

Методы оптимизации

Отчёт по лабораторной работе №1

Выполнили:

Евсеева Арина М3234

Шульпин Егор М3232

Гаврилюк Виталий М3232

Постановка задачи

1. Градиентный спуск с постоянным шагом (learning rate);
2. Градиентный спуск на основе одномерного поиска (dichotomy);
3. Метод Нелдера-Мида;
4. Анализ эффективности каждого из методов на примере квадратичных функций двух переменных:
 - Исследование сходимости и количества итераций для каждой функции с использованием разных методов;
 - Исследование работы методов в зависимости от выбранной начальной точки;
 - Иллюстрация работы методов с помощью графиков и таблиц;
5. Дополнительная реализация градиентного спуска на основе одномерного поиска (golden ratio);
6. Исследование реализованных методов в сложных случаях:
 - Исследование эффективности методов на функциях n переменных, в зависимости от размерности пространства n ;
 - Исследование эффективности методов на плохо обусловленных функциях двух переменных;
 - Исследование эффективности методов на функциях с зашумленными значениями и на мультимодальных функциях.

Градиентный спуск с постоянным шагом (learning rate)

Описание метода

Градиентный спуск - эвристический алгоритм, который, пользуясь градиентом функции, двигается в направлении скорейшего убывания функции, пошагово рассчитывая новые значения функции. При довольно маленьких изменениях значений (меньше выбранного \mathcal{E}) алгоритм завершает работу в точке найденного минимума функции.

Релизация метода

Алгоритм градиентного спуска:

1. Выбор начальной точки $x = (x_0, \dots, x_n)$ - начальное приближение.
2. На каждой итерации вычисляется новое значение аргумента функции:
$$x_{i+1} = x_i - \alpha \cdot \nabla f(x_i), \text{ где } \alpha - \text{выбранный шаг (learning rate).}$$
3. Условие останова: $\|f(x_{i+1}) - f(x_i)\| < \mathcal{E}$.

Градиентный спуск на основе дихотомии

Описание одномерного поиска (dichotomy)

Метод дихотомии - метод, основанный на делении текущего отрезка $[a, b]$, содержащего экстремум, на две равные части с последующим выбором одной из половин в качестве следующего отрезка. При довольно маленькой длине отрезка (меньше выбранного \mathcal{E}) алгоритм завершает работу.

Реализация одномерного поиска (dichotomy)

Алгоритм дихотомии:

1. Выбор границ отрезка $[a, b]$, вычисление начальных точек x_1 и x_2 :
$$x_1 = \frac{a+b}{2} - \delta, \quad x_2 = \frac{a+b}{2} + \delta.$$
2.
 - $f(x_1) \geq f(x_2)$: смещаем левую границу к x_1 ;
 - $f(x_1) < f(x_2)$: смещаем правую границу к x_2 .
3. Условие останова: $b - a < \mathcal{E}$.

Реализация градиентного спуска при помощи дихотомии

При фиксированном шаге мы сталкиваемся с проблемой выбора шага, который влияет на работоспособность градиентного спуска. Неправильный выбор шага (learning rate) приводит к неэффективности: при слишком маленьком замедляется процесс, при слишком быстром происходят перепрыгивания минимума. Чтобы избежать данной проблемы, можно использовать метод дихотомии для поиска оптимального шага: встаем вдоль определенного направления и подбираем шаг одномерным поиском на отрезке с довольно малыми значениями (в нашем случае $l = \mathcal{E}, r = 0.5$).

Метод Нелдера-Мида

Описание метода

Метод Нелдера-Мида - метод оптимизации функции от нескольких переменных, не использующий градиентов функции. Основной идеей метода является перемещение и деформация симплекса вокруг точки локального экстремума. Данный метод считается довольно надежным и практичным при работе с негладкими функциями, однако у него полностью отсутствует теория сходимости.

Реализация метода

Реализация метода Нелдера-Мида взята из библиотеки `scipy.optimize`, анализ эффективности которой будет рассмотрен далее.

Градиентный спуск на основе золотого сечения

Описание одномерного поиска (golden ratio)

Метод золотого сечения - метод, основанный на делении текущего отрезка $[a, b]$, содержащего экстремум, на две неравные части, для которых будет выполнено правило золотого сечения:

$$\frac{b - x_1}{x_1 - a} = \frac{x_2 - a}{x_1 - a} = \phi = \frac{1 + \sqrt{5}}{2}.$$

При довольно маленькой длине отрезка (меньше выбранного \mathcal{E}) алгоритм завершает работу.

Реализация одномерного поиска (golden ratio)

Алгоритм золотого сечения:

1. Выбор границ отрезка $[a, b]$, вычисление начальных точек x_1 и x_2 :

$$x_1 = b - \frac{b - a}{\phi}, \quad x_2 = a + \frac{b - a}{\phi}.$$

2.
 - $f(x_1) \geq f(x_2)$: смещаем левую границу к x_1 ;
 - $f(x_1) < f(x_2)$: смещаем правую границу к x_2 .
3. Условие останова: $b - a < \mathcal{E}$.

Реализация градиентного спуска на основе золотого сечения

В реализации градиентного спуска на основе дихотомии мы смогли избавиться от фиксированного шага, тем самым убрав проблему с неэффективностью. Однако метод золотого сечения имеет преимущество в более быстрой сходимости. Поэтому заменим поиск шага по направлению с метода дихотомии на метод золотого сечения, что позволит тратить меньше времени на вычисление функции.

Анализ эффективности методов

Выбор исследуемых функций

1) $f = x^2 \cdot y^2 \cdot \ln(4 \cdot x^2 + y^2)$

Данная функция содержит 4 минимума:

- $f(\frac{1}{2\sqrt{2}\sqrt[4]{e}}, \frac{1}{\sqrt{2}\sqrt[4]{e}}) = -\frac{1}{32e}$
- $f(\frac{1}{2\sqrt{2}\sqrt[4]{e}}, -\frac{1}{\sqrt{2}\sqrt[4]{e}}) = -\frac{1}{32e};$
- $f(-\frac{1}{2\sqrt{2}\sqrt[4]{e}}, \frac{1}{\sqrt{2}\sqrt[4]{e}}) = -\frac{1}{32e};$
- $f(-\frac{1}{2\sqrt{2}\sqrt[4]{e}}, -\frac{1}{\sqrt{2}\sqrt[4]{e}}) = -\frac{1}{32e}.$

Она хорошо подходит для проверки сходимости и количества затрачиваемых итераций.

2) $f = -\cos x \cdot \cos y \cdot \exp(-(x - \pi)^2 + (y - \pi)^2)$ (функция Изома)

Данная функция содержит 1 глобальный минимум:

- $f(\pi, \pi) = -1.$

Она подойдет для проверки работоспособности каждого метода при различном выборе начальных точек.

Исследование сходимости и количества итераций

Начальные параметры:

- $f = x^2 \cdot y^2 \cdot \ln(4 \cdot x^2 + y^2);$
- $\mathcal{E} \in \{10^{-3}, 10^{-5}, 10^{-7}\};$
- learning rate = 10^{-2} (при использовании фиксированного шага);
- $l = \mathcal{E}, r = 0.5$ (при использовании метода дихотомии или золотого сечения).

Количество итераций и обращений к минимизируемой функции в зависимости от выбранной точности:

eps	learning rate	golden ratio	dichotomy	nelder mead
1e -3	53 итер. 53 функ. 106 град.	2 итер. 36 функ. 4 град.	2 итер. 48 функ. 4 град.	34 итер. 101 функ. 68 град.
1e -5	192 итер. 192 функ. 384 град.	15 итер. 270 функ. 30 град.	15 итер. 360 функ. 30 град.	50 итер. 147 функ. 100 град.
1e -7	592 итер. 592 функ. 1184 град.	37 итер. 666 функ. 74 град.	37 итер. 888 функ. 74 град.	64 итер. 188 функ. 128 град.

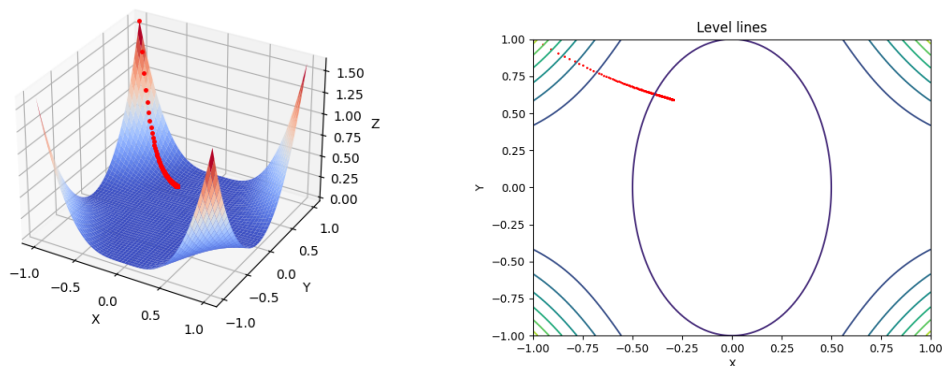


Рис. 1: Градиентный спуск с фиксированным шагом

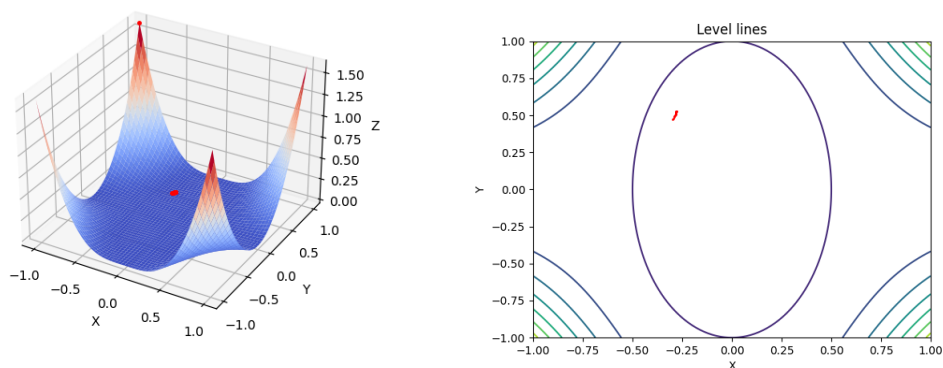


Рис. 2: Градиентный спуск на основе дихотомии

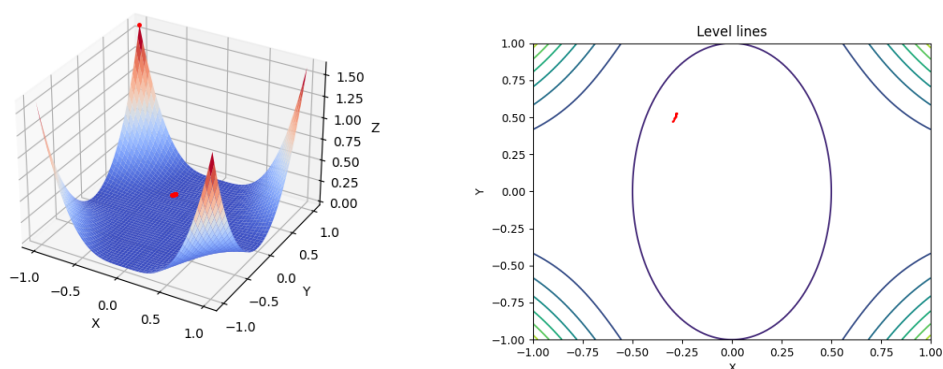


Рис. 3: Градиентный спуск на основе золотого сечения

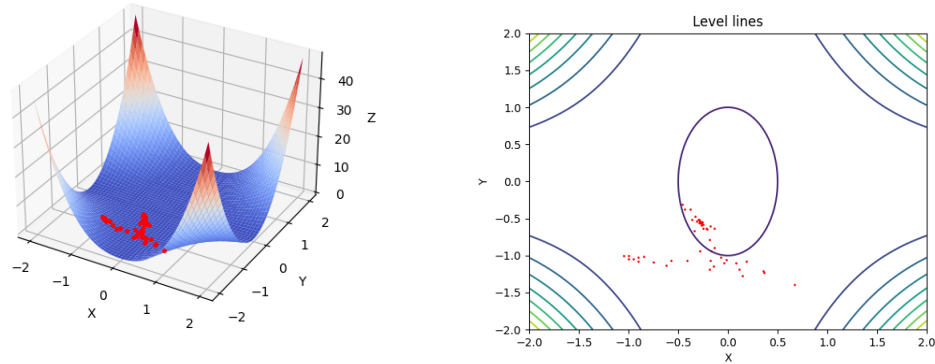


Рис. 4: Метод Нелдера-Мида

Сравнение вычисленных значений:

При запуске из точки $x = (-1, 1)$ ожидаемым значением является $(x, y) \approx (-0.275348, 0.550695)$, $f((x, y)) \approx -0.0114962$.

Значения, вычисленные с помощью разных методов:

eps	learning rate	dichotomy	golden ratio	nelder mead
1e -3	$f(-0.4284;$ $0.647) = 0.0109$	$f(-0.2942;$ $0.4802) = -0.011$	$f(-0.2942;$ $0.4802) = -0.011$	$f(-0.2755;$ $0.5504) = -0.0115$
1e -5	$f(-0.2941;$ $0.5913) = -0.011$	$f(-0.2793;$ $0.5313) = -0.0115$	$f(-0.2793;$ $0.5313) = -0.0115$	$f(-0.2754;$ $0.5507) = -0.0115$
1e -7	$f(-0.2724;$ $0.5664) = -0.0115$	$f(-0.2759;$ $0.5487) = -0.0115$	$f(-0.2759;$ $0.5487) = -0.0115$	$f(-0.2753;$ $0.5507) = -0.0115$

Исследование работоспособности в зависимости от выбранной начальной точки

Начальные параметры:

- $f = -\cos x \cdot \cos y \cdot \exp(-(x - \pi)^2 + (y - \pi)^2)$;
- $\mathcal{E} = 10^{-5}$;
- learning rate = 10^{-2} (при использовании фиксированного шага);
- $l = \mathcal{E}, r = 0.5$ (при использовании метода дихотомии или золотого сечения).

Выбор начальной точки

Функция Изома имеет один глобальный минимум: $f((\pi, \pi)) = -1$. Для каждого метода будем подбирать радиус его работоспособности с помощью бинарного поиска. Левая граница - $(0, 0)$, правая граница - (π, π) , $x = y$, тогда на каждой итерации будем проверять работу метода. В конце остановимся в точке, начиная с которой метод находит верный минимум.

Градиентный спуск с постоянным шагом

С помощью бинарного поиска найденная точка - $(1.89, 1.89)$ (результат округлен до сотых).
Рассчитаем радиус работоспособности: $R = \sqrt{|\pi - 1.89|^2 + |\pi - 1.89|^2} = \sqrt{2 \cdot 1.5665} \approx 1.77$.

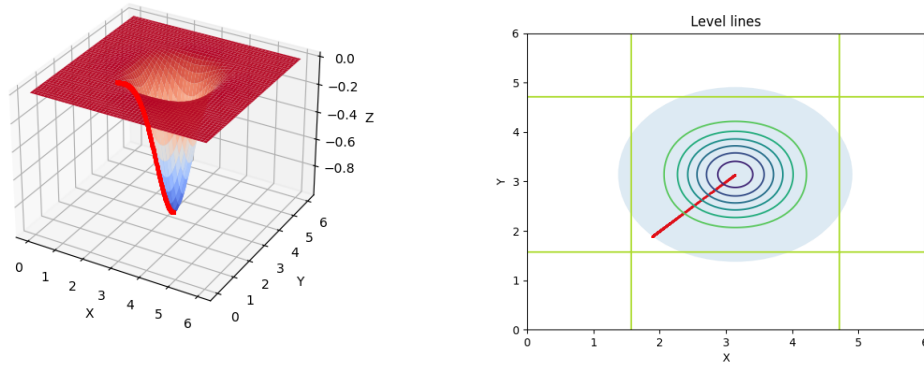


Рис. 5: Радиус работоспособности градиентного спуска с фиксированным шагом

Градиентный спуск на основе дихотомии/золотого сечения

С помощью бинарного поиска найденная точка - $(1.71, 1.71)$ (результат округлен до сотых).
Рассчитаем радиус работоспособности: $R = \sqrt{|\pi - 1.71|^2 + |\pi - 1.71|^2} = \sqrt{2 \cdot 2.0495} \approx 2.0246$.

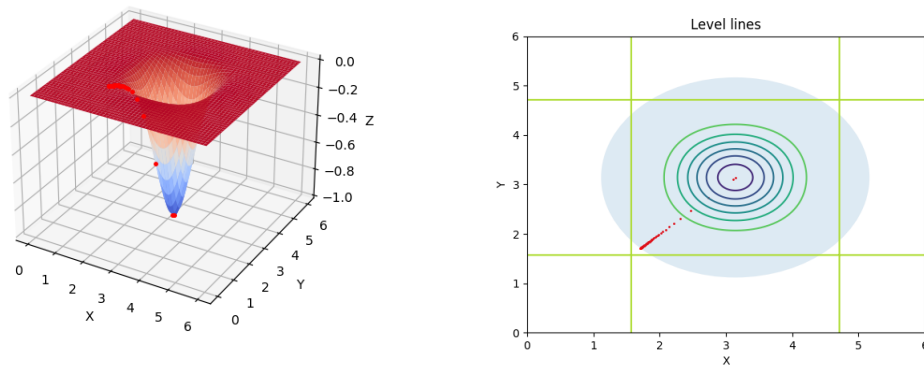


Рис. 6: Радиус работоспособности градиентного спуска на основе дихотомии/золотого сечения

Метод Нелдера-Мида

С помощью бинарного поиска найденная точка - $(1.56, 1.56)$ (результат округлен до сотых).
Рассчитаем радиус работоспособности: $R = \sqrt{|\pi - 1.56|^2 + |\pi - 1.56|^2} = \sqrt{2 \cdot 2.5014} \approx 2.2367$.

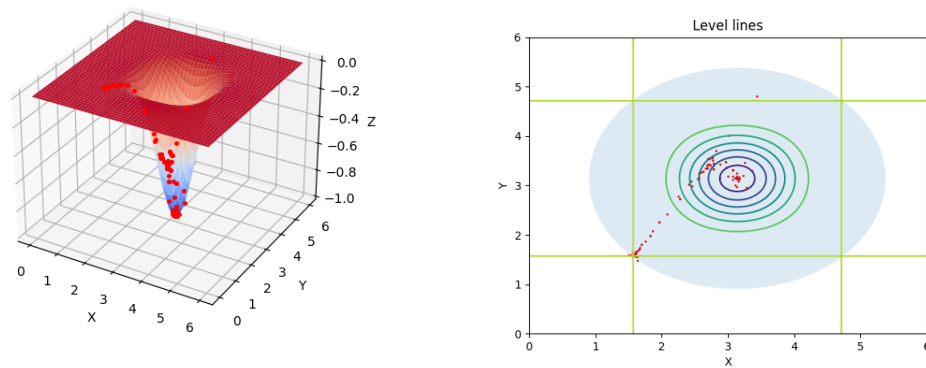


Рис. 7: Радиус работоспособности метода Нелдера-Мида

Полученные результаты

	Количество итераций	Радиус работоспособности
learning rate	767	1.77
dichotomy	63	2.0246
golden ratio	63	2.0246
nelder mead	65	2.2367

Исследование реализованных методов в сложных случаях

Исследование эффективности методов на функциях n переменных

Выбор функций:

- $f = x_1^2 \cdot x_2^2 \cdot \ln(4 \cdot x_1^2 + x_2^2)$ в \mathbb{R}^3 , начальная точка - $(-1, 1)$, ожидаемый минимум $= -\frac{1}{32e}$;
- $f = 7 \cdot x_1^2 + 3 \cdot x_2^2 + 2 \cdot x_3^2$ в \mathbb{R}^4 , начальная точка - $(-1, 5, 2)$, ожидаемый минимум $= 0$;
- $f = 15 \cdot x_1^2 \cdot \ln(x_1^2 + x_3^2) + x_2^2 + x_4^2$ в \mathbb{R}^5 , начальная точка - $(3, 2, 1, 0)$, ожидаемый минимум $= -\frac{15}{e}$;
- $f = 6 \cdot x_1^2 + x_2^2 + 2 \cdot x_3^2 + x_4^2 + 11 \cdot x_5^2$ в \mathbb{R}^6 , начальная точка - $(7, 2, 2, -1, 1)$, ожидаемый минимум $= 0$.

Градиентный спуск с фиксированным шагом

n	Количество итераций	Полученный минимум
3	192 итер. 192 функ. 384 град.	-0.011
4	137 итер. 137 функ. 411 град.	0.0001
5	241 итер. 241 функ. 964 град.	-5.5179
6	246 итер. 246 функ. 1230 град.	0.0002

Градиентный спуск на основе дихотомии/золотого сечения

n	Количество итераций	Полученный минимум
3	15 итер. 270 функ. 30 град.	-0.0115
4	11 итер. 198 функ. 33 град.	0
5	52 итер. 936 функ. 208 град.	-5.5182
6	29 итер. 522 функ. 145 град.	0

Метод Нелдера-Мида

n	Количество итераций	Полученный минимум
3	50 итер. 97 функ. 0 град.	-0.0115
4	104 итер. 190 функ. 0 град.	0
5	254 итер. 436 функ. 0 град.	-5.5182
6	375 итер. 610 функ. 0 град.	0

Полученные результаты:

При нахождении оптимумов в \mathbb{R}^n появляется большая проблема использования градиентного спуска: нахождение всех частных производных. При использовании фиксированного шага, помимо прочих проблем, на каждой итерации вычисляется n производных по каждому направлению, что очень сказывается на производительности. Большого количества вычислений градиентов можно избежать, если динамически находить шаг на каждой итерации, однако метод одномерного поиска тоже занимает большую часть времени при n переменных. Решением будет использование метода Нелдера-Мида, который не использует производные по каждому из направлений и вычисляет функцию на порядок меньше раз (как показано в примере).

Исследование эффективности методов на плохо обусловленных функциях двух переменных

Число обусловленности отвечает за величину изменения значения функции при небольшом изменении ее аргументов и считается как $\mu(x) = \max_x \frac{\varepsilon_y}{\varepsilon_x}$. Тогда плохо обусловленной можно назвать такую функцию, число обусловленности которой $\neq 1$, а ее линии уровня образуют эллипсы с малой полуосью, которая в несколько раз меньше большей. Хорошим примером такой функции является $f = 100 \cdot x^2 + y^2$, так как значение функции увеличивается во много раз больше, чем ее аргумент x .

Начальные параметры:

- $f = 100 \cdot x^2 + y^2$;
- $\mathcal{E} = 10^{-5}$;
- $(x_0, y_0) = (-1, 1)$;
- learning rate $\in \{10^{-2}, 10^{-3}\}$ (при использовании фиксированного шага);
- $l = \mathcal{E}, r = 0.5$ (при использовании метода дихотомии или золотого сечения).

Градиентный спуск с постоянным шагом

При использовании фиксированного шага $= 10^{-2}$ была выявлена проблема с работоспособностью: градиентный спуск двигался в неверном направлении. Это возникает из-за того, что на каждой итерации $x_i = (-1)^{i+1}$, не позволяя градиентному спуску двигаться по оси x .

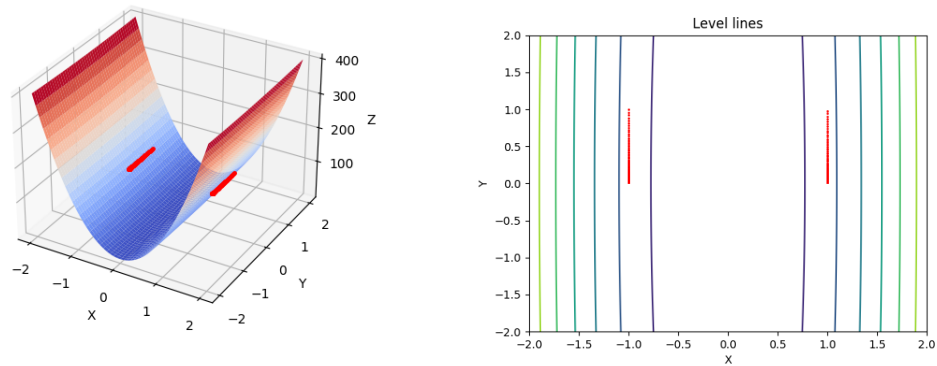


Рис. 8: Градиентный спуск с фиксированным шагом 10^{-2}

Данную проблему решает смена значения шага на 10^{-3} .

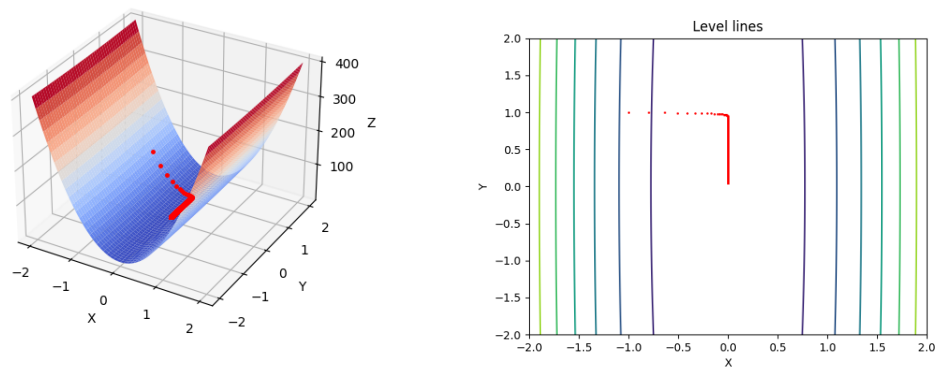


Рис. 9: Градиентный спуск с фиксированным шагом 10^{-3}

Такое решение является совершенно непрактичным, так как невозможно каждый раз предугадывать минимизируемую функцию и необходимый для нее шаг.

Градиентный шаг на основе дихотомии/золотого сечения

Хорошим решением предыдущей проблемы является динамический выбор шага, который можно находить с помощью методов одномерного поиска, например, дихотомии или золотого сечения.

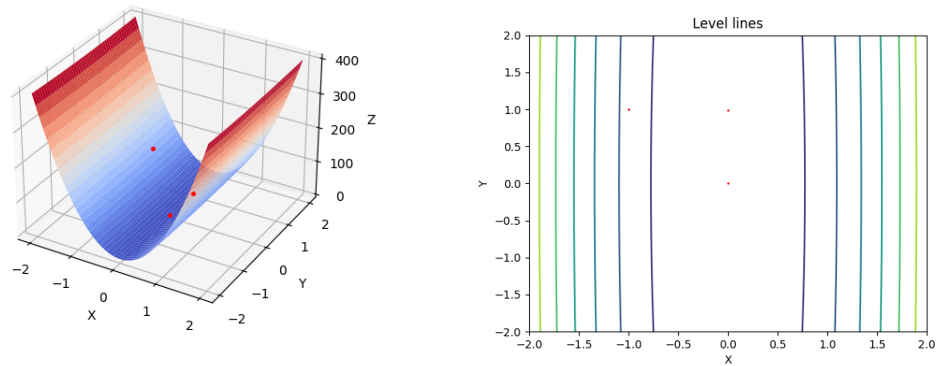


Рис. 10: Градиентный спуск на основе метода дихотомии/золотого сечения

Метод Нелдера-Мида

Еще одним способом избавиться от данной проблемы является выбор другого метода нахождения минимума функции. Метод Нелдера-Мида использует перемещение и деформирование симплекса, точки которого сравниваются по значению функции в них, не давая стоять на месте (в случае с данной функцией по оси x).

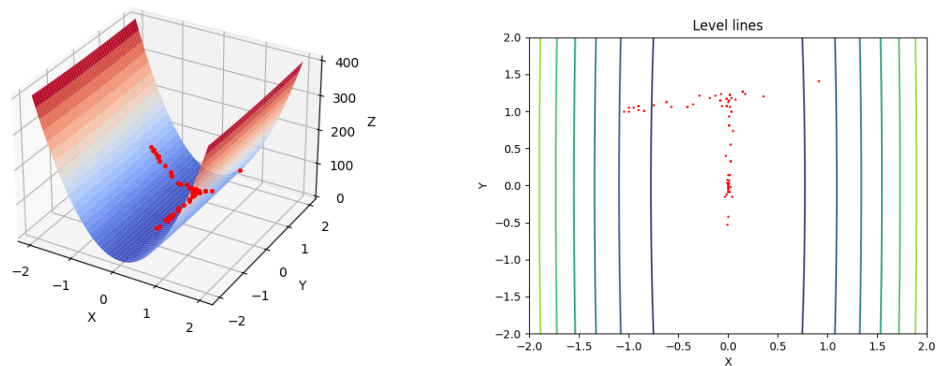


Рис. 11: Метод Нелдера-Мида

Исследование эффективности методов на функциях с зашумленными значениями

$$f = \begin{cases} x^2 + y^2 + \delta, & \text{где } \delta \in [0, 0.15], \text{ если } x \not\approx 0 \text{ и } y \not\approx 0; \\ 0, & \text{иначе.} \end{cases}$$

Градиентный спуск с фиксированным шагом

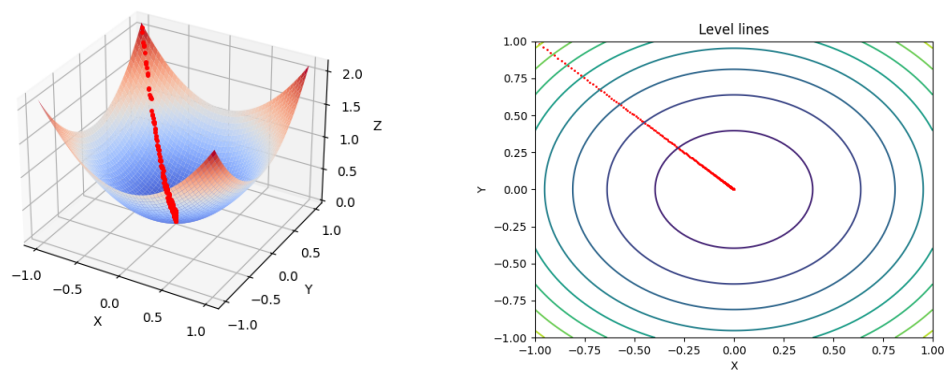


Рис. 12: Градиентный спуск с фиксированным шагом

Градиентный шаг на основе дихотомии/золотого сечения

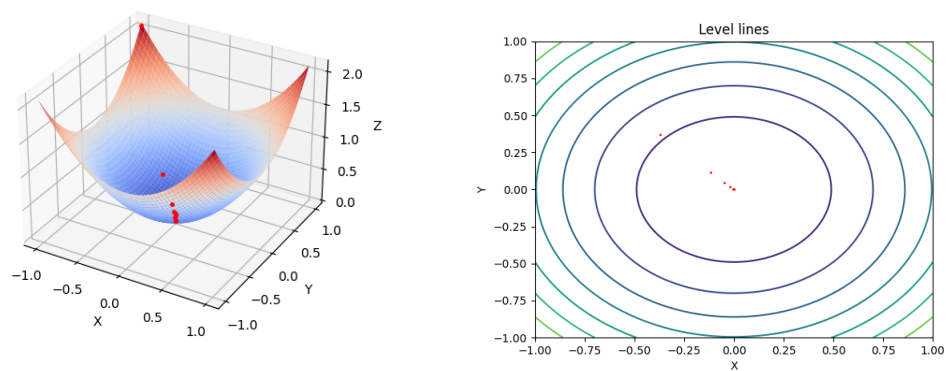


Рис. 13: Градиентный спуск на основе метода дихотомии/золотого сечения

Метод Нелдера-Мида

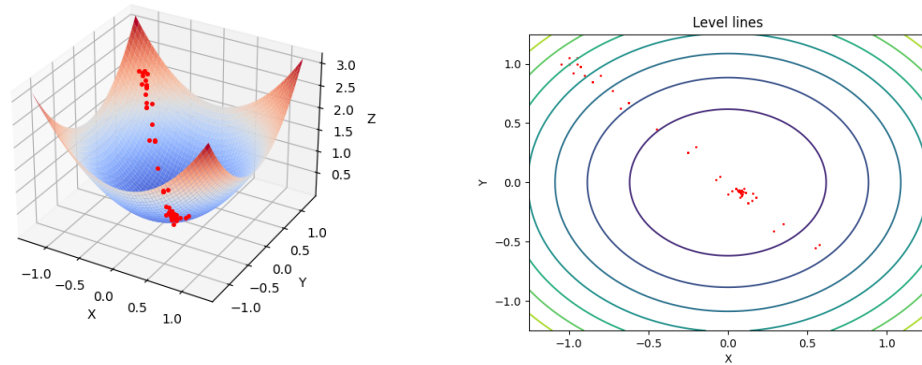


Рис. 14: Метод Нелдера-Мида

Полученные результаты:

Хотя все три метода справились с нахождением оптимума на функции из примера, лучшим методом для работы с функциями с зашумленными значениями является метод Нелдера-Мида, который не использует производные по направлениям.

Исследование эффективности методов на мультимодальных функциях

Мультимодальная функция - это функция, имеющая более чем один оптимум. Такие функции являются невыпуклыми, могут иметь ложные оптимумы или несколько глобальных сразу. Мультимодальной функцией двух переменных является функция Химмельблау $f = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$. Она имеет четыре разнзначных минимума:

- $f(3, 2) = 0$;
- $f(3.584428 \dots, -1.848126 \dots) = 0$;
- $f(-2.805118 \dots, 3.131312 \dots) = 0$;
- $f(-3.779310 \dots, -3.283186 \dots) = 0$.

Начальные параметры:

- $f = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$;
- $\mathcal{E} = 10^{-5}$;
- $(x_0, y_0) \in \{(0, 0), (0, -4)\}$;
- learning rate = 10^{-2} (при использовании фиксированного шага);
- $l = \mathcal{E}, r = 0.5$ (при использовании метода дихотомии или золотого сечения).

Градиентный спуск с фиксированным шагом

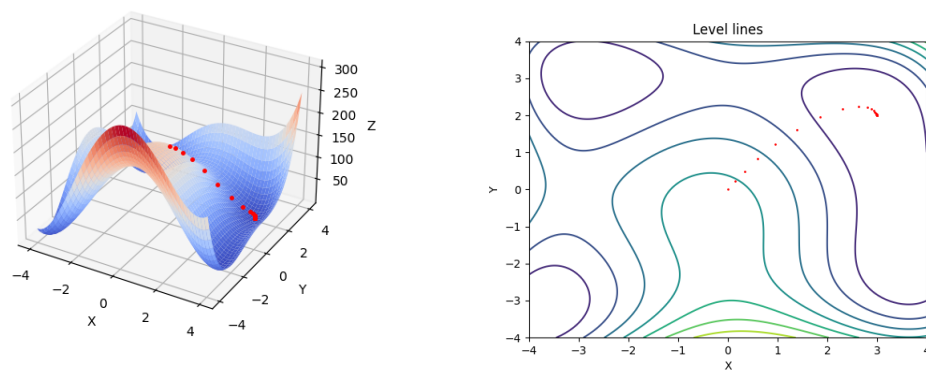


Рис. 15: Градиентный спуск с фиксированным шагом, $(x_0, y_0) = (0, 0)$

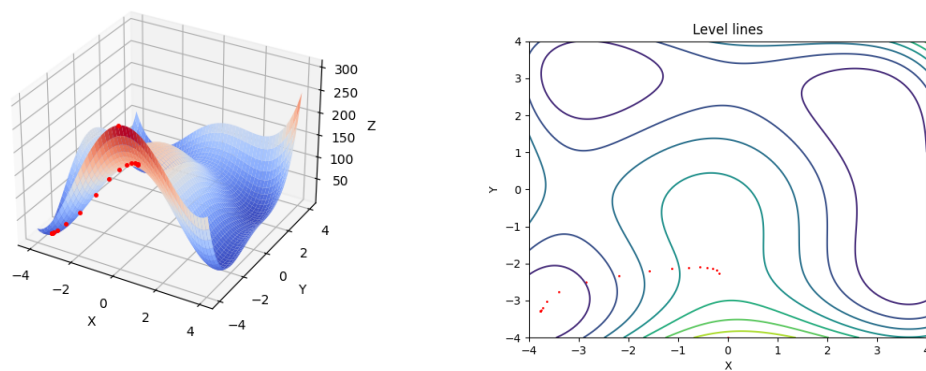


Рис. 16: Градиентный спуск с фиксированным шагом, $(x_0, y_0) = (0, -4)$

Градиентный шаг на основе дихотомии/золотого сечения

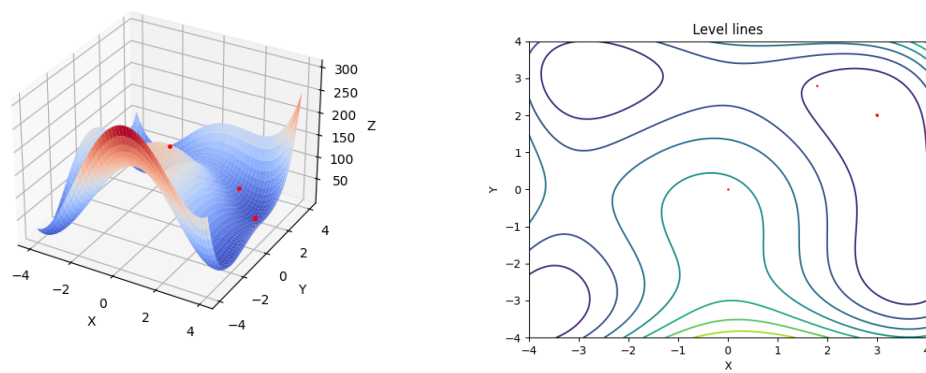


Рис. 17: Градиентный спуск на основе метода дихотомии/золотого сечения, $(x_0, y_0) = (0, 0)$

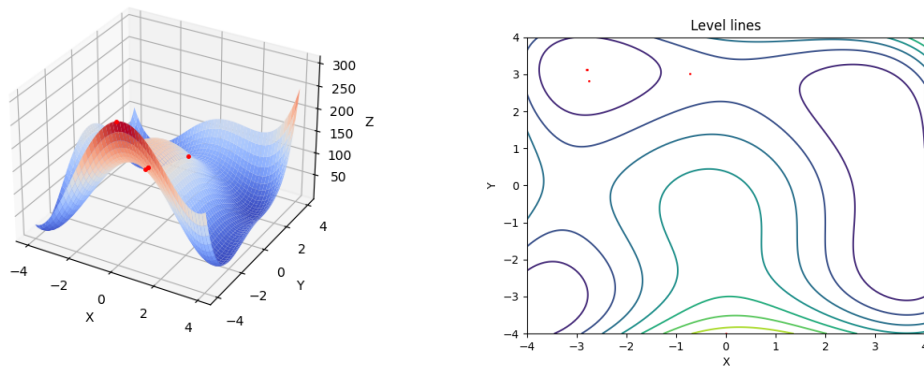


Рис. 18: Градиентный спуск на основе метода дихотомии/золотого сечения, $(x_0, y_0) = (0, -4)$

Метод Нелдера-Мида

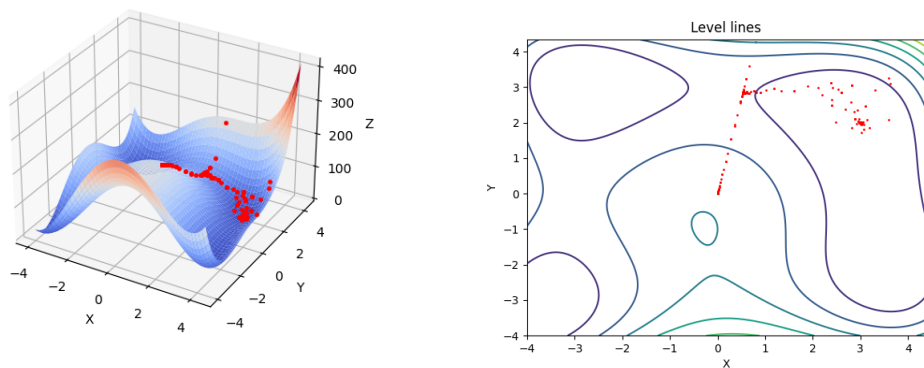


Рис. 19: Метод Нелдера-Мида, $(x_0, y_0) = (0, 0)$

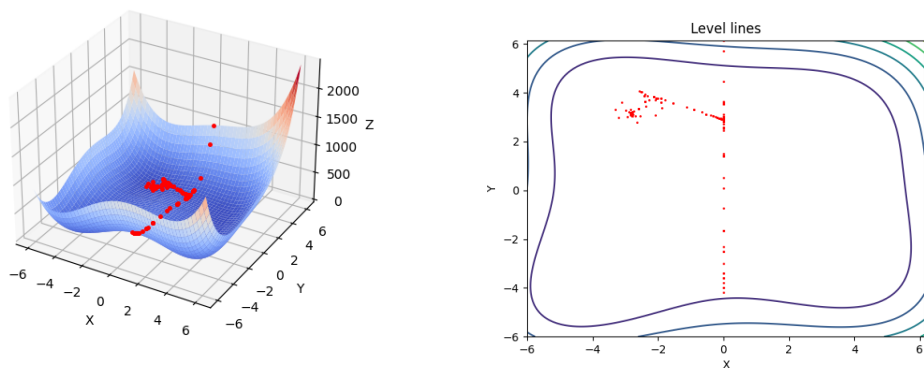


Рис. 20: Метод Нелдера-Мида, $(x_0, y_0) = (0, -4)$

Полученные результаты:

Проблема мультимодальных функций для градиентного спуска является наличие локальных минимумов, в которых он может остановиться, не дойдя до глобального. Однако в примере показана функция с несколькими глобальными минимумами, которые расположены рядом друг с другом. Для градиентного спуска это не является проблемой, так как он движется в конкретном направлении к одному из оптимумов, благодаря производным по направлению. Однако для метода Нелдера-Мида это вызывает трудности, так как симплексы могут попадать в окрестности разных минимумов.