Функции многих переменных

1. Частные производные и градиент

Понятие частной производной

Частная производная — это одно из обобщений понятия производной на случай функции нескольких переменных.

Частная производная функции $f(x, y)$ по x определяется как производная по x, взятая в смысле функции одной переменной, при условии постоянства оставшейся переменной y. Таким образом:

$$f'_x(x,y) = \lim_{h\to 0} \frac{f(x+h,y) - f(x,y)}{h}, \qquad f'_y(x,y) = \lim_{h\to 0} \frac{f(x,y+h) - f(x,y)}{h}.$$

Геометрический смысл частной производной

Пусть дана некоторая функция двух переменных $f : \mathbb{R}^2 \to \mathbb{R}$. Она, вообще говоря, определяет некоторую поверхность $z = f(x, y)$ в трехмерном пространстве.

Если в некоторой точке (x_0, y_0) функция дифференцируема как функция многих переменных, то в этой точке можно рассмотреть плоскость касательную к рассматриваемой поверхности.

Эта касательная плоскость пересекает координатные плоскости $x0z$ и $y0z$ по прямым, тангенсы угла между которыми и соответствующими координатными осями равны значениям частных производных в точке (x_0, y_0).

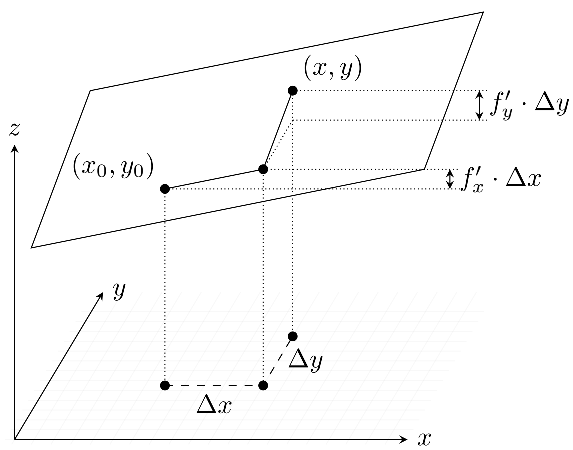

Рис. 1: Геометрический смысл частных производных.

Таким образом, график функции $f(x, y)$ в окрестности точки можно приблизить касательной плоскостью:

$$f(x_0 + \Delta x, y_0 + \Delta y) \approx f(x_0, y_0) + f'_x(x_0, y_0)\Delta x + f'_y(x_0, y_0)\Delta y.$$

Градиент

Если $f(x_1, \ldots, x_n)$ — функция n переменных $x_1, ..., x_n$, то n-мерный вектор из частных производных:

$$\operatorname{grad} f = \left(\frac{\partial f}{\partial x_1}, \ \ldots, \ \frac{\partial f}{\partial x_n} \right)$$

называется градиентом.

Линией уровня функции называется множество точек, в которых функция принимает одно и то же фиксированное значение. Оказывается, что градиент перпендикулярен линии уровня. Более подробное обсуждение этого факта будет произведено позднее.

2. Градиент в задачах оптимизации

Задачей оптимизации называется задача по нахождению экстремума функции, например минимума:

$$f(x_1, ..., x_n) \to \min .$$

Такая задача часто встречается в приложениях, например при выборе оптимальных параметров рекламной компании, а также в задачах классификации.

Если функция дифференцируема, то найти точки, подозрительные на экстремум, можно с помощью необходимого условия экстремума: все частные производные должны равняться нулю, а значит вектор градиента — нулевому вектору.

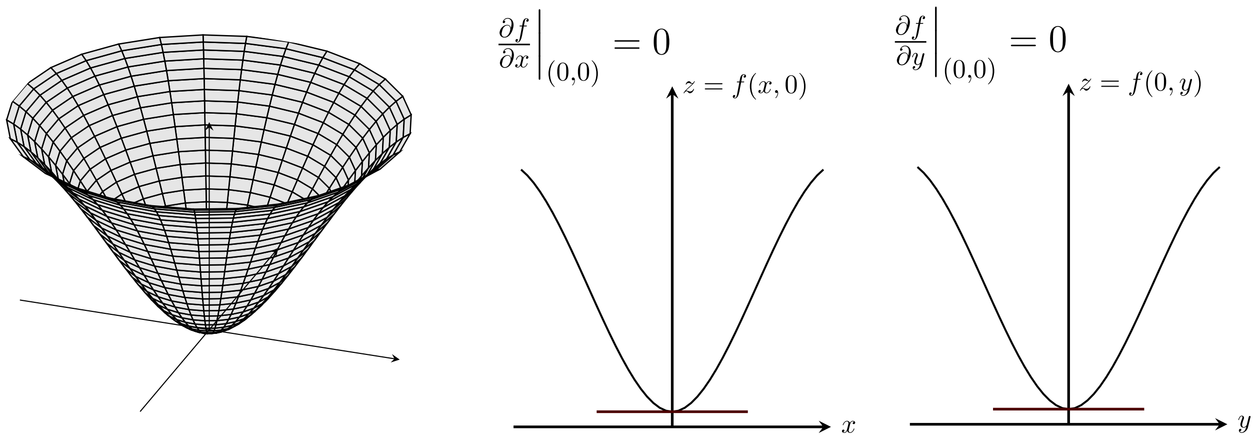

Рис. 2: Функция двух переменных достигает минимума в начале координат.

Но не всегда задачу можно решать аналитически. В таком случае используется численная оптимизация.

Наиболее простым в реализации из всех методов численной оптимизации является метод градиентного спуска. Это итерационный метод. Решение задачи начинается с выбора начального приближения $\vec{x}^{[0]}$. После вычисляется приблизительное значение $\vec{x}^1$, а затем $\vec{x}^2$ и так далее, согласно итерационной формуле:

$$\vec{x}^{[j+1]} = \vec{x}^{[j]} - \gamma^{[j]} \nabla F(\vec{x}^{[j]}), \qquad \text{где } \gamma^{[j]} \text{ — шаг градиентного спуска.}$$

Основная идея метода заключается в том, чтобы идти в направлении наискорейшего спуска, а это направление задаётся антиградиентом $-\nabla F$.

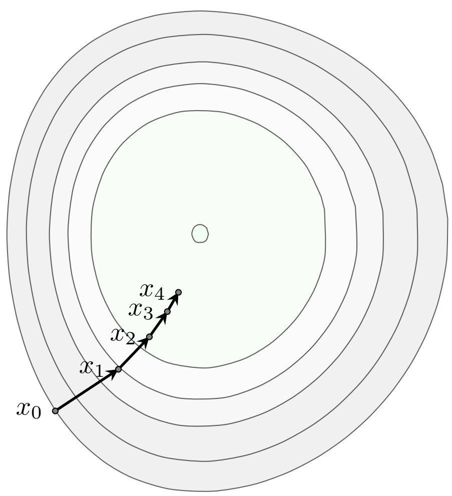

Рис. 3: Градиентный спуск

Можно привести следующую аналогию: Представьте, что вы приехали в незнакомое место и поселились где-то в низине. Гуляя по недалеким окрестностям, вы заблудились и теперь хотите попасть домой. Если вы не ориентируетесь в местности, логичным решением будет идти в каждый момент времени в направлении наискорейшего спуска, чтобы вернуться в свой дом.

3. Производная по направлению

Пусть $f(\vec{x}) = f(x_1, x_2, \ldots, x_n)$ — функция n переменных, $\vec{\ell} \in \mathbb{R}^n$, $|\vec{\ell}| = 1$, тогда частной производной в точке x_0 по направлению $\vec{\ell}$ называется

$$\frac{\partial f}{\partial \vec{\ell}}(\vec{x}_0) = \lim_{t \to 0} \frac{f(\vec{x}_0 + t \cdot \vec{\ell}) - f(\vec{x}_0)}{t}.$$

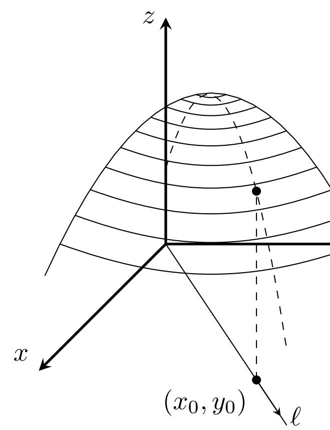

Рис. 4: К определению производной по направлению функции двух аргументов.

Непосредственно из определения становится понятен геометрический смысл производной по направлению. Производная по направлению показывает, насколько быстро функция изменяется при движении вдоль заданного направления. Производная по направлению координатной оси является частной производной по соответствующей координате.

4. Касательная плоскость и линейное приближение

Пусть функция $f(x, y)$ дифференцируема в точке (x_0, y_0), тогда в окрестности этой точки можно записать:

$$f(x_0 + \Delta x, y_0 + \Delta y) \approx f(x_0, y_0) + f'_x(x_0, y_0)\Delta x + f'_y(x_0, y_0)\Delta y.$$

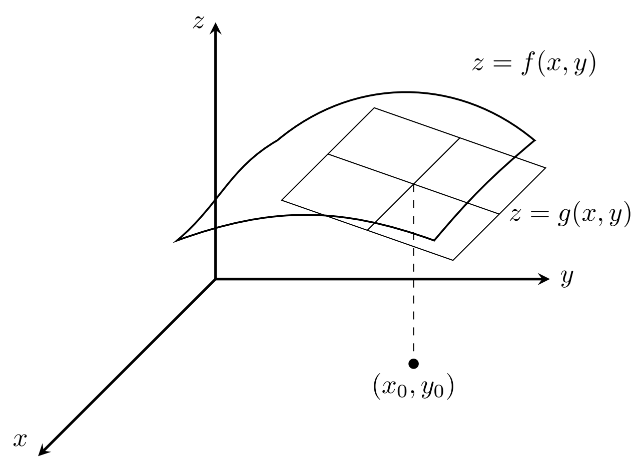

Рис. 5: Касательная плоскость к функции двух переменных.

Это выражение может быть переписано более удобных обозначениях:

$$\Delta f \approx \left\langle \nabla f(x_0, y_0), \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right\rangle.$$

Это есть не что иное, как уравнение касательной плоскости, записанное через скалярное произведение. Следует отметить, что линейное приближение тем лучше описывает поведение функции, чем ближе к точке (x_0, y_0) находится интересующее значение.

Линейное приближение часто используется для анализа сложных функций, в частности, при построении алгоритмов анализа данных.

5. Направление наискорейшего роста

Если функция $f(x, y)$ дифференцируема в точке (x_0, y_0), то в окрестности этой точки для функции можно воспользоваться линейным приближением:

$$\Delta f \approx \left\langle \nabla f(x_0, y_0), \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right\rangle.$$

Пусть $\vec{\ell} \in \mathbb{R}^2$, $|\vec{\ell}| = 1$, тогда приращения можно задать вдоль вектора $\vec{\ell}$:

$$\Delta x = t \cdot \ell_x, \qquad \Delta y = t \cdot \ell_y.$$

Если теперь подставить эти приращения в выражение для линейного приближения:

$$\Delta f \approx \left\langle \nabla f(x_0, y_0), \begin{pmatrix} t \cdot \ell_x \\ t \cdot \ell_y \end{pmatrix} \right\rangle = t \cdot \left\langle \nabla f(x_0, y_0), \vec{\ell} \right\rangle,$$

можно получить связь производной по направлению и градиентом:

$$\frac{\partial f}{\partial \vec{\ell}}(x_0, y_0) = \lim_{t \to 0} \frac{\Delta f}{t} = \left\langle \nabla f(x_0, y_0), \vec{\ell} \right\rangle$$

Таким образом, производная по направлению может быть вычислена как скалярное произведение градиента на соответствующий единичный вектор:

$$\frac{\partial f}{\partial \vec{\ell}}(x_0, y_0) = \left\langle \nabla f(x_0, y_0), \vec{\ell} \right\rangle.$$

Согласно данной формуле, направление максимального роста — это направление задаваемое градиентом. Действительно, скалярное произведение $\left\langle \nabla f(x_0, y_0), \vec{\ell} \right\rangle$ максимально в том случае, когда векторы $\nabla f(x_0, y_0)$ и $\vec{\ell}$ сонаправлены.