Математика для Data Science. Математический анализ.
Решения задач

Содержание

4.1 **Одномерный градиентный спуск** **1**
Задача 1 . . . 1
Задача 2 . . . 2
Задача 3 . . . 2
Задача 4 . . . 3
Задача 5 . . . 4

4.3 $\mathbb{R}^n$**: расстояния и векторы.** **6**
Задача 1 . . . 6
Задача 2 . . . 6
Задача 3 . . . 8
Задача 4 . . . 8

4.4 **Дифференциал** **9**
Задача 1 . . . 9
Задача 2 . . . 10
Задача 3 . . . 12

4.5 **Частная производная** **13**
Задача 1 . . . 13
Задача 2 . . . 14
Задача 3 . . . 14
Задача 4 . . . 15
Задача 5 . . . 16

4.6 **Направление и градиент** **17**
Задача 1 . . . 17
Задача 2 . . . 17
Задача 3 . . . 17

Замечание. Вот этим цветом отмечены ссылки на страницы внутри этого файла.

4.1 Одномерный градиентный спуск

Задача 1

Дана функция $f(x) = x^2$. Мы начинаем в точке $r_1 = 1024$.

1. В какой точке мы окажемся через 1 шаг градиентного спуска с learning rate $\frac{1}{4}$?
2. В какой точке мы окажемся через 2 шага градиентного спуска с learning rate $\frac{1}{4}$?
3. В какой точке мы окажемся через 13 шагов градиентного спуска с learning rate $\frac{1}{4}$?

Подсказка. Формула следующей точки градиентного спуска такая: $r_{i+1} = r_i - \lambda f'(r_i)$.

Решение. $r_{i+1} = r_i - \lambda f'(r_i)$. При этом так как $f(x) = x^2$, то $f'(x) = 2x$. А значит

$$r_{i+1} = r_i - \frac{1}{4} \cdot 2r_i = \frac{r_i}{2}.$$

То есть каждая следующее число равно половине предыдущего числа. А тогда мы получаем такую цепочку равенств:

$$r_{i+1} = \frac{r_i}{2} = \frac{r_{i-1}}{2^2} = \frac{r_{i-2}}{2^3} = \cdots = \frac{r_1}{2^i} = \frac{2^{10}}{2^i} = 2^{10-i}$$

1. Через 1 шаг градиентного спуска мы окажемся в точке $r_2 = \frac{r_1}{2} = \frac{1024}{2} = 512$.
2. Через 2 шага градиентного спуска мы окажемся в точке $r_3 = \frac{r_1}{4} = \frac{1024}{4} = 256$.
3. Через 13 шагов градиентного спуска мы окажемся в точке $r_{14} = \frac{1024}{2^{13}} = \frac{2^{10}}{2^{13}} = \frac{1}{2^3} = \frac{1}{8} = 0.125$.

Задача 2

В алгоритме градиентного спуска мы надеемся, что для всех i будет выполнено $f(r_i) > f(r_{i+1})$. То есть, что на каждом шаге значение функции уменьшается.

Обязательно ли это будет так? Другими словами, может ли так случиться, что мы делаем шаг градиентного спуска и попадаем в точку, в которой значение функции больше, чем в предыдущей?

Если может – объясните, почему, и приведите пример. Если не может – докажите, что не может.

Подсказка. Попробуйте рассмотреть простые функции, которые мы уже изучили, и разные начальные точки и learning rate.

Решение. Приведём пример, как может случиться, что мы делаем шаг градиентного спуска и попадаем в точку, в которой значение функции больше, чем в предыдущей.

Пусть наша функция $f(x) = x^2$, и мы начинаем градиентный спуск в точке $r_1 = 1$ с learning rate равным 2. Посчитаем производную нашей функции: $f'(x) = (x^2)' = 2x$. Тогда $r_2 = r_1 - 2f'(r_1) = r_1 - 2 \cdot 2r_1 = r_1 - 4r_1 = -3r_1 = -3$. Посчитаем значение функции в полученной нами точке: $f(r_2) = f(-3) = (-3)^2 = 9$. Итого мы получили $f(r_2) > f(r_1)$.

Задача 3

Пусть про функцию $f : \mathbb{R} \to \mathbb{R}$ и точку r_1 известно, что

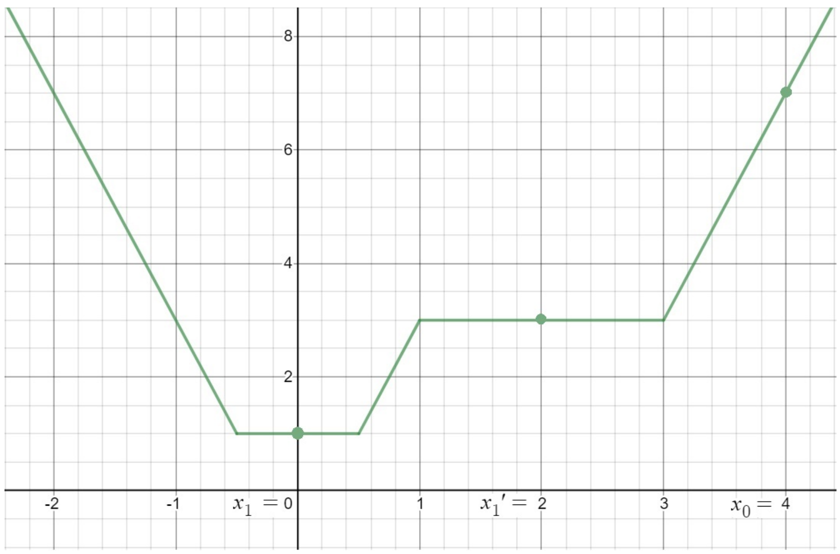
- градиентный спуск с learning rate $= 1$ и стартовой точкой r_1 достигает точки глобального минимума за 1 шаг, и
- градиентный спуск с learning rate $= \frac{1}{2}$ и стартовой точкой r_1 не достигает 1-окрестности точки глобального минимума ни на каком шаге (пояснение: 1-окрестность это ε-окрестность с $\varepsilon = 1$:))

Как может выглядеть график такой функции и где должна располагаться на нём точка r_1?

Подсказка. Можно попробовать придумать пример так называемой кусочно-линейной функции (её график представляет собой ломаную линию).

Решение.

В качестве примера подойдёт функция со следующим графиком:



В виде формулы наша функция запишется следующим образом:

$$f(x) = \begin{cases} -4x - 1 & \text{при } x \le -\frac{1}{2} \\ 1 & \text{при } -\frac{1}{2} < x \le \frac{1}{2} \\ 4x - 1 & \text{при } \frac{1}{2} < x \le 1 \\ 3 & \text{при } 1 < x \le 3 \\ 4x - 9 & \text{при } x > 3 \end{cases}$$

У такой функции бесконечное количество точек глобального минимума: ими являются все точки из отрезка $[-\frac{1}{2}, \frac{1}{2}]$. Действительно, рассмотрим любую точку из этого отрезка: $r_{min} \in [-\frac{1}{2}, \frac{1}{2}]$, тогда для любого x выполнено $f(x) \ge f(r_{min}) = 1$.

В качестве точки r_1 возьмём $r_1 = 4$. Тогда $f'(r_1) = 4$ (действительно, при $x > 3$ производная функции равна $f'(x) = (4x - 9)' = 4$).

Тогда в случае learning rate $= 1$ за один шаг градиентного спуска мы попадаем в точку $r_2 = r_1 - 1 \cdot f'(r_1) = 4 - 1 \cdot 4 = 0$, которая, как мы уже выяснили ранее, является точкой глобального минимума.

В случае learning rate $= \frac{1}{2}$ за один шаг мы попадаем в точку $r_2' = r_1 - \frac{1}{2} \cdot f'(r_1) = 4 - \frac{1}{2} \cdot 4 = 2$ (мы написали не r_2, а r_2', чтобы подчеркнуть, что в случае learning rate $= \frac{1}{2}$ мы получили другую точку, чем в случае learning rate $= 1$). В точке r_2' производная функции f равна нулю, значит, градиентный спуск в ней останавливается.

Заметим, что если $x > 1 + \frac{1}{2}$, то x не попадает в 1-окрестность точки глобального минимума (ведь самая большая точка глобального минимума это $\frac{1}{2}$). Следовательно, точка $r_2' = 2$, в которой останавливается градиентный спуск, не попадает в 1-окрестность точки глобального минимума.

Отметим, что в приведённом выше примере функция f не была дифференцируемой на всём $\mathbb{R}$: производная не существовала в точках $\{-\frac{1}{2}, \frac{1}{2}, 1, 3\}$. Однако в нашем случае градиентный спуск ни в какой момент не попадал ни в одну из этих точек, поэтому проблем у нас не возникло. Если всё же хочется привести пример всюду дифференцируемой функции, то можно воспользоваться так называемой процедурой «сглаживания», которая выходит за рамки нашего курса. Неформально говоря, мы сможем заменить изломы на графике на более округлые кривые.

Задача 4

Существует ли дифференцируемая функция $f : \mathbb{R} \to \mathbb{R}$ и точка r_1, такие что:

- функция f достигает глобального минимума,
- $f'(r_1) \ne 0$,
- градиентный спуск с начальной точкой r_1 и любым положительным learning rate не достигает 1-окрестности глобального минимума.

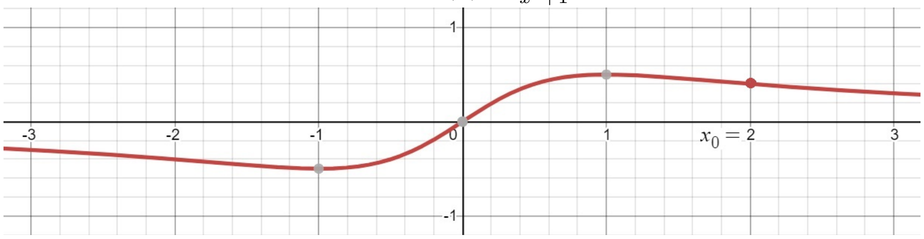
Если да — приведите пример функции f и точки r_1. Если нет — докажите, что такой функции не существует.

Если ответ «да», то достаточно объяснить, как выглядит график такой функции f и где должна располагаться на нём точка r_1. Формулу для $f(x)$ находить не обязательно.

Подсказка. Learning rate можно выразить так: $\lambda = \frac{r_n - r_{n+1}}{f'(r_n)}$. Можно ли придумать функцию f и точку r_1 так, что для всех n дробь $\frac{r_n - r_{min}}{f'(r_n)}$ была отрицательной (здесь r_{min} — точка локального минимума)?

Решение. Да, существуют.

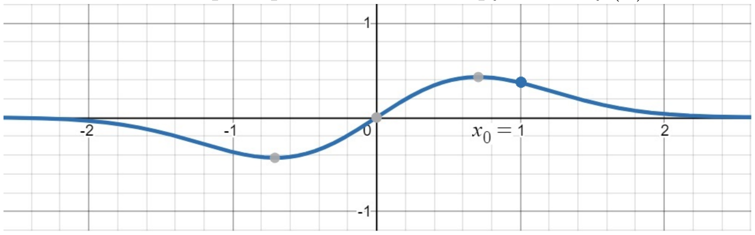
Например, можно взять функцию $f(x) = \frac{x}{x^2+1}$:



и точку $r_1 = 2$.

Действительно, $f'(x) = \frac{1-x^2}{(x^2+1)^2}$ и при $x \geq 2$ $f'(x) < 0$ и градиентный спуск с любым положительным learning rate будет уходить только вправо. А точка глобального минимума функции в точке $x = -1$, которая слева от $r_1 = 2$.

Или в качестве примера можно взять функцию $f(x) = xe^{-x^2}$:



и точку $r_1 = 1$.

В этом случае $f'(x) = e^{-x^2}(1 - 2x^2)$ и при $x \geq 1$ $f'(x) < 0$ и градиентный спуск с любым положительным learning rate будет уходить только вправо. А точка глобального минимума функции в точке $x = -\frac{1}{\sqrt{2}}$, которая слева от $r_1 = 1$.

Задача 5

1. Существует ли функция f, у которой есть бесконечно много локальных минимумов, но нет глобального минимума?

2. Существует ли функция $f : \mathbb{R} \to [0, +\infty)$, у которой есть бесконечно много локальных минимумов, но нет глобального минимума?

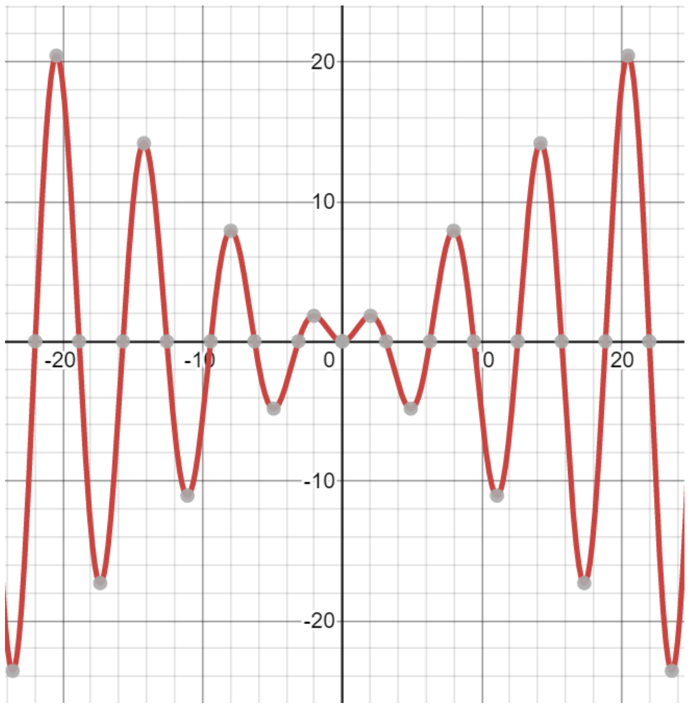
Если да — приведите пример функции f. Если нет — докажите.

Если ответ «да», то достаточно объяснить, как выглядит график такой функции f. Формулу для $f(x)$ находить не обязательно.

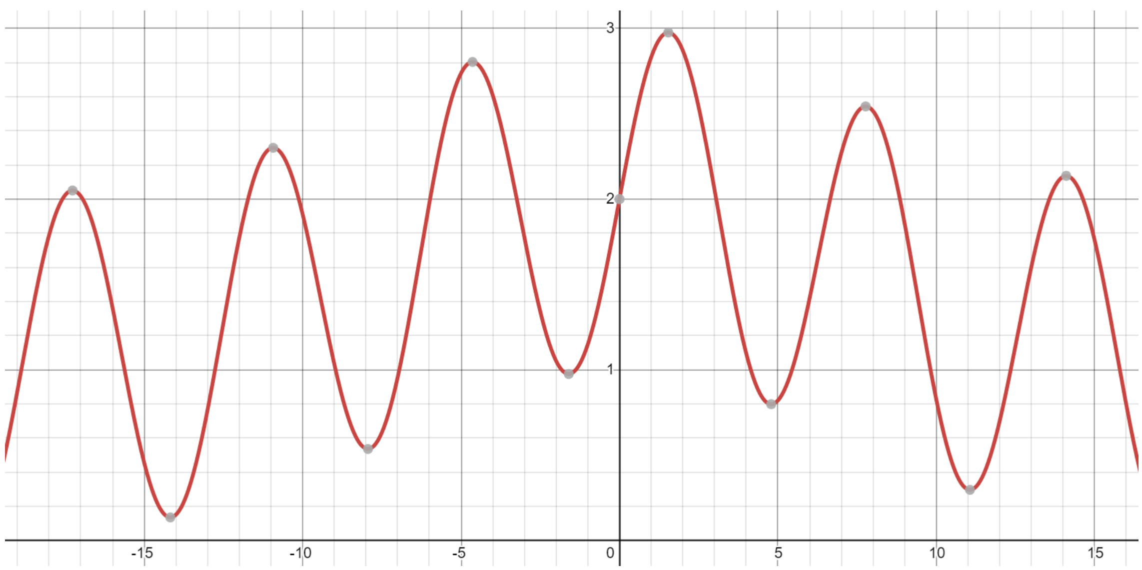
Подсказка. Вспомните, у какой простой функции бесконечно много локальных минимумов. Во втором пункте попробуйте «скомбинировать» такую функцию с функцией, которая приближается к нулю, но не достигает его.

Решение.

1. Например, можно взять функцию $f(x) = x \cdot \sin x$. График у неё такой:

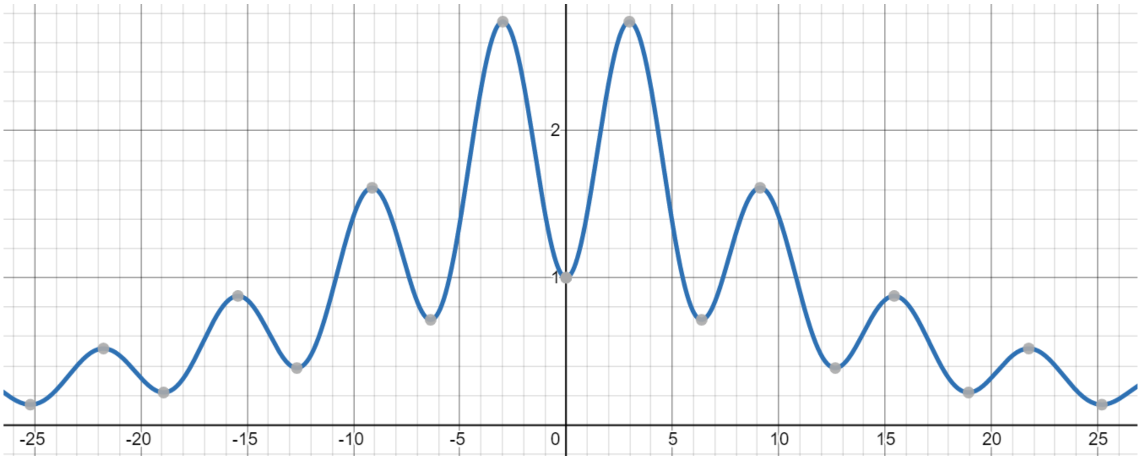


2. Можно взять функцию $f(x) = e^{-x^2/100} + \sin x + 1$.



При всех $x \in \mathbb{R}$ выполнено $f(x) > 0$. Производная $f'(x) = -\frac{x}{50} \cdot e^{-x^2/100} + \cos x$ будет менять знак в бесконечном количестве точек, значит, точек локального минимума будет бесконечно много.

Другой пример: $f(x) = \frac{2-\cos(x)}{1+\frac{x^2}{100}}$.



4.3 $\mathbb{R}^n$: расстояния и векторы.

Задача 1

По аналогии с определением ε-окрестности точки для случая $\mathbb{R}^1$, мы определяем ε-окрестность точки $a \in \mathbb{R}^n$.

Определение. ε-окрестностью точки $a \in \mathbb{R}^n$ называется множество всех таких точек $x \in \mathbb{R}^n$, что расстояние от x до a меньше ε.

1. Как выглядит ε-окрестность точки $a \in \mathbb{R}^2$?
2. Как выглядит ε-окрестность точки $a \in \mathbb{R}^3$?
3. Как записать условие "x лежит в ε-окрестности точки a" в виде условия на координаты точки $x = (x_1, \dots, x_n)$?

Подсказка. Для начала можно вспомнить, какая фигура получится, если рассмотреть все точки, удалённые от a ровно на расстояние ε.

Решение. Определение окрестности можно эквивалентно записать так: ε-окрестность точки $a \in \mathbb{R}^n$ — это $\{x \in \mathbb{R}^n | d(x, a) < \varepsilon\}$.

1. В случае $a \in \mathbb{R}^2$ ε-окрестность точки a — это внутренняя часть круга с центром в точке a и радиусом ε.
2. В случае $a \in \mathbb{R}^3$ ε-окрестность точки a — это внутренняя часть шара с центром в точке a и радиусом ε.
3. Если точка $x = (x_1, \dots, x_n)$ лежит в ε-окрестности точки $a = (a_1, \dots, a_n)$, то $d(x, a) < \varepsilon$.

   В координатах последнее неравенство запишется так:

$$\sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2 + \cdots + (x_n - a_n)^2} < \varepsilon$$

   Или, возведя это неравенство в квадрат, получим:

$$(x_1 - a_1)^2 + (x_2 - a_2)^2 + \cdots + (x_n - a_n)^2 < \varepsilon^2$$

Задача 2

Дана последовательность $\{z_i\}$ точек в $\mathbb{R}^n$. Обозначим k-ую координату точки z_i за z_{ik}. То есть по координатам точки нашей последовательности $\{z_i\}$ запишутся так: $z_1 = (z_{11}, z_{12}, \dots, z_{1n})$, $z_2 = (z_{21}, z_{22}, \dots, z_{2n})$, … $z_i = (z_{i1}, z_{i2}, \dots, z_{in})$, … Заметим, что $z_i \in \mathbb{R}^n$ это точки из многомерного пространства, а вот координаты $z_{ik} \in \mathbb{R}$ это просто обычные действительные числа.

Докажите, что эта последовательность сходится к точке $a = (a_1, \dots, a_n)$ если и только если одновременно выполнены следующие n условий:

- последовательность первых координат точек z_i сходится к a_1, то есть $\lim_{i\to\infty} z_{i1} = a_1$,
- последовательность вторых координат точек z_i сходится к a_2, то есть $\lim_{i\to\infty} z_{i2} = a_2$,
- ...
- последовательность n-ых координат точек z_i сходится к a_n, то есть $\lim_{i\to\infty} z_{in} = a_n$,.

Для понимания может быть полезно сначала решить эту задачу в случае $n = 2$.

Комментарий. В этой задаче вы можете пользоваться следующим утверждением. Последовательность неотрицательных чисел стремится к нулю тогда и только тогда, когда последовательность корней из этих чисел стремится нулю. Другими словами, если для любого i выполнено $y_i > 0$, то

$$\lim_{i\to\infty}(y_i) = 0 \Leftrightarrow \lim_{i\to\infty}(\sqrt{y_i}) = 0.$$

Подсказка. Полезным будет расписать формулу расстояния между точками с использованием координат.

Решение. Запишем в наших обозначениях, что мы хотим доказать:

$$\lim_{i\to\infty} z_i = a \;\Leftrightarrow\; \lim_{i\to\infty} z_{ik} = a_k \text{ для всех } k \in \{1, 2, \dots, n\}$$

Справа налево. Докажем, что если выполнена правая часть формулы выше, то выполнена и левая. Итак, пусть $\lim_{i\to\infty} z_{ik} = a_k$ для всех $k \in \{1, 2, \dots, n\}$.

Мы же хотим доказать, что $\lim_{i\to\infty} z_i = a$, то есть что для любого $\varepsilon > 0$ $\exists N : \forall i \geq N$ $d(z_i, a) < \varepsilon$. Последнее неравенство можно переписать в эквивалентном виде:

$$d(z_i, a) < \varepsilon \;\Leftrightarrow\; \sqrt{(z_{i1} - a_1)^2 + \dots + (z_{in} - a_n)^2} < \varepsilon \;\Leftrightarrow\; (z_{i1} - a_1)^2 + \dots + (z_{in} - a_n)^2 < \varepsilon^2$$

Но, поскольку $\lim_{i\to\infty} z_{ik} = a_k$, то найдётся номер N_k такой, что для любого $i \geq N_k$ выполнено

$$|z_{ik} - a_k| < \frac{\varepsilon}{\sqrt{n}}$$

А тогда, взяв $N = \max(N_1, \dots, N_n)$, мы получим, что при всех $i \geq N$ и всех k выполнено

$$|z_{ik} - a_k| < \frac{\varepsilon}{\sqrt{n}} \;\Leftrightarrow\; (z_{ik} - a_k)^2 < \frac{\varepsilon^2}{n}$$

Просуммировав все n таких неравенств, мы получим

$$(z_{i1} - a_1)^2 + \dots + (z_{in} - a_n)^2 < n \cdot \frac{\varepsilon^2}{n} = \varepsilon^2$$

Таким образом мы доказали нужное нам утверждение и показали, что $\lim_{i\to\infty} z_i = a$.

Слева направо. Теперь докажем, что следствие верно и в другую сторону. А именно, пусть известно, что $\lim_{i\to\infty} z_i = a$.

Докажем, что $\lim_{i\to\infty} z_{ik} = a_k$ для всех $k \in \{1, 2, \dots, n\}$. Воспользуемся методом от противного. То есть, пусть для некоторого k не выполнено $\lim_{i\to\infty} z_{ik} = a_k$. В кванторах это запишется так:

$$\exists \varepsilon > 0 \; \forall N \; \exists i \geq N \; |z_{ik} - a_k| \geq \varepsilon.$$

А тогда, возведя последнее неравенство в квадрат, получим, что для любого N найдётся $i \geq N$ такой что выполнено

$$(z_{ik} - a_k)^2 \geq \varepsilon^2$$

Следовательно, для этого i выполнено

$$(d(z_i, a))^2 = (z_{i1} - a_1)^2 + \dots + (z_{in} - a_n)^2 \geq (z_{ik} - a_k)^2 \geq \varepsilon^2$$

Тогда $(d(z_i, a))^2$ не может стремиться к нулю. А значит по утверждению из условия к нулю не стремится и $d(z_i, a)$. Таким образом мы получили противоречие с тем, что $\lim_{i\to\infty} z_i = a$. То есть наше изначальное предположение было неверно, и должно быть выполнено $\lim_{i\to\infty} z_{ik} = a_k$ для всех $k \in \{1, 2, \dots, n\}$.

Задача 3

Дайте определения:

1. Точки локального минимума функции $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$,
2. Точки глобального минимума функции $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$,
3. Предела функции $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ в точке x_0,
4. Непрерывности функции $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ в точке x_0.

Для понимания может быть полезно сначала дать эти определения в случае $\mathbb{R}^2$, то есть $n = 2$.
По сути, в этой задаче мы просим вас обобщить понятия, которые мы уже определяли для случая $\mathbb{R}^1$ в первых двух неделях. Вот что может пригодиться:

1. Определение локального минимума функции для случая $\mathbb{R}^1$. В определении для случая $\mathbb{R}^1$ мы использовали ε-окрестность в $\mathbb{R}^1$. На прошлых шагах мы научились определять ε-окрестность в $\mathbb{R}^n$.
2. Определение глобального минимума функции для случая $\mathbb{R}^1$.
3. Определение предела функции для случая $\mathbb{R}^1$. Предел функции для случая $\mathbb{R}^1$ определялся через предел последовательности в $\mathbb{R}^1$. На прошлых шагах мы научились определять предел последовательности в $\mathbb{R}^n$.
4. Определение непрерывности функции в точке для случая $\mathbb{R}^1$.

Подсказка. Посмотрите на соответствующие определения в одномерном случае и подумайте, что в них нужно изменить (и нужно ли вообще).

Решение.

1. Точка $x_0 \in D$ называется *точкой локального минимума*, если существует такой $\varepsilon > 0$, что $f(x) \geq f(x_0)$ для всех $x \in D \cap U_\varepsilon(x_0)$, где за $U_\varepsilon(x_0)$ мы обозначили ε-окрестность точки x_0.
2. Точка x_0 называется *точкой (глобального) минимума* функции f, если $f(x_0) \leq f(x)$ для всех $x \in D$.
3. Число $a \in \mathbb{R}^n$ называется *пределом функции* f в точке x_0, если $\lim\limits_{i\to\infty} f(x_i) = a$ для любой последовательности $\{x_i\}$, такой что $\lim\limits_{i\to\infty}(x_i) = x_0$ и $x_i \in D \setminus \{x_0\}$ для всех i (мы предполагаем, что хотя бы одна такая последовательность существует).
4. Число $a \in \mathbb{R}^n$ называется *пределом функции* f в точке 0, если $\lim\limits_{i\to\infty} f(x_i) = a$ для любой последовательности $\{x_i\}$, такой что $\lim\limits_{i\to\infty}(x_i) = x_0$ и $x_i \in D \setminus \{x_0\}$ для всех i (мы предполагаем, что хотя бы одна такая последовательность существует).
5. Функция f называется *непрерывной* в точке x_0, если f определена в точке x_0 и $\lim\limits_{x\to x_0} f(x)$ существует и равен $f(x_0)$.

Задача 4

Пусть $x, y \in \mathbb{R}^n$. Докажите, что $d(x, y) = ||x - y||$.
Попробуйте понять, какая геометрическая интерпретация есть у этого равенства.
Заметьте, что в левой части равенства мы используем расстояние между точками x и y, а в правой — длину вектора $x - y$.

Подсказка. Запишите определения.

Решение. Пусть $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$. По определению

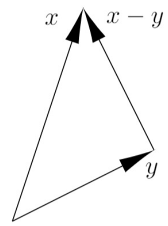
$$d(x,y) = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2 + \dots + (x_n-y_n)^2}.$$

Запишем теперь, что представляет из себя вектор $x-y$: $x-y=(x_1-y_1, x_2-y_2, \dots, x_n-y_n)$. Тогда его длина равна

$$||x-y|| = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2 + \dots + (x_n-y_n)^2}.$$

Прекрасно, мы получили одинаковые выражения! Значит, $d(x,y) = ||x-y||$.

Геометрически разность векторов представляется так:



Действительно, мы видим, что длина вектора $x-y$ равна расстоянию между точками x и y.

4.4 Дифференциал

Задача 1

Пример. Докажем, что для любой точки $x=(x_1,x_2)$ дифференциал функции $f(x_1,x_2)=3x_1$ равен

$$d_x f(\Delta x_1, \Delta x_2) = 3\ \Delta x_1 + 0\ \Delta x_2 = 3\Delta x_1.$$

Подставим f и $d_x f$ в наше определение дифференциала:

$$\lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{f(x_1+\Delta x_1, x_2+\Delta x_2) - (f(x) + a_1\Delta x_1 + a_2\Delta x_2)}{||(\Delta x_1,\Delta x_2)||} =$$
$$= \lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{3(x_1+\Delta x_1) - (3x_1 + 3\Delta x_1 + 0\Delta x_2)}{||(\Delta x_1,\Delta x_2)||} =$$
$$= \lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{0}{||(\Delta x_1,\Delta x_2)||} = 0.$$

Что и требовалось доказать.

Задача.

1. В начале этого урока мы нестрого показали, что дифференциал функции $f(x_1,x_2) = 4+3x_1+5x_2$ в каждой точке равен $3\ \Delta x_1 + 5\ \Delta x_2$. Докажите это утверждение строго, используя определение дифференциала.

2. Найдите дифференциал функции $f(x_1,x_2) = p+qx_1+rx_2$ в каждой точке.

Подсказка. Задача решается аналогично примеру.

Решение.

1. Докажем, что для любой точки $x=(x_1,x_2)$ дифференциал функции $f(x_1,x_2)=4+3x_1+5x_2$ равен

   $$d_x f(\Delta x_1, \Delta x_2) = 3\ \Delta x_1 + 5\ \Delta x_2$$

   в каждой точке.

   Подставим f и $d_x f$ в наше определение дифференциала:

$$
\lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{f(x_1+\Delta x_1, x_2+\Delta x_2) - (f(x) + a_1\Delta x_1 + a_2\Delta x_2)}{||(\Delta x_1, \Delta x_2)||} =
$$

$$
= \lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{4 + 3(x_1+\Delta x_1) + 5(x_2+\Delta x_2) - (4 + 3x_1 + 5x_2 + 3\Delta x_1 + 5\Delta x_2)}{||(\Delta x_1, \Delta x_2)||} =
$$

$$
= \lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{0}{||(\Delta x_1, \Delta x_2)||} = 0.
$$

2. Докажем, что для любой точки $x = (x_1, x_2)$ дифференциал функции $f(x_1, x_2) = p + qx_1 + rx_2$ равен

$$
d_x f(\Delta x_1, \Delta x_2) = q\ \Delta x_1 + r\ \Delta x_2
$$

в каждой точке.

Подставим f и $d_x f$ в наше определение дифференциала:

$$
\lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{f(x_1+\Delta x_1, x_2+\Delta x_2) - (f(x) + a_1\Delta x_1 + a_2\Delta x_2)}{||(\Delta x_1, \Delta x_2)||} =
$$

$$
= \lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{p + q(x_1+\Delta x_1) + r(x_2+\Delta x_2) - (p + qx_1 + rx_2 + q\Delta x_1 + r\Delta x_2)}{||(\Delta x_1, \Delta x_2)||} =
$$

$$
= \lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{0}{||(\Delta x_1, \Delta x_2)||} = 0.
$$

Задача 2

В первой половине урока мы нестрого нашли дифференциалы функций $x_1^2 + x_2^2$ и $x_1 x_2$. На этом шаге мы сделаем это строго.

Пример. Докажем, что для любой точки $x = (x_1, x_2)$ дифференциал функции $f(x_1, x_2) = x_1^2$ равен

$$
d_x f(\Delta x_1, \Delta x_2) = 2x_1\Delta x_1 + 0\ \Delta x_2 = 2x_1\Delta x_1.
$$

Подставим f и $d_x f$ в наше определение дифференциала:

$$
\lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{f(x_1+\Delta x_1, x_2+\Delta x_2) - (f(x) + a_1\Delta x_1 + a_2\Delta x_2)}{||(\Delta x_1, \Delta x_2)||} =
$$

$$
= \lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{(x_1+\Delta x_1)^2 - (x_1^2 + 2x_1\Delta x_1 + 0\Delta x_2)}{||(\Delta x_1, \Delta x_2)||} =
$$

$$
= \lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{\cancel{x_1^2} + \cancel{2x_1\Delta x_1} + \Delta x_1^2 - \cancel{x_1^2} - \cancel{2x_1\Delta x_1}}{||(\Delta x_1, \Delta x_2)||} = \lim_{(\Delta x_1,\Delta x_2)\to(0,0)} \frac{\Delta x_1^2}{||(\Delta x_1, \Delta x_2)||}.
$$

А этот предел равен нулю. Докажем это.

Неформальное доказательство. Числитель по модулю не больше, чем квадрат длины вектора $(\Delta x_1, \Delta x_2)$. Значит, сама дробь по модулю не больше длины вектора $(\Delta x_1, \Delta x_2)$. Когда $(\Delta x_1, \Delta x_2)$ стремится к нулю, длина вектора $(\Delta x_1, \Delta x_2)$ тоже стремится к нулю.

Формальное доказательство. Рассмотрим любую последовательность векторов $\{(\Delta x_{1n}, \Delta x_{2n})\}$ сходящуюся к точке $(0, 0)$. Нам нужно доказать, что тогда последовательность $\{\frac{\Delta x_{1n}^2}{||(\Delta x_{1n}, \Delta x_{2n})||}\}$ сходится к нулю.

1. Так как выполнено $|\Delta x_{1n}| \le ||(\Delta x_{1n}, \Delta x_{2n})||$, имеем $\frac{\Delta x_{1n}^2}{||(\Delta x_{1n}, \Delta x_{2n})||} \le \frac{||(\Delta x_{1n}, \Delta x_{2n})||^2}{||(\Delta x_{1n}, \Delta x_{2n})||} = ||(\Delta x_{1n}, \Delta x_{2n})||$
2. Так как последовательность векторов $\{(\Delta x_{1n}, \Delta x_{2n})\}$ сходится к точке $(0, 0)$, последовательность $||(\Delta x_{1n}, \Delta x_{2n})||$ сходится к нулю.
3. Так как $0 \le \frac{\Delta x_{1n}^2}{||(\Delta x_{1n}, \Delta x_{2n})||} \le ||(\Delta x_{1n}, \Delta x_{2n})||$ и последовательность $||(\Delta x_{1n}, \Delta x_{2n})||$ сходится к нулю, последовательность $\{\frac{\Delta x_{1n}^2}{||(\Delta x_{1n}, \Delta x_{2n})||}\}$ тоже сходится к нулю.

Что и требовалось доказать.

Комментарий. Несколько шагов назад мы нестрого говорили, что член со второй степенью Δx_1 можно игнорировать. Рассуждение выше формализует это утверждение.

Задача.

1. Докажите, что дифференциал функции $f(x_1, x_2) = x_1^2 + x_2^2$ в точке (x_1, x_2) равен $2x_1\Delta x_1 + 2x_2\Delta x_2$.
2. Докажите, что дифференциал функции $f(x_1, x_2) = x_1 x_2$ в точке (x_1, x_2) равен $x_2\ \Delta x_1 + x_1\Delta x_2$.

Подсказка 1.

1. Чтобы доказать первое утверждение, нужно будет доказать, что $\lim\limits_{(\Delta x_1, \Delta x_2)\to(0,0)} \frac{\Delta x_1^2 + \Delta x_2^2}{||(\Delta x_1, \Delta x_2)||} = 0$
2. Чтобы доказать второе утверждение, нужно будет доказать, что $\lim\limits_{(\Delta x_1, \Delta x_2)\to(0,0)} \frac{\Delta x_1 \Delta x_2}{||(\Delta x_1, \Delta x_2)||} = 0$

Решение.

1. Докажем, что дифференциал функции $f(x_1, x_2) = x_1^2 + x_2^2$ в точке (x_1, x_2) равен $2x_1\Delta x_1 + 2x_2\Delta x_2$.

   Подставим f и $d_x f$ в наше определение дифференциала:

$$\lim_{(\Delta x_1, \Delta x_2)\to(0,0)} \frac{f(x_1 + \Delta x_1, x_2 + \Delta x_2) - (f(x) + a_1\Delta x_1 + a_2\Delta x_2)}{||(\Delta x_1, \Delta x_2)||} =$$
$$= \lim_{(\Delta x_1, \Delta x_2)\to(0,0)} \frac{(x_1 + \Delta x_1)^2 + (x_2 + \Delta x_2)^2 - (x_1^2 + x_2^2 + 2x_1\Delta x_1 + 2x_2\Delta x_2)}{||(\Delta x_1, \Delta x_2)||} =$$
$$= \lim_{(\Delta x_1, \Delta x_2)\to(0,0)} \frac{\cancel{x_1^2} + \cancel{2x_1\Delta x_1} + \Delta x_1^2 + \cancel{x_2^2} + \cancel{2x_2\Delta x_2} + \Delta x_2^2 - \cancel{x_1^2} - \cancel{x_2^2} - \cancel{2x_1\Delta x_1} - \cancel{2x_2\Delta x_2}}{||(\Delta x_1, \Delta x_2)||} =$$
$$= \lim_{(\Delta x_1, \Delta x_2)\to(0,0)} \frac{\Delta x_1^2 + \Delta x_2^2}{||(\Delta x_1, \Delta x_2)||} = \lim_{(\Delta x_1, \Delta x_2)\to(0,0)} \frac{\Delta x_1^2 + \Delta x_2^2}{\sqrt{\Delta x_1^2 + \Delta x_2^2}} = \lim_{(\Delta x_1, \Delta x_2)\to(0,0)} \sqrt{\Delta x_1^2 + \Delta x_2^2} = 0.$$

   Что и требовалось доказать.

2. Докажем, что дифференциал функции $f(x_1, x_2) = x_1 x_2$ в точке (x_1, x_2) равен $x_2\ \Delta x_1 + x_1\Delta x_2$.

   Подставим f и $d_x f$ в наше определение дифференциала:

$$\lim_{(\Delta x_1, \Delta x_2)\to(0,0)} \frac{f(x_1 + \Delta x_1, x_2 + \Delta x_2) - (f(x) + a_1\Delta x_1 + a_2\Delta x_2)}{||(\Delta x_1, \Delta x_2)||} =$$
$$= \lim_{(\Delta x_1, \Delta x_2)\to(0,0)} \frac{(x_1 + \Delta x_1)(x_2 + \Delta x_2) - (x_1 x_2 + x_2\Delta x_1 + x_1\Delta x_2)}{||(\Delta x_1, \Delta x_2)||} =$$
$$= \lim_{(\Delta x_1, \Delta x_2)\to(0,0)} \frac{\cancel{x_1 x_2} + \cancel{x_1\Delta x_2} + \cancel{x_2\Delta x_1} + \Delta x_1\Delta x_2 - \cancel{x_1 x_2} - \cancel{x_2\Delta x_1} - \cancel{x_1\Delta x_2}}{||(\Delta x_1, \Delta x_2)||} =$$
$$= \lim_{(\Delta x_1, \Delta x_2)\to(0,0)} \frac{\Delta x_1\Delta x_2}{||(\Delta x_1, \Delta x_2)||}$$

   А этот предел равен нулю. Докажем это.

   Рассмотрим любую последовательность векторов $\{(\Delta x_{1n}, \Delta x_{2n})\}$ сходящуюся к точке $(0, 0)$. Нам нужно доказать, что тогда последовательность $\{\frac{\Delta x_{1n}\Delta x_{2n}}{||(\Delta x_{1n}, \Delta x_{2n})||}\}$ сходится к нулю.

   (a) $|\Delta x_{1n}| = \sqrt{|\Delta x_{1n}|^2} \le \sqrt{\Delta x_{1n}^2 + \Delta x_{2n}^2} = ||(\Delta x_{1n}, \Delta x_{2n})||$ и аналогично $|\Delta x_{2n}| = \sqrt{|\Delta x_{2n}|^2} \le \sqrt{\Delta x_{1n}^2 + \Delta x_{2n}^2} = ||(\Delta x_{1n}, \Delta x_{2n})||$. Поэтому $|\Delta x_{1n}\Delta x_{2n}| \le ||(\Delta x_{1n}, \Delta x_{2n})||^2$.

(b) Так как выполнено $|\Delta x_{1n}\Delta x_{2n}| \le ||(\Delta x_{1n}, \Delta x_{2n})||^2$, имеем $\frac{|\Delta x_{1n}\Delta x_{2n}|}{||(\Delta x_{1n},\Delta x_{2n})||} \le \frac{||(\Delta x_{1n},\Delta x_{2n})||^2}{||(\Delta x_{1n},\Delta x_{2n})||} = ||(\Delta x_{1n}, \Delta x_{2n})||$

(c) Так как последовательность векторов $\{(\Delta x_{1n}, \Delta x_{2n})\}$ сходится к точке $(0, 0)$, последовательность $||(\Delta x_{1n}, \Delta x_{2n})||$ сходится к нулю.

(d) Так как $0 \le \frac{|\Delta x_{1n}\Delta x_{2n}}{||(\Delta x_{1n},\Delta x_{2n})||} \le \frac{1}{2}||(\Delta x_{1n}, \Delta x_{2n})||$ и последовательность $\frac{1}{2}||(\Delta x_{1n}, \Delta x_{2n})||$ сходится к нулю, последовательность $\{\frac{\Delta x_{1n}\Delta x_{2n}}{||(\Delta x_{1n},\Delta x_{2n})||}\}$ тоже сходится к нулю.

Что и требовалось доказать.

Задача 3

Пример.

- Пусть функция $f : \mathbb{R}^n \to \mathbb{R}$ дифференцируема в точке x, её дифференциал в точке x это $d_x f(\Delta x) := a_1\Delta x_1 + \cdots + a_n\ \Delta x_n$.

- Зафиксируем любое число $c \in \mathbb{R}$.

Докажем, что функция (cf) дифференцируема в точке x, и её дифференциал в точке x это

$$d_x(cf)(\Delta x) := ca_1\Delta x_1 + \cdots + ca_n\ \Delta x_n.$$

Доказательство. Подставим функцию $ca_1\Delta x_1 + \cdots + ca_n\ \Delta x_n$ в определение дифференциала функции (cf). Если полученный предел окажется равным нулю, то это и будет значить, что $ca_1\Delta x_1 + \cdots + ca_n\ \Delta x_n$ – дифференциал функции (cf) в точке x (в частности, мы получим, что дифференциал cf в точке x существует).

$$\lim_{\Delta x\ \to 0} \frac{(cf)(x + \Delta x) - ((cf)(x) + ca_1\Delta x_1 + \cdots + ca_n\ \Delta x_n)}{||\Delta x||} =$$
$$= \lim_{\Delta x\ \to 0} \frac{c \cdot f(x + \Delta x) - (c \cdot f(x) + c \cdot d_x f(\Delta x))}{||\Delta x||} =$$
$$= c \cdot \lim_{\Delta x\ \to 0} \frac{f(x + \Delta x) - (f(x) + d_x f(\Delta x))}{||\Delta x||} = c \cdot 0 = 0.$$

Что и требовалось доказать. Предпоследнее равенство выполнено, потому что $d_x f(\Delta x)$ это дифференциал функции f.

Комментарий. Тем самым, дифференциал ведёт себя как производная – при умножении функции на константу c, дифференциал умножается на c.

Задача.

- Пусть функция $f : \mathbb{R}^n \to \mathbb{R}$ дифференцируема в точке x, её дифференциал в точке x это $d_x f(\Delta x) := a_1\Delta x_1 + \cdots + a_n\Delta x_n$.

- Пусть функция $g : \mathbb{R}^n \to \mathbb{R}$ дифференцируема в точке x, её дифференциал в точке x это $d_x g(\Delta x) = b_1\Delta x_1 + \cdots + b_n\Delta x_n$.

Докажите, что сумма этих двух функций тоже дифференцируема в точке x, и её дифференциал в точке x равен

$$d_x f(\Delta x) + d_x g(\Delta x) = (a_1 + b_1)\Delta x_1 + \cdots + (a_n + b_n)\Delta x_n.$$

Другими словами, докажите, что $d_x(f + g)(\Delta x) = d_x f(\Delta x) + d_x g(\Delta x)$.

Комментарий. Тем самым, дифференциал ведёт себя как производная – если f и g дифференцируемы в точке x, то $(f + g)$ тоже дифференцируема в точке x, и её дифференциал равен сумме дифференциала f и дифференциала g.

Подсказка. Требуется доказать, что $\lim\limits_{\Delta x\ \to 0} \frac{(f+g)(x+\Delta x)-((f+g)(x)+d_x f(\Delta x)+d_x g(\Delta x))}{||\Delta x||}$

Решение. Подставим функцию $(a_1 + b_1)\Delta x_1 + \cdots + (a_n + b_n)\Delta x_n$ в определение дифференциала функции $(f + g)$. Если полученный предел окажется равным нулю, то это и будет значить, что $(a_1 + b_1)\Delta x_1 + \cdots +$

$(a_n+b_n)\Delta x_n$ – дифференциал функции $(f+g)$ в точке x (в частности, мы получим, что дифференциал $f+g$ в точке x существует).

$$\lim_{\Delta x\ \to 0}\frac{(f+g)(x+\Delta x)-((f+g)(x)+(a_1+b_1)\Delta x_1+\dots+(a_n+b_n)\Delta x_n)}{||\Delta x||}=$$
$$=\lim_{\Delta x\ \to 0}\frac{f(x+\Delta x)+g(x+\Delta x)-(f(x)+g(x)+d_xf(\Delta x)+d_xg(\Delta x))}{||\Delta x||}=$$
$$=\lim_{\Delta x\ \to 0}\frac{f(x+\Delta x)-(f(x)+d_xf(\Delta x))}{||\Delta x||}+\lim_{\Delta x\ \to 0}\frac{g(x+\Delta x)-(g(x)+d_xg(\Delta x))}{||\Delta x||}=0+0=0.$$

Что и требовалось доказать.

4.5 Частная производная

Задача 1

1. Чему равняется частная производная функции $f(x_1,x_2,x_3)=x_1^2x_2-x_1x_2x_3+x_2^6x_3^3+x_3+10$ по x_2 в точке $(4,1,-3)$?
2. Чему равняется частная производная функции $g(x_1,x_2)=x_1^4-x_1^{x_2+2}-\operatorname{tg}x_2$ по x_1 в точке $(x_1,x_2)=(2,3)$?
3. Чему равняется частная производная функции $h(x_1,x_2,x_3)=\ln(\cos(\frac{\sqrt{\frac{x_1^2+x_3^2-x_1x_3}{2\sin(\operatorname{arctg}(x_3-x_1))}}+x_1-x_3^{sin x_1}}{e^{x_3}})+\sin(\frac{2^{\operatorname{tg}(x_1x_3)}}{\ln(x_1x_3)}))+2x_2$ по второй координате в точке $(x_1,x_2,x_3)=(e^{\sqrt{75}},\cos 197,\operatorname{arctg}1453)$?

Решение.

1. Подставим $x_1=4$ и $x_3=-3$ в $f(x_1,x_2,x_3)=x_1^2x_2-x_1x_2x_3+x_2^6x_3^3+x_3+10$. Получим
$$f(x_2)=4^2x_2-4x_2(-3)+x_2^6(-3)^3+(-3)+10=-27x_2^6+28x_2+7$$
Производная этой функции по x_2 равна:
$$-27\cdot 6x_2^5+28.$$
Подставляя значение $x_2=1$, получаем ответ
$$-27\cdot 6\cdot 1^5+28=-134.$$
2. Подставим $x_2=3$ в $g(x_1,x_2)=x_1^4-x_1^{x_2+2}-\operatorname{tg}x_2$. Получим
$$g(x_2)=x_1^4-x_1^{3+2}-\operatorname{tg}3=-x_1^5+x_1^4-\operatorname{tg}3$$
Производная этой функции по x_1 равна:
$$-5x_1^4+4x_1^3.$$
Подставляя значение $x_1=2$, получаем ответ
$$-5\cdot 2^4+4\cdot 2^3=-80+32=-48.$$
3. Подставим конкретные значения x_1 и x_3 в функцию $h(x_1,x_2,x_3)=\ln(\cos(\frac{\sqrt{\frac{x_1^2+x_3^2-x_1x_3}{2\sin(\operatorname{arctg}(x_3-x_1))}}+x_1-x_3^{sin x_1}}{e^{x_3}})+\sin(\frac{2^{\operatorname{tg}(x_1x_3)}}{\ln(x_1x_3)}))+2x_2$. Получится функция вида $c+2x_2$, где $c\in\mathbb{R}$ это константа. Производная функции $c+2x_2$ по x_2 это 2 (для любого x_2). Поэтому ответ 2.

Задача 2

1. Чему равняется частная производная функции $f(x_1, x_2, \dots, x_n) = 7 + x_1 + 2x_2 + \dots + nx_n$ по x_5 в точке $(10, 10, \dots, 10)$?

2. Чему равняется частная производная функции $g(x_1, x_2, \dots, x_9) = x_1 x_2 \dots x_9$ по x_4 в точке $(2, 2, \dots, 2)$?

Решение.

1. Подставим значения всех x_i кроме x_5 в $f(x_1, x_2, \dots, x_n) = 7 + x_1 + 2x_2 + \dots + nx_n$. Получится функция вида $c+5x_5$, где $c \in \mathbb{R}$ это константа. Производная функции $c+5x_5$ по x_5 это 5 (для любого x_5). Поэтому ответ 5.

2. Подставим значения всех x_i кроме x_4 в функцию $g(x_1, x_2, \dots, x_9) = x_1 x_2 \dots x_9$. Получится функция $g(x_4) = 2^3 x_4 2^5 = 2^8 x_4 = 256 x_4$. Производная функции $256x_4$ по x_4 это 256 (для любого x_4). Поэтому ответ 256.

Задача 3

Теорема. Дана функция f от n переменных. Пусть у f в точке x существует дифференциал $d_x f(\Delta x_1, \dots, \Delta x_n) = a_1 \Delta x_1 + \dots + a_n \ \Delta x_n$ и частные производные $\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}$. Тогда

$$a_1 = \frac{\partial f}{\partial x_1}, \dots, a_n \ = \frac{\partial f}{\partial x_n}.$$

То есть для любого $j = 1, \dots, n$ число a_j равно частной производной функции f по j-ой координате, вычисленной в точке x. Другими словами:

$$d_x f(\Delta x_1, \dots, \Delta x_n) = \frac{\partial f}{\partial x_1} \Delta x_1 + \frac{\partial f}{\partial x_2} \Delta x_2 \ + \dots + \frac{\partial f}{\partial x_n} \Delta x_n,$$

где все частные производные вычислены в точке x.
Задача. Докажите теорему.

Решение. Ничем принципиально не отличается от доказательства для случая $n = 2$.
Будем доказывать равенство $a_1 = \frac{\partial f}{\partial x_1}$.
Часть 1.
Вспомним определение дифференциала. Пусть f – функция от двух переменных. Функция $d_x f(\Delta x_1, \dots, \Delta x_n) = a_1 \Delta x_1 + \dots + a_n \Delta x_n$ называется дифференциалом функции f в точке x, если следующий предел существует и равен нулю:

$$\lim_{(\Delta x_1, \dots, \Delta x_n) \to (0, \dots, 0)} \frac{f(x_1 + \Delta x_1, \dots, x_n + \Delta x_n) - (f(x) + a_1 \Delta x_1 + \dots + a_n \Delta x_n)}{||(\Delta x_1, \dots, \Delta x_n)||} = 0.$$

Предел функции определяется через пределы всевозможных последовательностей – то есть равенство из определения предела выполнено для любой последовательности $\{(\Delta x_{1i}, \dots, \Delta x_{ni})\}$, сходящейся к точке $(0, \dots, 0)$. Выберем какую-нибудь конкретную последовательность $\{(\Delta x_{1i}, \dots, \Delta x_{ni})\} = \{(\Delta x_{1i}, 0, 0, \dots, 0)\}$, сходящуюся к $(0, \dots, 0)$, такую что все $x_{1i} > 0$ (а все $x_{2i} = \dots = x_{ni} = 0$).

Давайте подставлять нашу последовательность $\{(\Delta x_{1i}, \dots, \Delta x_{ni})\}$ в выражение из определения дифференциала. Для этого сначала заметим, что при $x_{1i} > 0$ и $x_{2i} = \dots = x_{ni} = 0$ выполнено:

$$||(\Delta x_{1i}, \dots, \Delta x_{ni})|| = \sqrt{\Delta x_{1i}^2} = \Delta x_{1i}.$$

Подставив $||(\Delta x_{1i}, \dots, \Delta x_{ni})|| = \Delta x_{1i}$ и $\Delta x_{2i} = \dots = \Delta x_{ni} = 0$ в выражение из определения дифференциала, получаем следующее условие (предел в этом выражении это уже не предел функции, а предел последовательности):

$$0 = \lim_{(\Delta x_{1i},\ldots,\Delta x_{ni})\to(0,\ldots,0)} \frac{f(x_1 \ + \Delta x_{1i}, \ldots, x_n + \Delta x_{ni}) - (f(x) + a_1\Delta x_{1i} + \cdots + a_n \Delta x_{ni})}{||(\Delta x_{1i}, \ldots, \Delta x_{ni})||} =$$

$$= \lim_{\Delta x_{1i} \ \to 0} \frac{f(x_1 \ + \Delta x_{1i}, x_2, \ldots, x_n) - (f(x) + a_1\Delta x_{1i})}{\Delta x_{1i}}$$

Посмотрим на второй из этих пределов. Вынесем член $\frac{a_1 \Delta x_{1i}}{\Delta x_{1i}}$ из дроби и получим

$$0 = \lim_{\Delta x_{1i} \ \to 0} \left(\frac{f(x_1 \ + \Delta x_{1i}, x_2, \ldots, x_n) - f(x)}{\Delta x_{1i}} - a_1 \right).$$

Что равносильно

$$\lim_{\Delta x_{1i} \ \to 0} \frac{f(x_1 \ + \Delta x_{1i}, x_2, \ldots, x_n) - f(x)}{\Delta x_{1i}} = a_1.$$

Запомним это равенство.

Часть 2.

Теперь посмотрим на частную производную по первой координате в точке x. По определению частной производной

$$\lim_{\Delta x_1 \ \to 0} \frac{f(x_1 + \Delta x_1, x_2, \ldots, x_n) - f(x_1, \ldots, x_n)}{\Delta x_1} = \frac{\partial f}{\partial x_1}.$$

Это предел функции. Он определяется через пределы всевозможных последовательностей. Рассмотрим последовательность $\{x_{1i}\}$ (эти x_{1i} в точности те же, что и в Части 1 нашего доказательства). По определению предела функции, для этой последовательности должно быть выполнено

$$\lim_{\Delta x_{1i} \ \to 0} \frac{f(x_1 \ + \Delta x_{1i}, x_2, \ldots, x_n) - f(x)}{\Delta x_{1i}} = \frac{\partial f}{\partial x_1}.$$

Часть 3.

Левая часть последнего равенства из Части 1 совпадает с левой частью последнего равенства Части 2. Значит, $a_1 = \frac{\partial f}{\partial x_1}$. Что и требовалось доказать.

Аналогично для любого $1 < k \leq n$ доказывается, что a_k равно частной производной функции f по k-ой координате, вычисленной в точке x.

Мы доказали теорему.

Задача 4

Пример.

Вычислим дифференциал функции

$$f(x_1, x_2, x_3) = x_1 x_2^3 + x_2 x_3^2$$

в точке $(4, -2, 1)$. Воспользуемся теоремой с предыдущего шага и вычислим сначала частные производные:

- $\left.\dfrac{\partial(x_1x_2^3 + x_2x_3^2)}{\partial x_1}\right|_{(4,-2,1)} = x_2^3 = -8$
- $\left.\dfrac{\partial(x_1x_2^3 + x_2x_3^2)}{\partial x_2}\right|_{(4,-2,1)} = 3x_1x_2^2 + x_3^2 = 49.$
- $\left.\dfrac{\partial(x_1x_2^3 + x_2x_3^2)}{\partial x_3}\right|_{(4,-2,1)} = 2x_2x_3 = -4.$

Следовательно, дифференциал f в этой точке будет равен

$$d_x f(\Delta x_1, \Delta x_2) = -8\Delta x_1 + 49\Delta x_2 - 4\Delta x_3.$$

Задача.

1. Найдите дифференциал функции $g(x_1, x_2, x_3) = -x_1^4 x_3 + x_1 x_2^3 x_3 + x_1 x_3^2$ в точке $(2, 5, 4)$
2. Найдите дифференциал функции $h(x_1, x_2, \ldots, x_n) = x_1^2 x_2 + x_2^2 x_3 + \cdots + x_{n-1}^2 x_n + x_n^2 x_1$ в точке $(2, 2, \ldots, 2)$

Решение.

1. Найдём частные производные $g(x_1, x_2, x_3) = -x_1^4 x_3 + x_1 x_2^3 x_3 + x_1 x_3^2$ в точке $(2, 5, 4)$
   - $\left.\dfrac{\partial(-x_1^4 x_3 + x_1 x_2^3 x_3 + x_1 x_3^2)}{\partial x_1}\right|_{(2,5,4)} = -4x_1^3 \cdot 4 + 5^3 \cdot 4 + 4^2 = -16 \cdot 2^3 + 4 \cdot 5^3 + 16 = 388$
   - $\left.\dfrac{\partial(-x_1^4 x_3 + x_1 x_2^3 x_3 + x_1 x_3^2)}{\partial x_2}\right|_{(2,5,4)} = 0 + 2 \cdot 3x_2^2 \cdot 4 + 0 = 24 \cdot 5^2 = 600.$
   - $\left.\dfrac{\partial(-x_1^4 x_3 + x_1 x_2^3 x_3 + x_1 x_3^2)}{\partial x_3}\right|_{(2,5,4)} = -2^4 + 2 \cdot 5^3 + 2 \cdot 2x_3 = -16 + 250 + 16 = 250.$

   Следовательно в точке $(2, 5, 4)$ дифференциал g равен $d_x g(\Delta x_1, \Delta x_2, \Delta x_3) = 388\Delta x_1 + 600\Delta x_2 + 250\Delta x_3$.
2. Найдём дифференциал функции $h(x_1, x_2, \ldots, x_n) = x_1^2 x_2 + x_2^2 x_3 + \cdots + x_{n-1}^2 x_n + x_n^2 x_1$ в точке $(2, 2, \ldots, 2)$ В формуле h каждый x_i входит ровно в два слагаемых: в слагаемое $x_{i-1}^2 x_i$ и в слагаемое $x_i^2 x_{i+1}$. Это очевидно для $1 < i < n$ (например x_2 входит только в слагаемые $x_1^2 x_2$ и $x_2^2 x_3$). Для $i = 1$ нужно заменить индекс $i - 1$ на n, а для $i = n$ нужно заменить индекс $i + 1$ на 1. Итого для любого i выполнено

$$\left.\frac{\partial(x_1^2 x_2 + x_2^2 x_3 + \cdots + x_{n-1}^2 x_n + x_n^2 x_1)}{\partial x_i}\right|_{(2,\ldots,2)} = \left.\frac{\partial(x_{i-1}^2 x_i + x_i^2 x_{i+1})}{\partial x_i}\right|_{(2,\ldots,2)} = 2^2 + 2x_i \cdot 2 = 4 + 8 = 12.$$

Задача 5

Как вы помните, в случае функции одной переменной верно следующее утверждение:

Утверждение. Если функция f непрерывна и дифференцируема на отрезке $[a, b]$, то локальный минимум или максимум не может достигаться в точках $x \in (a, b)$, в которых $f'(x) \neq 0$.

Сформулируем аналогичное утверждение для функции от нескольких переменных.

Теорема. Дана функция f от n переменных. Пусть f определена в некоторой окрестности точки x, и в точке x у f существуют частные производные по всем координатам. Тогда x может быть точкой локального минимума или максимума только если все частные производные равны нулю.

Другими словами, если для какого-то k частная производная f по k-ой координате не равна нулю, то x не может быть точкой локального минимума или максимума.

Следствие. Пусть в точке x так же существует дифференциал $d_x f$. Точка x может быть точкой локального минимума или максимума, только если $d_x f = 0$ (то есть $d_x f(\Delta x_1, \ldots, \Delta x_n) = 0$ для любых $\Delta x_1, \ldots, \Delta x_n$).

Геометрически думать про это можно так: точка x может быть локальным минимумом или максимумом, если касательная плоскость к графику f в точке x будет горизонтальной (см геометрическую интерпретацию дифференциала с этого шага)

Задача. Докажите теорему и следствие.

Комментарий. Уровень строгости доказательства остаётся на ваше усмотрение (можно ограничиться уровнем строгости рассуждения с этого шага)

Решение.

Доказательство теоремы. Докажем, что если в точке x частная производная $\frac{\partial(f(x))}{\partial x_1}$ не равна нулю, то x не может быть точкой локального минимума или максимума.

Пусть, например, $\frac{\partial(f(x))}{\partial x_1}$ в точке x больше 0. То есть $\frac{\partial(f(x))}{\partial x_1} = a$ для некоторого числа $a > 0$. По нашему неформальному определению частной производной, для небольших Δx_1 выполнено $f(x_1 + \Delta x_1, x_2, \ldots, x_n) \approx f(x) + a\Delta x_1$.

- Если $\Delta x_1 > 0$, то, так как $a > 0$, выполнено $a\Delta x_1 > 0$. А значит, $f(x_1 + \Delta x_1, x_2, \ldots, x_n) < f(x) + a\Delta x \approx f(x_1 + \Delta x_1, x_2, \ldots, x_n)$ при достаточно маленьких Δx_1. То есть значение функции f в точке $(x_1 + \Delta x_1, x_2, \ldots, x_n)$ больше, чем значение f в точке x. Значит, точка x не может быть точкой локального максимума.

- Если $\Delta x_1 < 0$, то, так как $a > 0$, выполнено $a\Delta x_1 < 0$. А значит, $f(x_1 + \Delta x_1, x_2, \dots, x_n) > f(x) + a\Delta x \approx\ f(x_1 + \Delta x_1, x_2, \dots, x_n)$ при достаточно маленьких (по модулю) Δx_1. То есть значение функции f в точке $(x_1 + \Delta x_1, x_2, \dots, x_n)$ меньше, чем значение f в точке x. Значит, точка x не может быть точкой локального минимума.

Аналогично для любого $1 < k \leq n$ доказывается, что если частная производная $\frac{\partial(f(x))}{\partial x_k}$ не равна нулю, то x не может быть точкой локального минимума или максимума. Тем самым, теорема доказана.

Доказательство следствия. Дифференциал выражается через частные производные. Точка x может быть точкой локального минимума или максимума, только если все частные производные в этой точке равны нулю. Если все частные производные равны нулю, то $d_x f(\Delta x_1, \dots, \Delta x_n) = 0 \cdot \Delta x_1 + \dots + 0 \cdot \Delta x_n = 0$.

4.6 Направление и градиент

Задача 1

Дан произвольный вектор $a := (a_1, \dots, a_n) \in \mathbb{R}^n$, не равный нулевому вектору. Докажите, что вектор $(\frac{-a_1}{||a||},\ \frac{-a_2}{||a||}, \dots\ , \frac{-a_n}{||a||}) = -\frac{a}{||a||}$ действительно является направлением – то есть, что его длина действительно равна 1.

Решение.
Давайте найдём длину вектора $(\frac{-a_1}{||a||},\ \frac{-a_2}{||a||}, \dots\ , \frac{-a_n}{||a||})$. Она равна

$$\sqrt{\left(\frac{-a_1}{||a||}\right)^2 + \dots + \left(\frac{-a_n}{||a||}\right)^2} = \sqrt{\frac{a_1^2 + \dots + a_n^2}{||a||^2}} = \sqrt{\frac{||a||^2}{||a||^2}} = 1.$$

Задача 2

Как мы уже видели, у любого ненулевого вектора есть длина и направление. Докажите, что они однозначно задают вектор.

Другими словами, если дано положительное число l и вектор e длины 1, то существует и единственен вектор a, такой что длина a это l, а направление a это e.

Задача для проверки. Известно, что вектор a имеет направление $\left(\frac{3}{5}, \frac{-4}{5}\right)$ и длину 10. Восстановите вектор a по этой информации.

Подсказка. Можно ли a выразить через e и l?

Решение. Положим $a := l \cdot e = (le_1, \dots, le_2)$. Тогда $||a|| = ||l \cdot e|| = l \cdot ||e|| = l \cdot 1 = l$, то есть у a действительно длина l.

Направление a по определению равно $\frac{a}{||a||} = \frac{l \cdot e}{l} = e$, то есть направление a это действительно e.

Докажем единственность от противного. Пусть найдётся вектор $b \neq a$ такой, что $||a|| = ||b|| = l$ и направление a и b это e. Запишем, что означает равенство направлений: $\frac{a}{||a||} = \frac{b}{||b||}$, но, поскольку $||a|| = ||b||$, получаем, что $a = b$. Мы пришли к противоречию..

Задача 3

Пример. Дана функция $f(x_1, x_2) = x_1 x_2$, мы находимся в точке $x = (5, 12)$. Найдём значение функции в точке, в которой мы окажемся, совершив из начальной точки шаг длины 2 по направлению, противоположному направлению градиента. Будет ли значение f в этой точке меньше, чем значение f в точке x?

- Сначала вычислим градиент:
$$\nabla f(x) = (x_2, x_1) = (12, 5).$$

- Следовательно, шаг мы будет делать в направлении
$$\frac{-\nabla f(x)}{||\nabla f(x)||} = (-\frac{12}{\sqrt{5^2 + 12^2}}, -\frac{5}{\sqrt{5^2 + 12^2}}) = (-\frac{12}{13}, -\frac{5}{13}).$$

- Так как длина шага должна быть равна 2, мы сместимся на вектор $2 \cdot (-\frac{12}{13}, -\frac{5}{13}) = (-\frac{24}{13}, -\frac{10}{13})$, и попадём в точку $(5 - \frac{24}{13}, 12 - \frac{10}{13})$.

- Значение функции в получившейся точке тогда будет равно

$$f(5 - \frac{24}{13}, 12 - \frac{10}{13}) = \frac{41}{13} \cdot \frac{146}{13} \approx 35.42,$$

 что действительно меньше, чем $f(5, 12) = 5 \cdot 12 = 60$.

Задача. Дана функция $f(x_1, x_2) = (x_1 + x_2)^2$, мы находимся в точке $(3, 4)$. Значение в этой точке равно $f(3, 4) = (3 + 4)^2 = 49$.

Мы делаем шаг длины $2\sqrt{2}$ в направлении, противоположном направлению градиента.

Каково значение функции f в точке, в которой мы оказались?

Подсказка. Сначала надо посчитать градиент (частные производные по x_1 и по x_2), затем разделить полученный вектор на его длину. Так мы получим направление градиента. Затем в получившемся векторе надо поменять знак у каждой координаты, чтобы получить направление, противоположное направлению градиента. И наконец прибавить получившийся вектор, умноженный на $2\sqrt{2}$, к точке a.

Решение.

Посчитаем градиент: $\nabla f(x_1, x_2) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2})$.

$$\frac{\partial f}{\partial x_1} = \frac{\partial (x_1 + x_2)^2}{\partial x_1} = \frac{\partial (x_1^2 + x_2^2 + 2x_1 x_2)}{\partial x_1} = \frac{\partial (x_1^2)}{\partial x_1} + \frac{\partial (2x_1 x_2)}{\partial x_1} = 2x_1 + 2x_2.$$

Аналогично получаем, что $\frac{\partial f}{\partial x_2} = 2(x_1 + x_2)$.

Итого мы получили, что $\nabla f(x_1, x_2) = (2(x_1 + x_2), 2(x_1 + 2))$. В точке $a = (3, 4)$ градиент $\nabla f(a) = (2 \cdot (3 + 4), 2 \cdot (3 + 4)) = (14, 14)$.

Найдём направление градиента:

$$\frac{\nabla f(a)}{||\nabla f(a)||} = \frac{(14, 14)}{\sqrt{14^2 + 14^2}} = \frac{(14, 14)}{14 \cdot \sqrt{2}} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

Тогда направление, противоположное направлению градиента, это $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$.

Теперь мы из точки $a = (3, 4)$ делаем шаг длины $2\sqrt{2}$ в направлении, противоположном направлению градиента, и попадаем в точку $a + 2\sqrt{2} \cdot \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) = (3, 4) + (-2, -2) = (1, 2)$. И в этой точке $f(1, 2) = (1 + 2)^2 = 9$ (видим, что мы получили число, меньшее изначального значения 49).