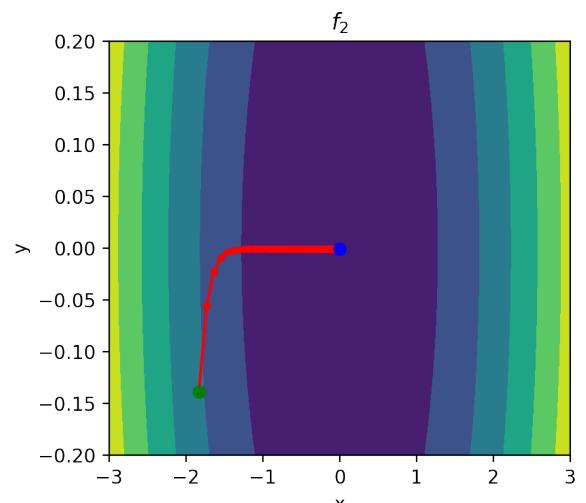
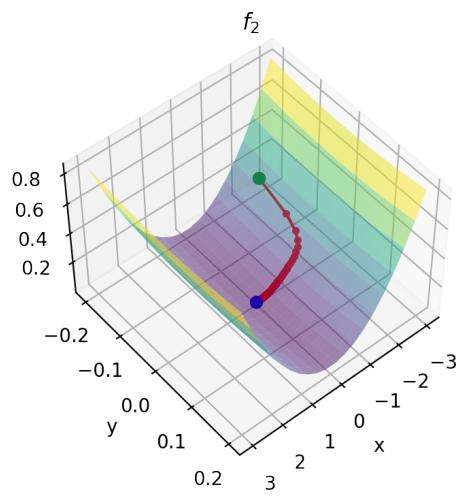
1.  $x_0 = (-0.2 \ -3)$



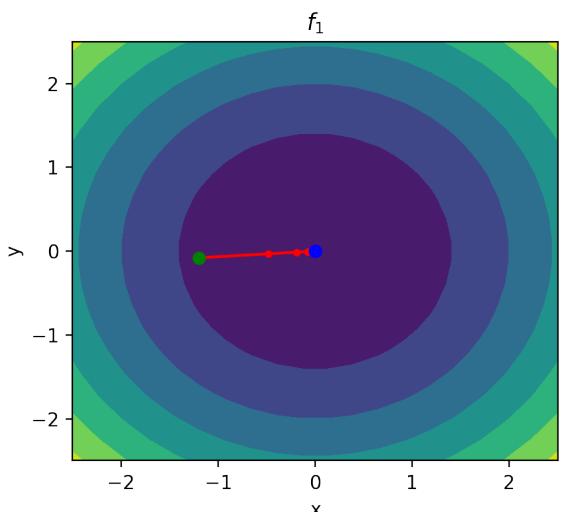
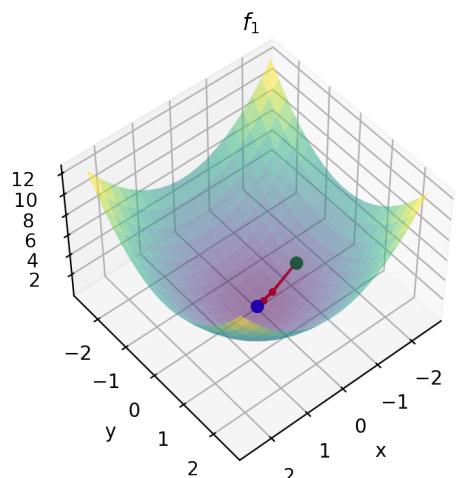


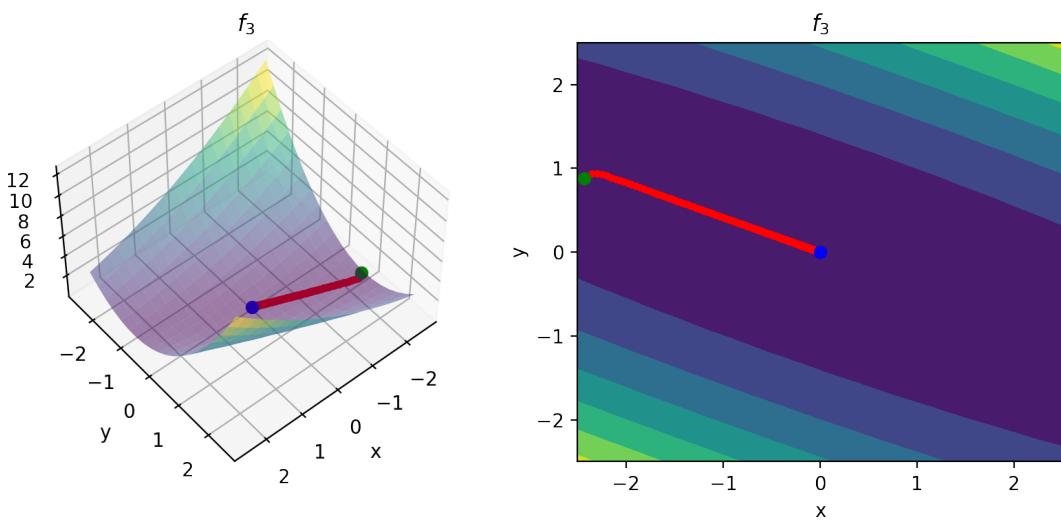
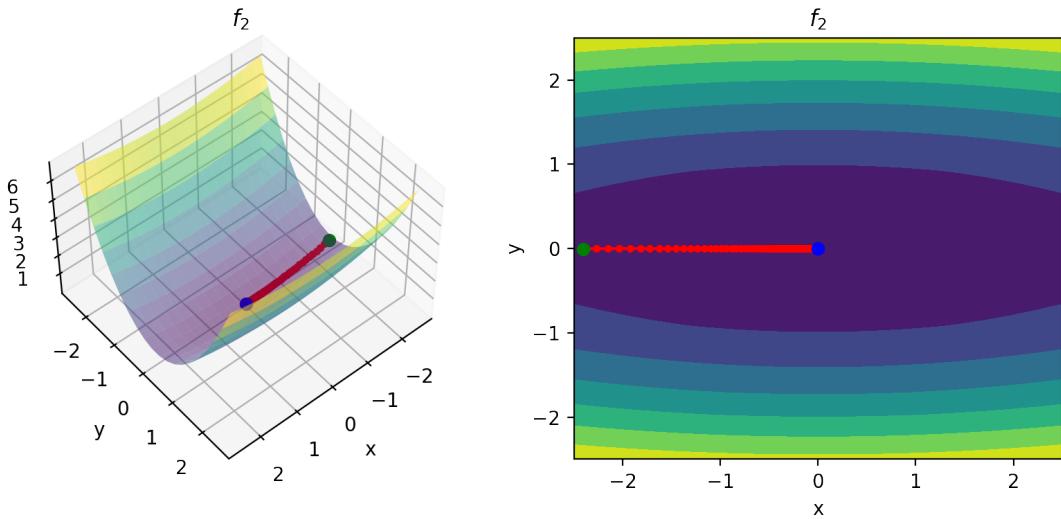
2.  $x_0 = (-2.16 -2.16)$





3.  $x_0 = (-3 -0.2)$





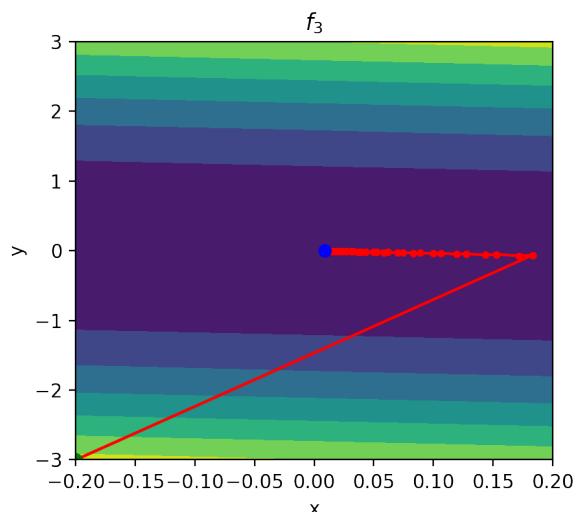
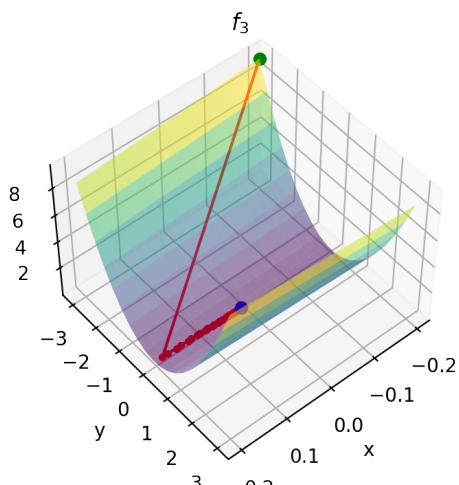
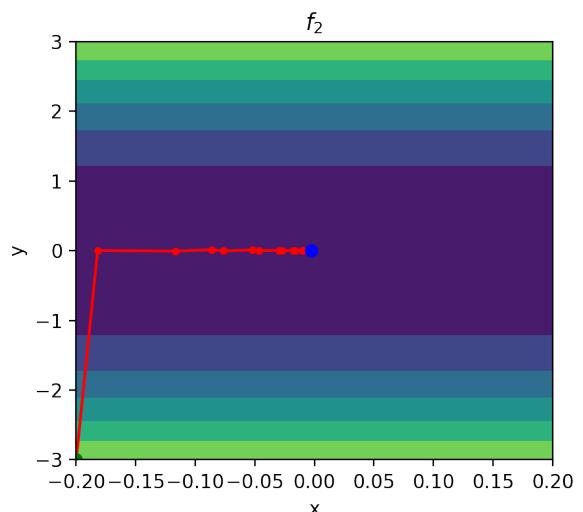
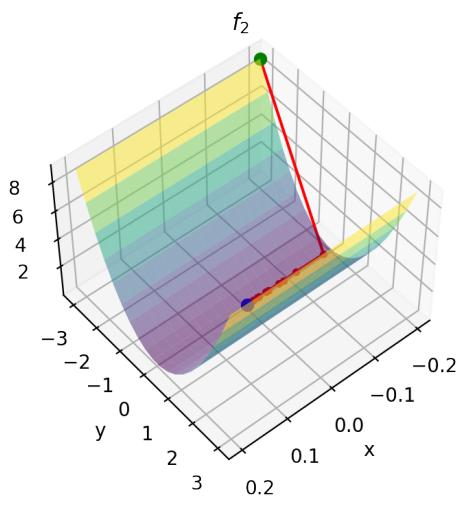
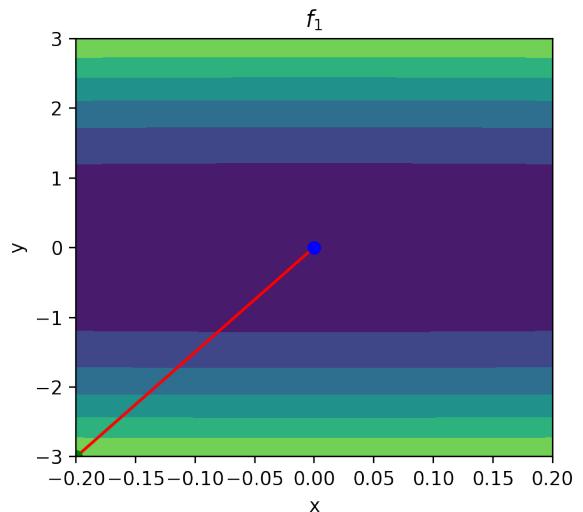
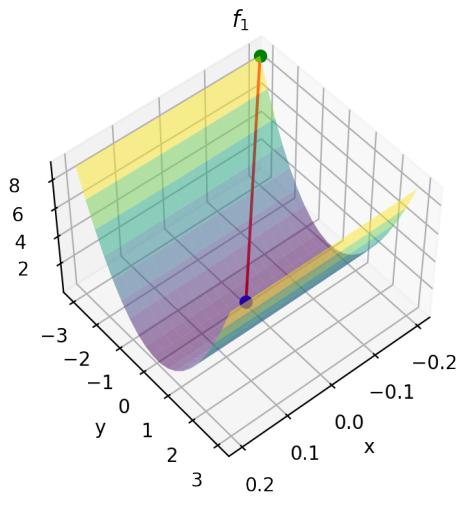
**Количество итераций для каждой функции в зависимости от начальной точки**

Функция	$x_1$	$x_2$	$x_3$
$f_1$	11	11	11
$f_2$	85	128	134
$f_3$	255	264	304

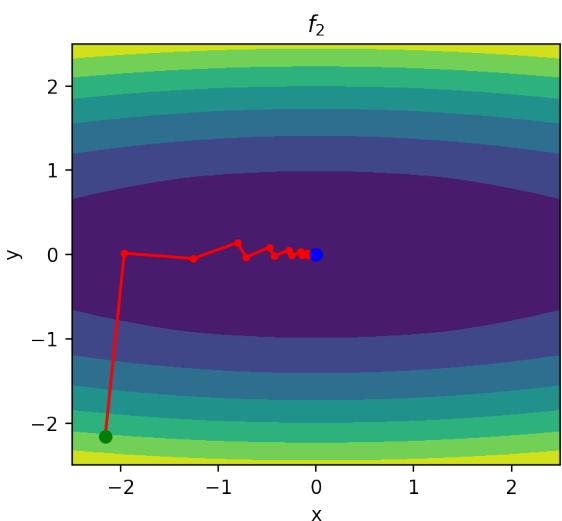
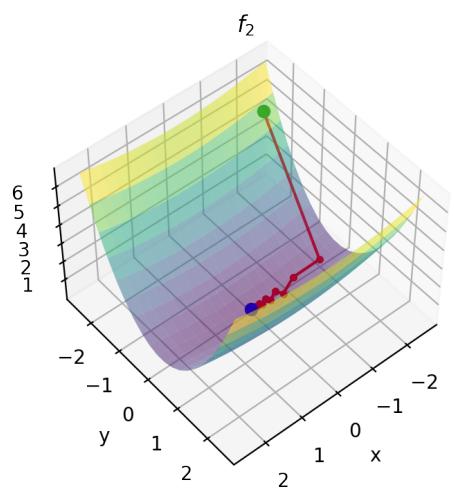
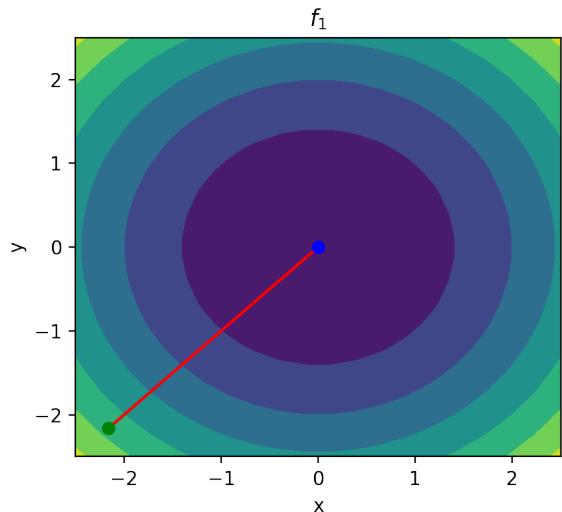
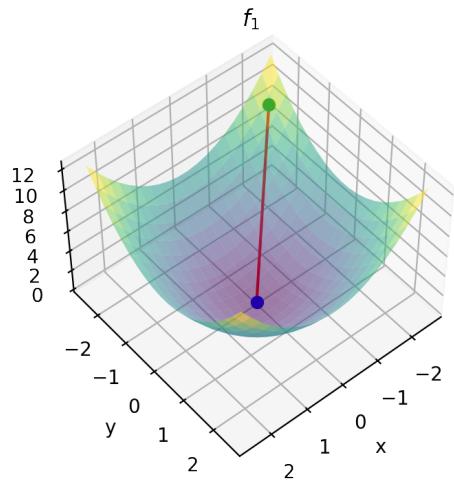
Для функции  $f_1$  количество итераций постоянно для выбранных точек. Для функций  $f_1$  и  $f_2$  количество итераций возрастает при удалении начальной точки от точки минимума  $(0, 0)$  по оси  $x$ . Можно сделать вывод, что количество итераций необходимых для сходимости зависит от выбора начальной точки.

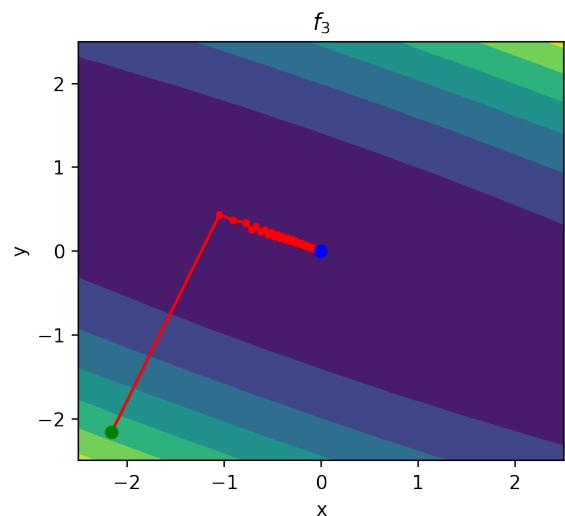
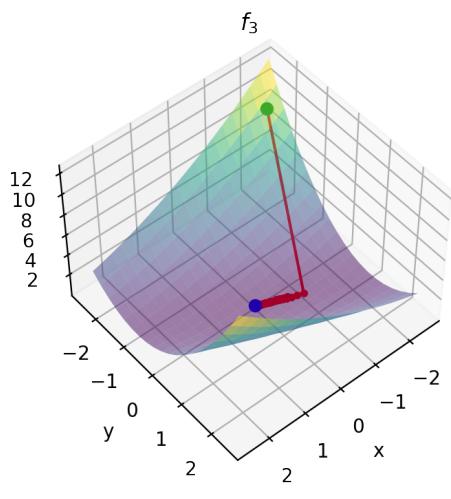
### 3.5.2 Градиентный спуск с методом золотого сечения

$$1. \quad x_0 = (-0.2 \ -3)$$

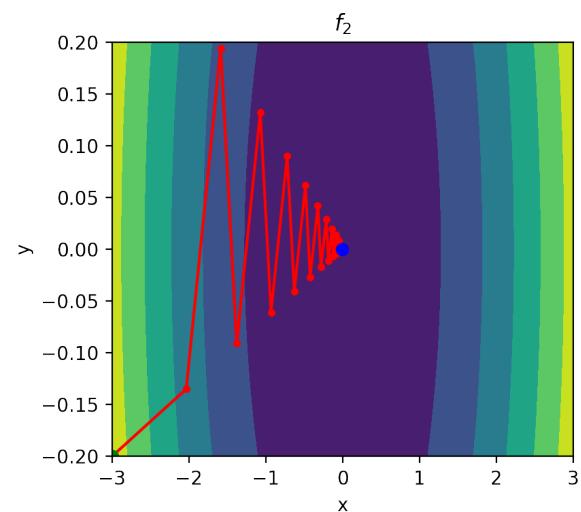
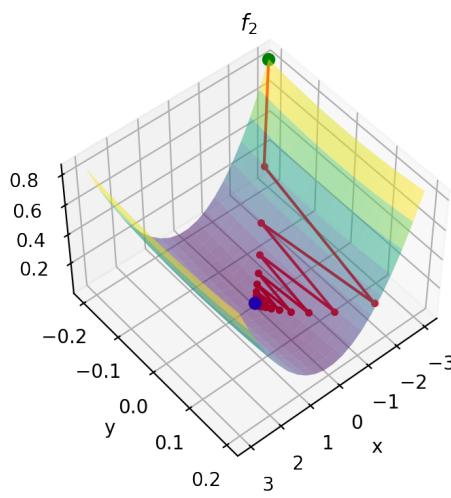
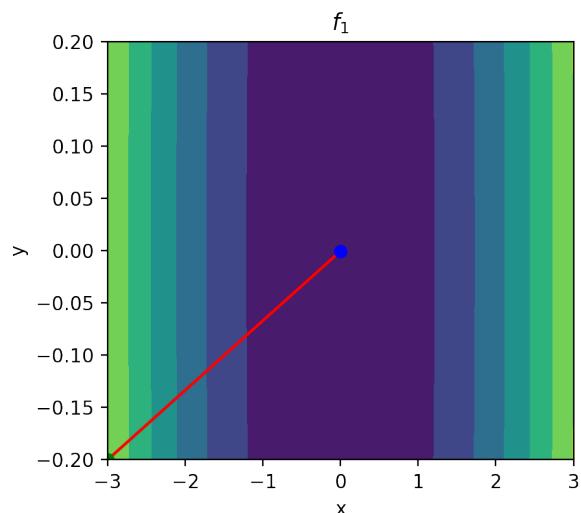
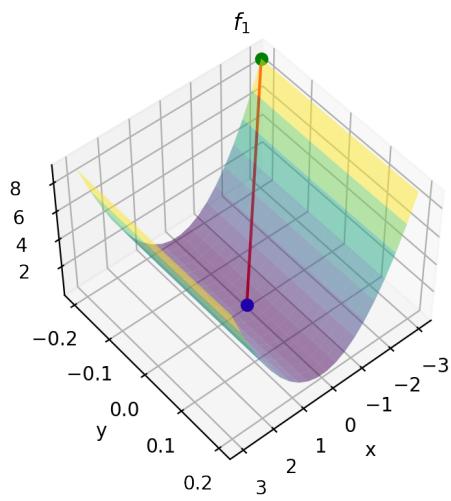


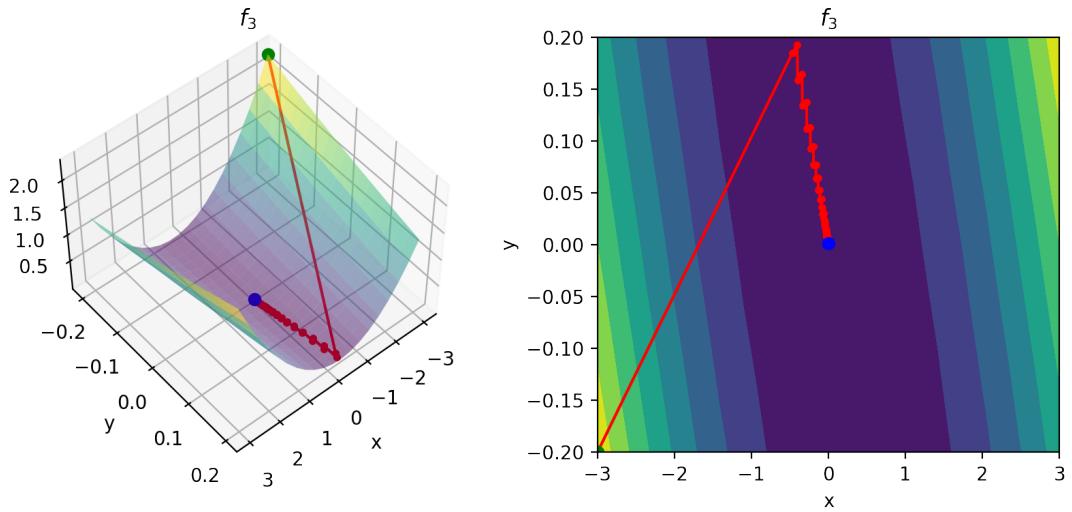
2.  $x_0 = (-2.16 -2.16)$





3.  $x_0 = (-3 -0.2)$





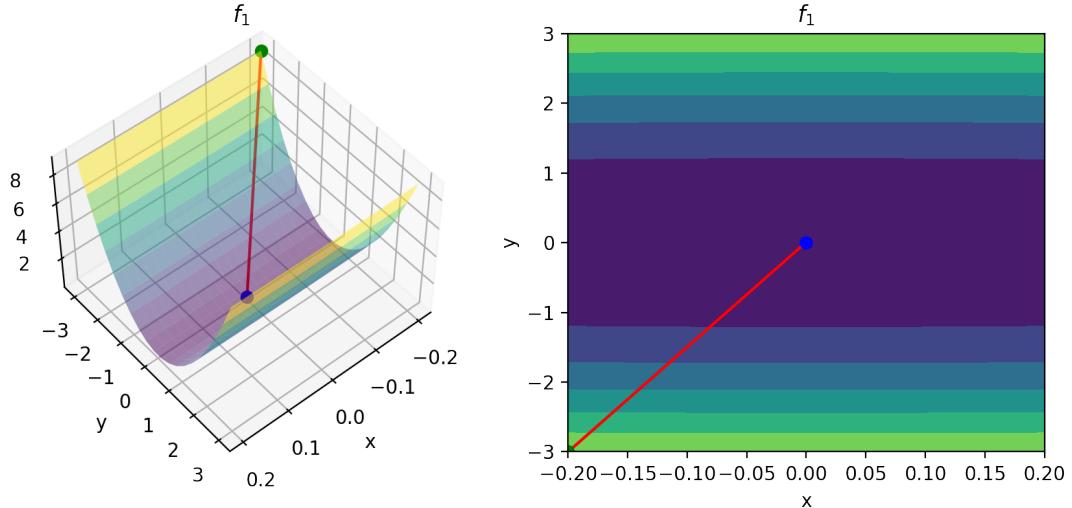
**Количество итераций для каждой функции в зависимости от начальной точки**

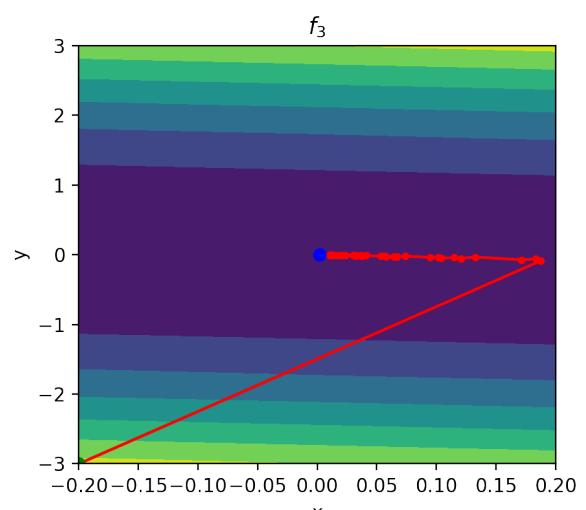
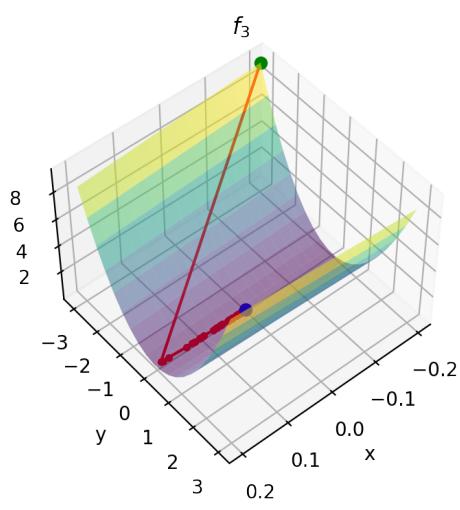
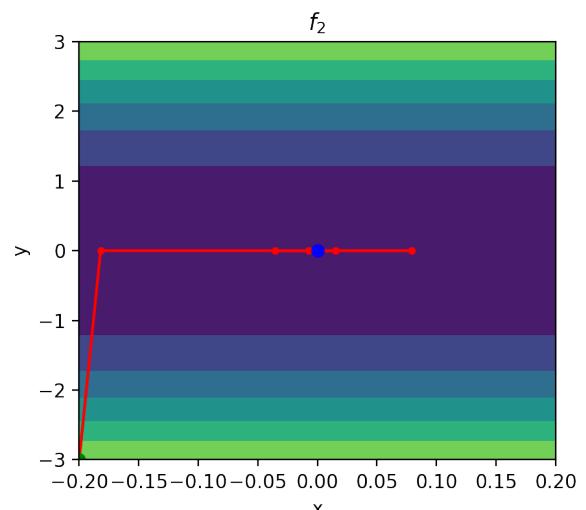
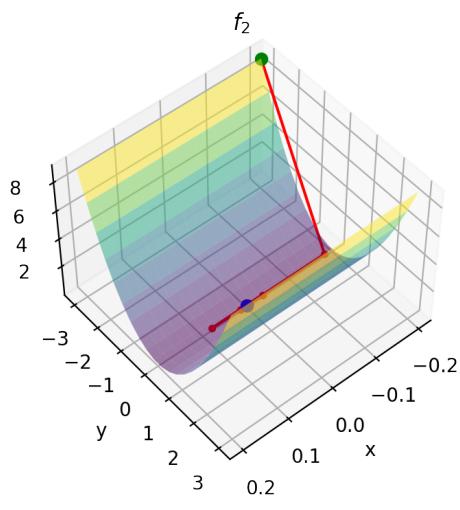
Функция	$x_1$	$x_2$	$x_3$
$f_1$	2	2	2
$f_2$	17	25	31
$f_3$	63	61	73

Для функции  $f_1$  количество итераций постоянно для выбранных точек. Для функций  $f_1$  и  $f_2$  количество итераций возрастает при удалении начальной точки от точки минимума  $(0, 0)$  по оси  $x$ , за исключением функции  $f_3$  для точек  $x_1$  и  $x_2$ , где количество итераций почти одинаковое. Можно сделать вывод, что количество итераций необходимых для сходимости зависит от начальной точки.

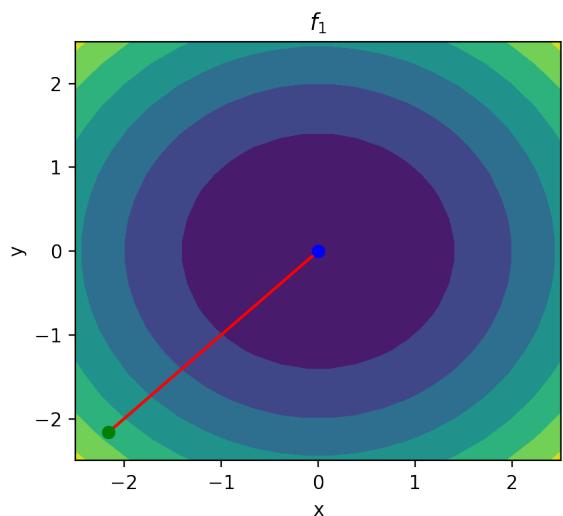
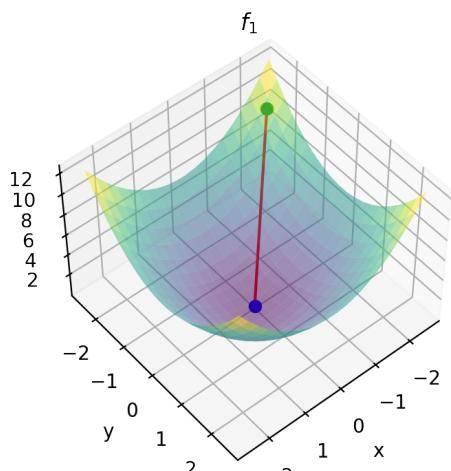
### 3.5.3 Градиентный спуск с условием Вольфе

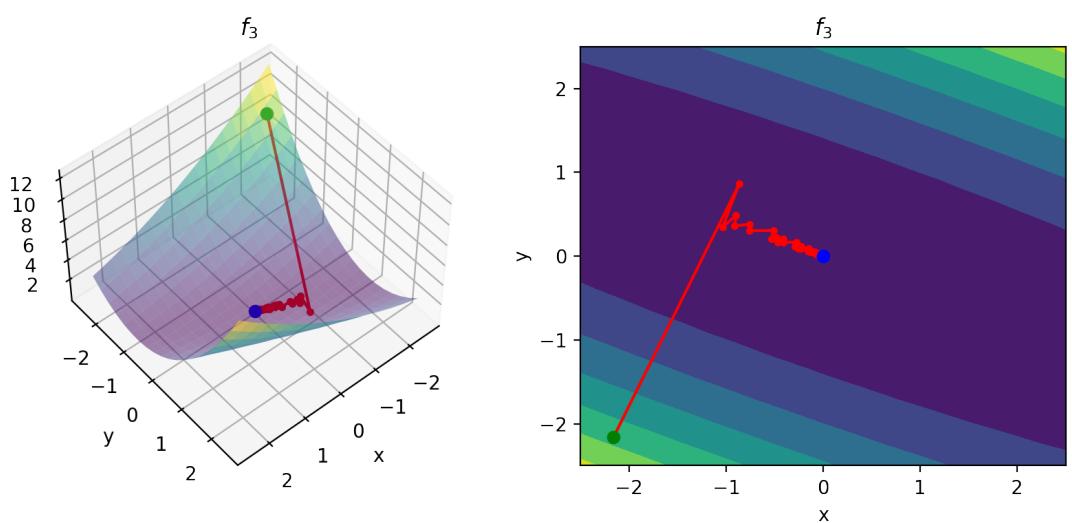
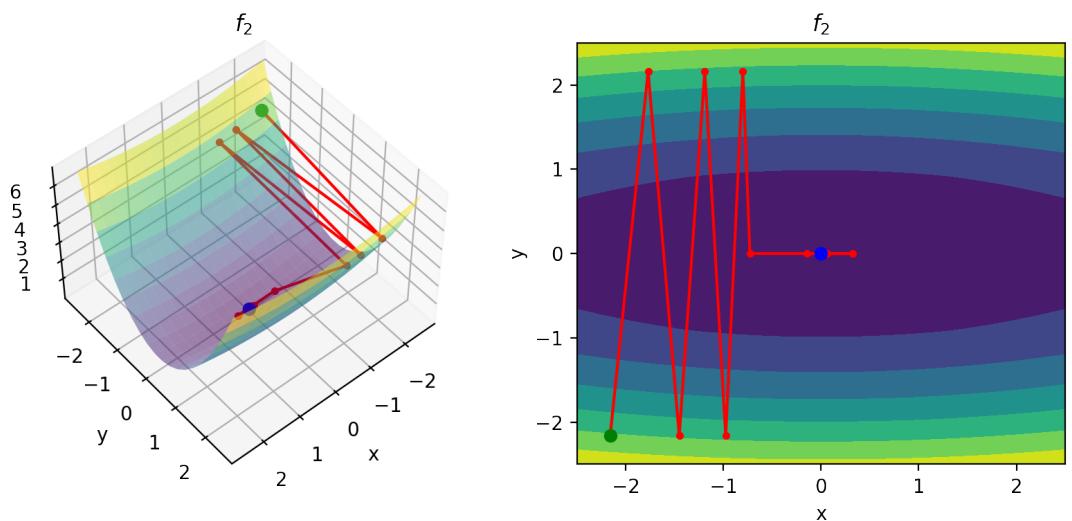
1.  $x_0 = (-0.2 \ -3)$



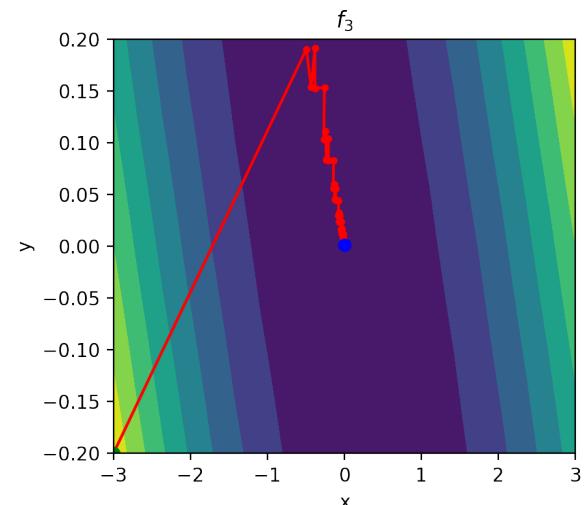
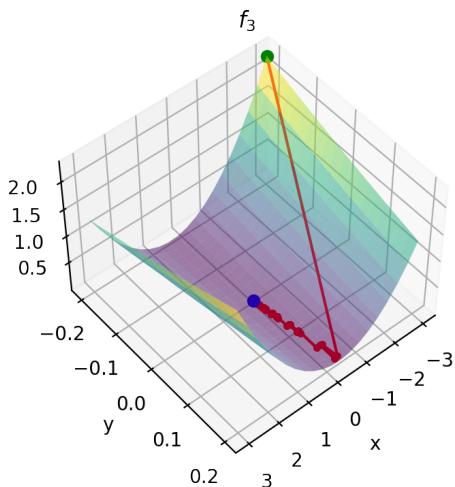
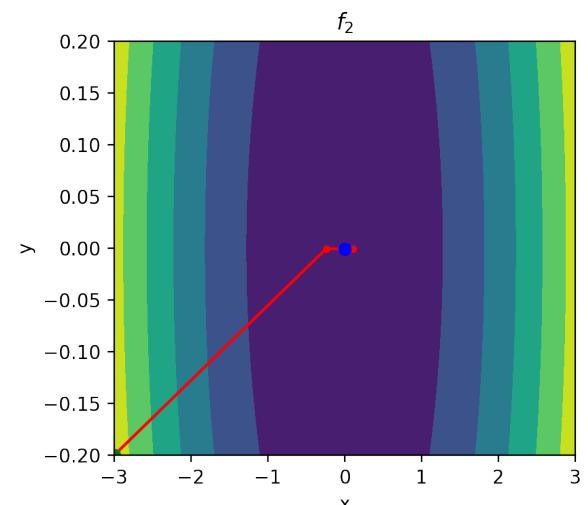
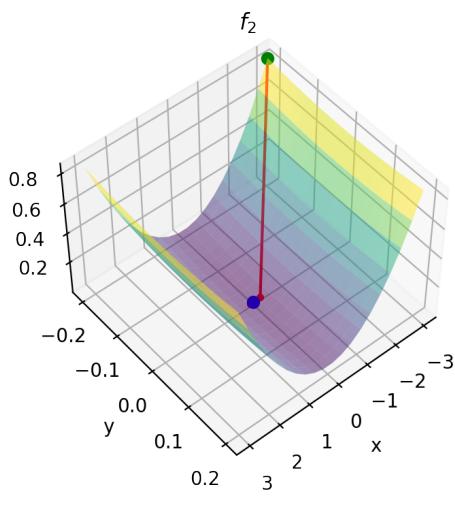
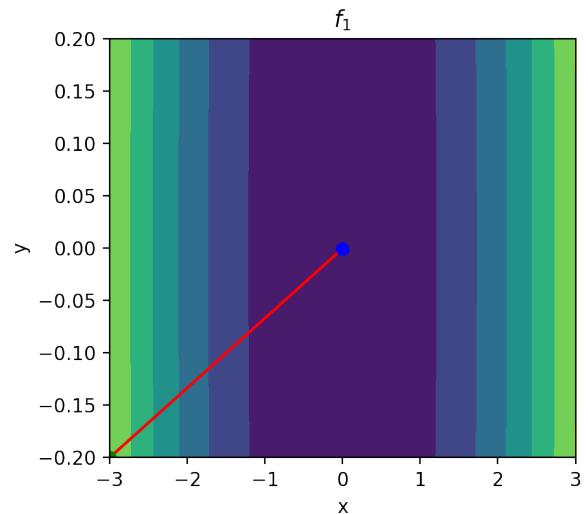
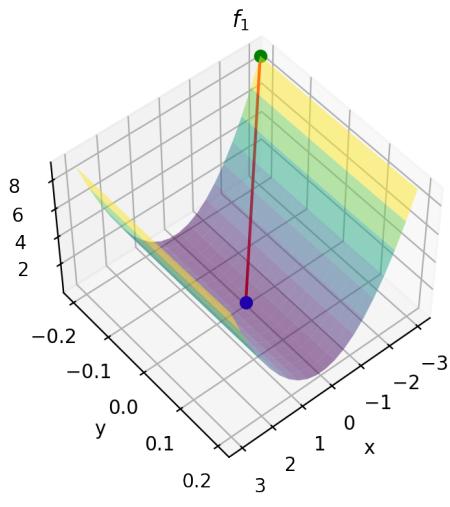


2.  $x_0 = (-2.16 \ -2.16)$





$$3. \quad x_0 = (-3 -0.2)$$



### Количество итераций для каждой функции в зависимости от начальной точки

Функция	$x_1$	$x_2$	$x_3$
$f_1$	1	1	1
$f_2$	8	14	21
$f_3$	52	51	62

Для функции  $f_1$  количество итераций постоянно для выбранных точек. Для функций  $f_1$  и  $f_2$  количество итераций возрастает при удалении начальной точки от точки минимума  $(0, 0)$  по оси  $x$ , за исключением функции  $f_3$  для точек  $x_1$  и  $x_2$ , где количество итераций почти одинаковое. Можно сделать вывод, что количество итераций необходимых для сходимости зависит от начальной точки.

Таким образом, можно сделать вывод что для всех методов количество итераций необходимых для сходимости зависит от выбранной точки для плохо обусловленных функций, при чём количество итераций возрастает с удалением от точки минимума по менее масштабируемой оси.

## 3.6 Анализ влияния масштабирования осей плохо обусловленной функции

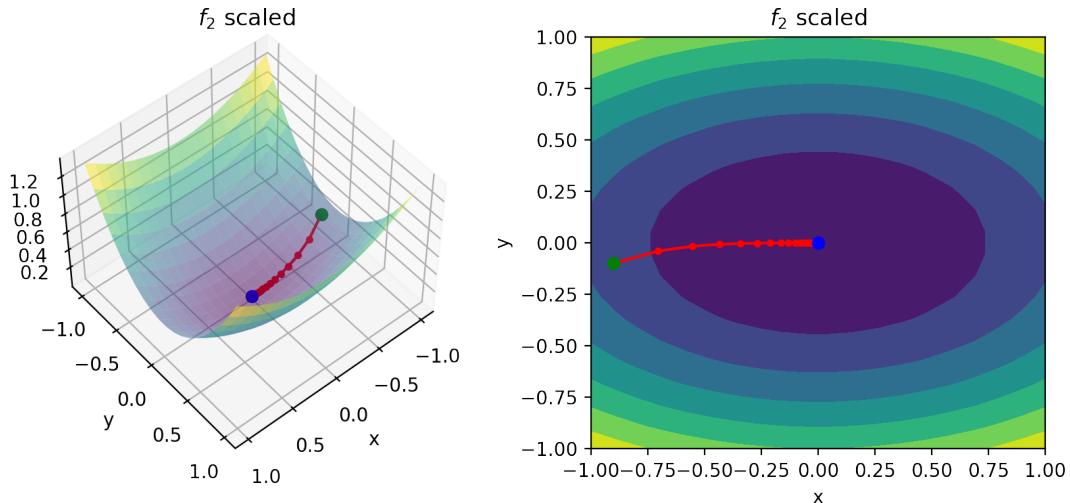
Рассмотрим влияние масштабирования на примере плохо обусловленных функций  $f_2$  и  $f_3$ . Для этого вместо функции  $f(x, y)$  будем рассматривать функцию  $f(\beta x, \gamma y)$ . Рассмотрим, работу методов для разных значений параметра  $\beta = 1, 1.4, 2$  и  $\gamma = 1, 0.2, 0.1$ . Далее будем рассмотрена работа методов с масштабируемыми осями. В каждой таблице, значение параметров  $\beta$  и  $\gamma$  будет указано в первой строке.

### 3.6.1 Градиентный спуск с постоянным шагом

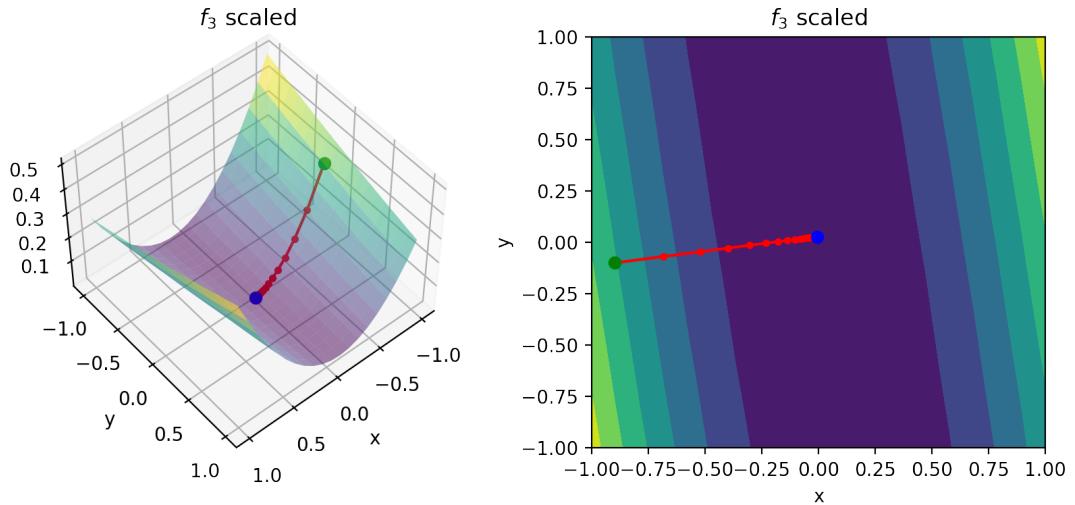
#### Количество итераций до сходимости в зависимости от параметров $\beta$ и $\gamma$

Функция	1, 1	1, 0.2	1, 0.1	1.4, 1	1.4, 0.2	1.4, 0.1	2, 1	2, 0.2	2, 0.1
$f_2$	112	133	297	62	131	297	32	131	297
$f_3$	246	565	57	155	428	29	101	261	13

Траектория на функции  $f_2$  с  $\beta = 2$  и  $\gamma = 1$



Траектория на функции  $f_3$  с  $\beta = 1.4$  и  $\gamma = 0.1$

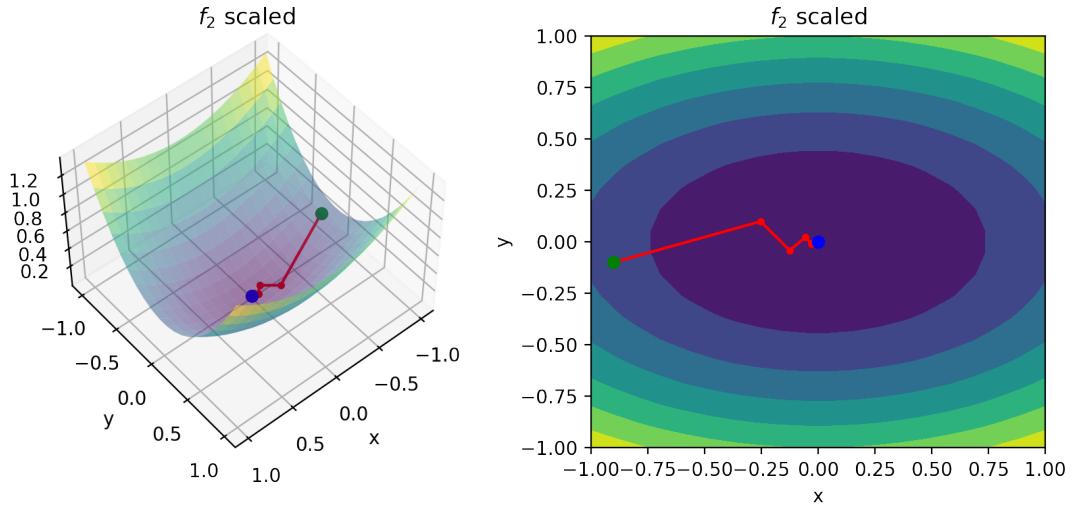


### 3.6.2 Градиентный спуск с методом золотого сечения

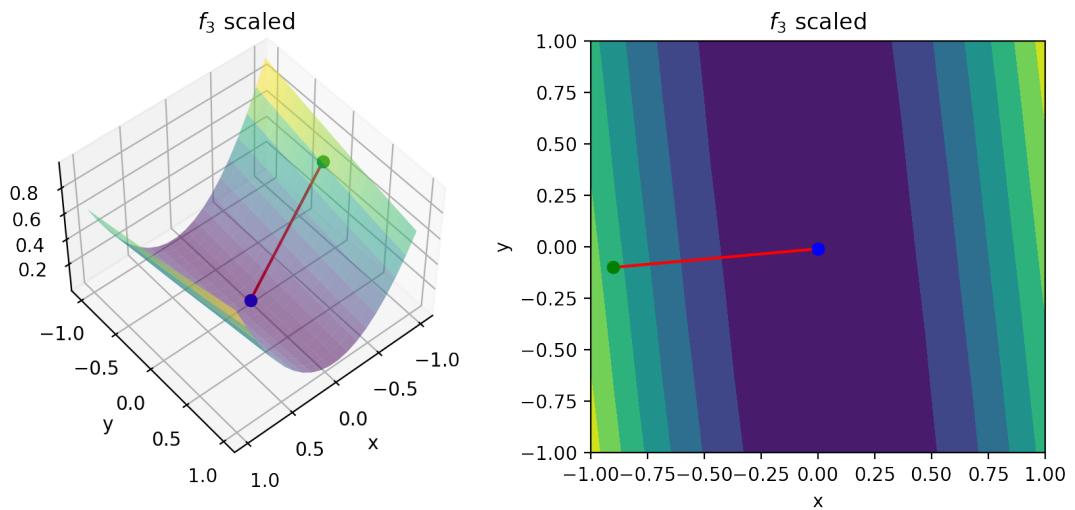
Количество итераций до сходимости в зависимости от параметров  $\beta$  и  $\gamma$

Функция	1, 1	1, 0.2	1, 0.1	1.4, 1	1.4, 0.2	1.4, 0.1	2, 1	2, 0.2	2, 0.1
$f_2$	37	53	149	20	53	149	10	53	149
$f_3$	72	256	266	54	209	6	38	156	1

Траектория на функции  $f_2$  с  $\beta = 2$  и  $\gamma = 1$



Траектория на функции  $f_3$  с  $\beta = 2$  и  $\gamma = 0.1$

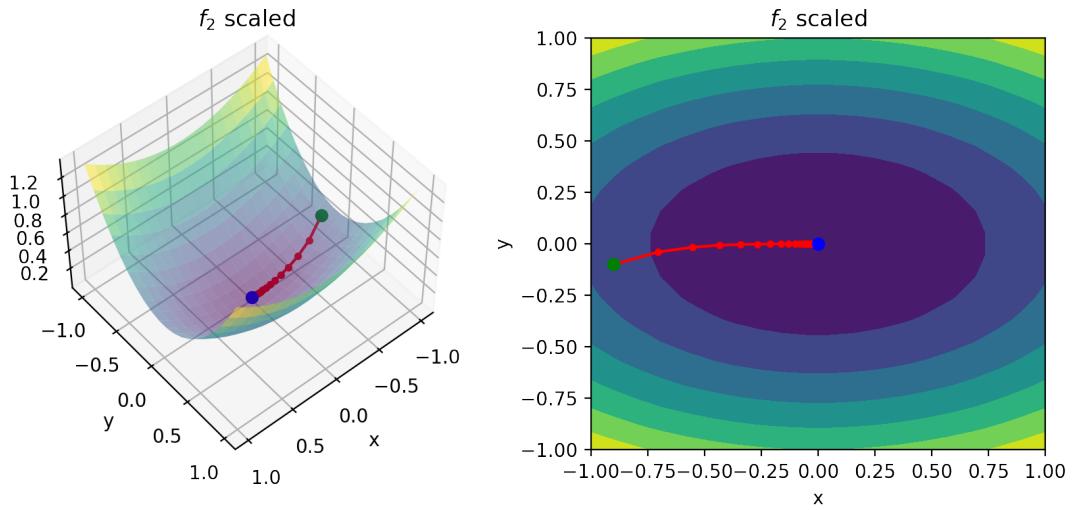


### 3.6.3 Градиентный спуск с условием Вольфе

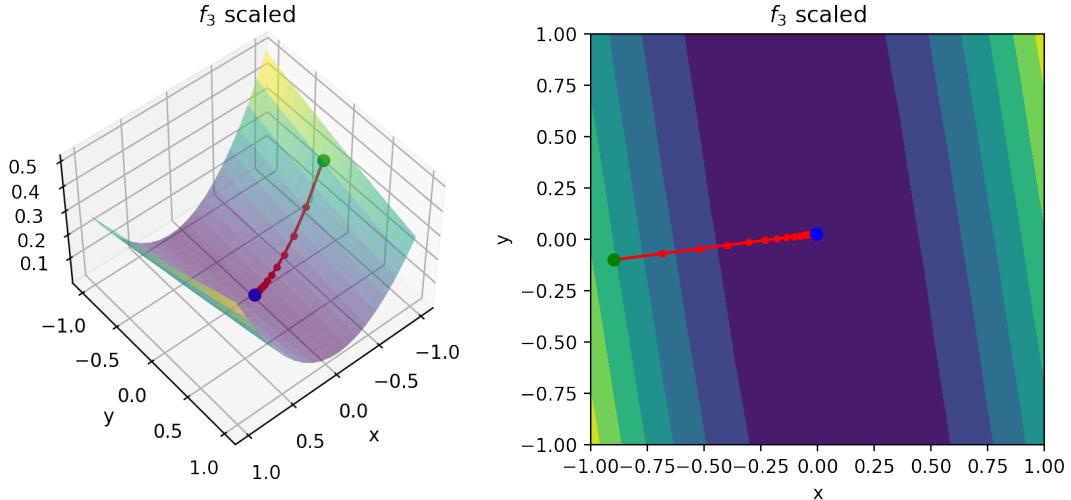
Количество итераций до сходимости в зависимости от параметров  $\beta$  и  $\gamma$

Функция	1, 1	1, 0.2	1, 0.1	1.4, 1	1.4, 0.2	1.4, 0.1	2, 1	2, 0.2	2, 0.1
$f_2$	20	10	18	11	11	26	9	18	35
$f_3$	49	56	46	50	67	15	46	45	15

Траектория на функции  $f_2$  с  $\beta = 2$  и  $\gamma = 1$



Траектория на функции  $f_3$  с  $\beta = 1.4$  и  $\gamma = 0.1$



На основе полученных данных, можно сделать вывод что масштабирование осей, может существенно снизить количество итераций необходимых для сходимости любого из представленных методов. Также, траектория градиентного спуска перестает быть зигзагообразной и становится равномерной при масштабировании осей.

## 4 Исследование сходимости в зависимости от числа обусловленности функции и размерности пространства

Пусть  $T(n, k)$  - количество итераций необходимых градиентному спуску для сходимости, где  $n$  - размерность пространства,  $k$  - число обусловленности.

Исследуем зависимость  $T(n, k)$  от  $n$  и  $k$  для представленных методов, построив соответствующие графики. Для этого проведем множественный эксперимент для различных значений  $n$  и  $k$ , а полученные значения усредним.

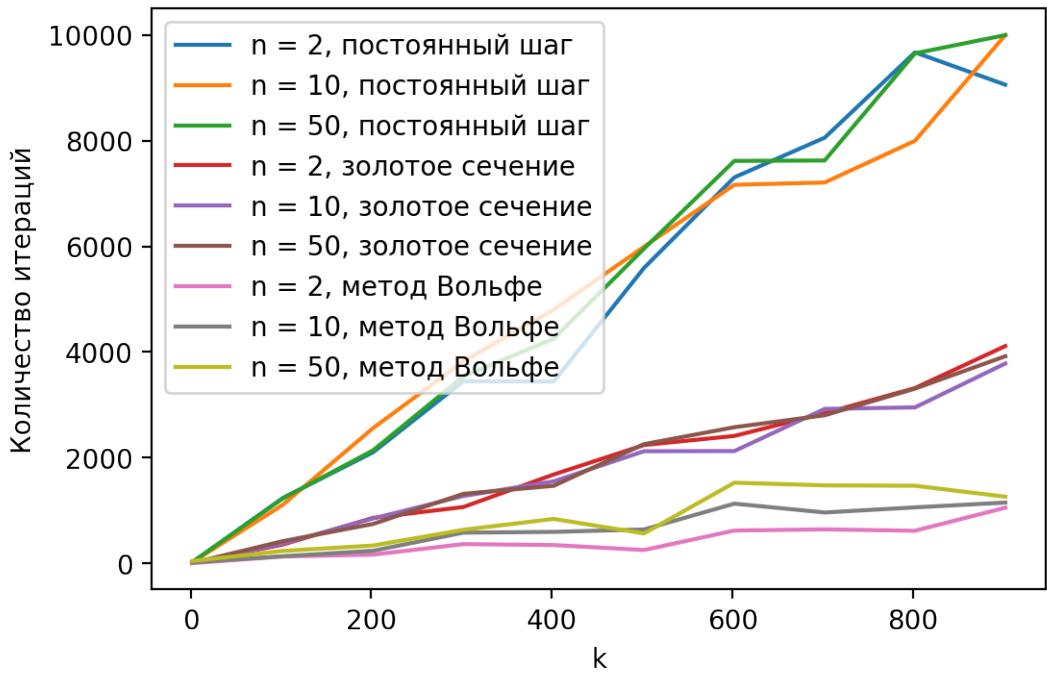
Для проведения эксперимента необходимо сгенерировать функцию с числом обусловленности  $k$  и размерность  $n$ .

Добиться этого можно сгенерировав случайный вектор  $b$  размера  $n$  и случайную диагональную матрицу  $A$  размера  $n$  на  $n$ , где наибольшее значение на диагонали равно 1, а наименьшее  $\frac{1}{k}$ . Функция  $f(x) = Ax^T - bx^T$  будет иметь число обусловленности  $k$ , как отношение наибольшего сингулярного числа к наименьшему матрицы  $A$ .

### 4.0.1 Зависимость $T(n, k)$ от $k$

Построим график зависимости  $T(n, k)$  от  $k$ . Рассмотрим  $k$  равное  $2, 102 \dots 902$ , для каждого значения  $k$  рассмотрим количество итераций необходимых для сходимости при  $n = 1, 2, 50$ .

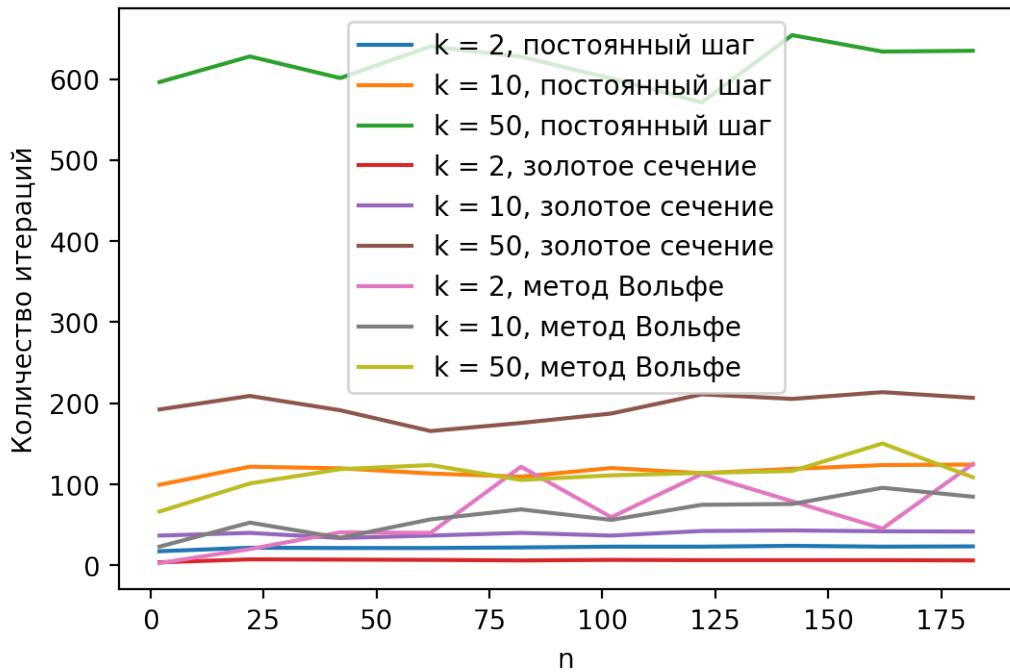
На основе графика, представленного ниже, можно сделать вывод, что количество итераций необходимых для сходимости линейно зависит от числа обусловленности, для всех методов. Также, градиентный спуск с методом Вольфе требует наименьшее количество итераций для сходимости, а градиентный спуск с постоянным шагом наибольшее.



#### 4.0.2 Зависимость $T(n, k)$ от $n$

Построим график зависимости  $T(n, k)$  от  $n$  для значений  $n$  от 2. Построим график зависимости  $T(n, k)$  от  $k$ . Рассмотрим  $n$  равное 2, 22, 42 … 182, для каждого значения  $n$  рассмотрим количество итераций необходимых для сходимости при  $k = 2, 10, 50$ .

На основе графика, представленного ниже, можно сделать вывод, что количество итераций необходимых для сходимости не зависит от размерности функции, для всех методов.



## 5 Выводы

В ходе работы была проанализирована работа методов градиентного спуска с постоянным шагом, с использованием метода золотого сечения и с условием Вольфе.

Сформулируем общие выводы:

1. Сходимость градиентного спуска с постоянным шагом зависит от выбора шага
2. Для улучшения сходимости могут быть применены модификации градиентного спуска: подбор шага методом золотого сечения или подбор шага удовлетворяющему условию Вольфе. При этом могут потребоваться дополнительные вычисления значений функции и градиента функции.
3. Градиентный спуск с условием Вольфе является наиболее оптимальным методом среди рассмотренных, с точки зрения количества итераций для сходимости.
4. Количество итераций необходимых для сходимости линейно растет с увеличением числа обусловленности функции.
5. С помощью масштабирования осей плохо обусловленной функции можно снизить количество итераций необходимых для сходимости.
6. Скорость сходимости зависит от выбора начальной точки.
7. Количество итераций необходимых для сходимости не зависит от размерности функции.