Математический анализ, тема 3. Функции нескольких переменных

Функция нескольких переменных

Функция (от) нескольких переменных записывается как $u = f(x_1, x_2, \dots, x_n)$.

В этом курсе чаще встречаются функции от двух переменных.

Функция двух переменных — это такая зависимость переменной z от переменных x и y, где каждой уникальной паре (x, y) соответствует единственное значение z.

Записывают её как $z = f(x, y)$ или $z = z(x, y)$. Переменная z — зависимая, а аргументы x и y — независимые.

Другое определение:

Функция двух переменных — это отображение множества $\mathbb{R}^2$ в $\mathbb{R}$, где каждому элементу первого множества соответствует только один элемент второго.

$\mathbb{R}^2$ — это множество всевозможных пар действительных чисел. Его можно воспринимать как множество координат всех точек двумерной плоскости. $\mathbb{R}$ — обычное множество действительных чисел.

Чтобы найти значение функции $z = f(x, y)$, нужно подставить в неё пару x и y и вычислить. Если нужно вычислить сразу много значений, то удобно оформить их в виде таблицы: в первой строке пишут значения аргументов x, в первом столбце — значения y, а во внутренних ячейках таблицы — соответствующие значения функции z.

График функции двух переменных — это множество точек в пространстве с координатами (x, y, z), где $z = f(x, y)$.

В общем случае график функции двух переменных — это некая поверхность в пространстве.

То есть каждой точке из плоскости $x\mathrm{O}y$ функция ставит в соответствие высоту z над или под плоскостью $x\mathrm{O}y$ с соответствующим знаком.

Получается набор точек в пространстве с тремя координатами.

Один из самых простых графиков в 3D — это плоскость, её уравнение: $Ax + By + Cz + D = 0$, где любой из коэффициентов может быть нулевым. Исключением будет лишь ситуация $A = B = C = 0$, тогда плоскости не получится. Если $C \neq 0$, то уравнение можно переписать в виде функции: $z = -\frac{A}{C}x - \frac{B}{C}y - \frac{D}{C}$. Видно, что это уравнение является линейным.

Функция двух переменных всегда «наследует» характер зависимости от каждого своего аргумента: например, если в уравнении был x^2 или y^2 — жди на графике параболических изгибов.

Линии уровня (изолинии) — это сечения поверхности $z = f(x, y)$ различными горизонтальными плоскостями вида $z = C$, где $C \in \mathbb{R}$.

Это горизонтальные «срезы» поверхности на различных высотах. На разной высоте срезы будут выглядеть по-разному, и если взять несколько штук и выложить на плоскость $x\mathrm{O}y$, то получится хорошая визуализация функции двух переменных без построения 3D-графика.

Частная производная

Частная производная функции двух переменных по одной из них — это скорость роста функции в направлении соответствующей оси.

У функции двух переменных есть две частные производные:

- по иксу, обозначается как z'_x или $\frac{\partial z}{\partial x}$;
- по игреку, обозначается как z'_y или $\frac{\partial z}{\partial y}$.

Частная производная — это функция, а частная производная в точке — это число.

Знак частной производной в точке можно определять на глаз по графику. Для этого нужно посмотреть, растёт функция или убывает по мере продвижения в положительную сторону по соответствующей оси. Если функция в точке растёт по мере увеличения икса, то $z'_x > 0$, а если убывает, то $z'_x < 0$. Аналогично про z'_y.

Чтобы вычислить частную производную, можно использовать правила дифференцирования функции одной переменной, а также правила для суммы, разности, произведения, частного и сложной функции. Но у функции двух переменных:

- когда мы ищем частную производную по x, переменная y считается константой;
- когда мы ищем частную производную по y, переменная x считается константой.

Таблица производных для функции одной переменной

$f(x)$	$f'(x)$
C	0
x^n	$n \cdot x^{n-1}$
e^x	e^x
a^x	$a^x \cdot \ln a$
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\operatorname{tg} x$	$\frac{1}{\cos^2 x}$
$\operatorname{ctg} x$	$-\frac{1}{\sin^2 x}$
$\ln x$	$\frac{1}{x}$
$\log_a x$	$\frac{1}{x \cdot \ln a}$

Практический смысл частных производных: они помогают выяснить, какая из переменных в данной точке сильнее влияет на итоговое значение функции. Чем больше модуль частной производной в данной точке, тем сильнее она влияет на значение функции.

Частная производная второго порядка — это частная производная, взятая от частной производной исходной функции.

У функции двух переменных будет две частные производные (первого порядка) и четыре производные второго порядка,

Обозначения в зависимости от того, в каком порядке шло дифференцирование:

- оба раза по переменной икс: z''_{xx} или $\dfrac{\partial^2 z}{\partial x^2}$;
- сначала по переменной икс, потом по игрек: z''_{xy} или $\dfrac{\partial^2 z}{\partial x \partial y}$;
- сначала по переменной игрек, потом по икс: z''_{yx} или $\dfrac{\partial^2 z}{\partial y \partial x}$;
- оба раза по переменной игрек: z''_{yy} или $\dfrac{\partial^2 z}{\partial y^2}$.

Частные производные z''_{xy} и z''_{yx} называются **смешанными производными**, и чаще всего совпадают (в нашем курсе они всегда будут равны).

Экстремумы функции нескольких переменных

ε-окрестность точки $A(x_0, y_0) \in x\mathrm{O}y$ — это круг радиуса $\varepsilon > 0$ с центром в данной точке A.

Круг отличается от окружности — это фигура, которая содержит и саму окружность, и все точки внутри неё.

Функция $z = f(x, y)$ имеет **(локальный) максимум** в точке $A(x_0, y_0) \in x\mathrm{O}y$, если для некоторой ε-окрестности точки A верно, что $f(x_0, y_0) > f(x, y)$.

Точка A называется **точкой максимума,** а значение функции в ней — **максимумом.**

Функция $z = f(x, y)$ имеет **(локальный) минимум** в точке $A(x_0, y_0) \in x\mathrm{O}y$, если для некоторой ε-окрестности точки A верно, что $f(x, y) > f(x_0, y_0)$.

Точка A называется **точкой минимума,** а значение функции в ней — **минимумом.**

Проще говоря, максимум находится выше всех в некоторой круговой окрестности, а минимум — ниже всех.

Точки экстремума — общее название для точек минимума и максимума, в случае функции двух переменных это точки плоскости $x\mathrm{O}y$.

Экстремумы — общее название минимумов и максимумов. Это про значение функции, и в случае функции двух переменных — про координату z.

Помимо локальных экстремумов бывают ещё и **глобальные** — те, которые выше или ниже вообще всех остальных точек из области определения. Если в формулировке нет уточнения, то имеются в виду локальные экстремумы.

Необходимое условие экстремума: $N(x_0, y_0)$ — точка возможного экстремума функции $z = f(x, y)$, если

$$\begin{cases} z'_x(x_0, y_0) = 0 \text{ или не существует,} \\ z'_y(x_0, y_0) = 0 \text{ или не существует.} \end{cases}$$

Без выполнения этого условия точка точно не будет экстремумом. Если условие выполняется, то эта точка будет кандидатом на точку экстремума.

Если хотя бы какая-то из первых частных производных функции никогда не равна нулю, то система не будет иметь решений и у функции не будет даже кандидатов на экстремум, и как следствие — самих экстремумов.

Обозначения:

$$A = \frac{\partial^2 z}{\partial x^2}(x_0, y_0) = z''_{xx}(x_0, y_0);$$

$$B = \frac{\partial^2 z}{\partial x \partial y}(x_0, y_0) = z''_{xy}(x_0, y_0);$$

$$C = \frac{\partial^2 z}{\partial y^2}(x_0, y_0) = z''_{yy}(x_0, y_0);$$

$$\Delta = AC - B^2.$$

Δ-тест: Пусть точка $N(x_0, y_0)$ удовлетворяет необходимому условию существования экстремума (то есть обе первые производные в ней равны нулю). Тогда:

- если $\Delta > 0$, то функция имеет экстремум в точке N, причём при $A < 0$ это **максимум**, а при $A > 0$ — **минимум**;
- если $\Delta < 0$, то экстремума в точке N нет;
- если $\Delta = 0$, то вывод о наличии экстремума сделать нельзя и нужны дополнительные исследования.

Здесь первый пункт является **достаточным условием** экстремума в точке. То есть если он выполнен, то кандидат является точкой экстремума.

Алгоритм поиска экстремальных точек у функции двух переменных в большинстве случаев выглядит так:

1. Решаем систему $\begin{cases} z'_x(x_0, y_0) = 0, \\ z'_y(x_0, y_0) = 0. \end{cases}$

2. Проверяем, какие из кандидатов являются реальными точками экстремума: вычисляем $\Delta = AC - B^2$, и по Δ-тесту делаем вывод о точке экстремума и её типе.

Задача оптимизации и градиентный спуск

Градиент функции f — вектор, состоящий из частных производных: $\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$. Обозначается как ∇f.

Чтобы найти градиент функции в точке (x_0, y_0), нужно рассчитать в ней значения частных производных:

$$\nabla f = \left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0)\right).$$

Антиградиент — вектор с направлением, противоположным градиенту.

Направление наискорейшего роста функции — по градиенту. Направление, по которому функция быстрее всего убывает, — по антиградиенту.

Алгоритм нахождения точек минимума функции с помощью градиентного спуска:

1) Выбрать начальную точку.

2) Сдвинуться из этой точки против направления градиента. Получить новую точку.

3) Повторять второй шаг, пока не подойдём к минимуму достаточно близко.

Начальную точку обычно выбирают случайно. Иногда по задаче понятно, какую точку лучше взять начальной.

Координаты новой точки вычисляются по формуле: $(x_{\text{new}}, y_{\text{new}}) = (x_{\text{old}}, y_{\text{old}}) - \nabla f(x_{\text{old}}, y_{\text{old}})$.

Метод градиентного спуска используется, чтобы найти точки минимума функции, когда найти нули частных производных и применить Δ-тест сложно.

Реализация градиентного спуска в Python

При реализации градиентного спуска при подсчёте новой точки размер шага нужно изменять: $x^k = x^{k-1} - \nabla f(x^{k-1}) \cdot \gamma$.

Коэффициент скорости спуска γ подбирают отдельно под каждую функцию так, чтобы значение функции уменьшалось, причём довольно быстро. Популярный рецепт подбора — перебирать отрицательные степени десятки.

Критерий остановки метода: если изменение значения функции на последнем шаге меньше заданной точности, то считается, что мы нашли минимум и нужно остановить вычисления.

Формальная запись критерия остановки: $|f(x_{\text{new}}) - f(x_{\text{old}})| < \epsilon$, где: x_{old} — точка до шага, x_{new} — точка после шага, ϵ — заданная точность.

Циклы — это такие выражения, которые позволяют запускать одну и ту же часть кода, пока не удовлетворится какое-то условие.

Работа цикла `while` на примере программы, которая выводит числа от 1 до 3 :

```
n = 1

while n < 4:
 print(n)
 n = n + 1
```

Конструкция выполняет кусок кода под словом `while` до тех пор, пока не выполнится указанное условие. Условие возле слова `while` называется **условием выхода**. Кусок кода, который выполняется много раз, называется **телом цикла.** Один проход по телу цикла называется **итерацией**.

Чтобы застраховаться от бесконечных циклов, можно искусственно ограничивать количество итераций с помощью дополнительного условия выхода, например: `while 2 < 3 and i < 10`.

Градиентный спуск реализуется в Python с помощью цикла `while` : условие выхода — это критерий остановки, тело цикла — подсчёт координат новой точки по формуле $x^k = x^{k-1} - \nabla f(x^{k-1}) \cdot \gamma$.