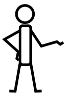
When my horse is running good,
I don't stop to give him sugar.
William Faulkner

МА-91. *ПРИРАЩЕНИЯ И ДИФФЕРЕНЦИАЛЫ*

§1. Приращение функции

Роль частных производных выявляется, когда речь заходит о *приращении функции*



$$\triangle f = f(\boldsymbol{x} + \triangle \boldsymbol{x}) - f(\boldsymbol{x}),$$

что, с точки зрения наглядности, удобнее рассмотреть на примере двух переменных, $z = f(x, y)$. Перепишем для этого случая приращение

$$\triangle z = f(x + \triangle x, y + \triangle y) - f(x, y).$$

в виде

$$\triangle z = \frac{f(x + \triangle x, y + \triangle y) - f(x, y + \triangle y)}{\triangle x} \triangle x + \frac{f(x, y + \triangle y) - f(x, y)}{\triangle y} \triangle y,$$

откуда следует

$$\triangle z = f'_x(x, y + \triangle y)\triangle x + o(\triangle x) + f'_y(x, y)\triangle y + o(\triangle y),$$

и в итоге

$$\boxed{\triangle z = \frac{\partial f(x, y)}{\partial x} \triangle x + \frac{\partial f(x, y)}{\partial y} \triangle y + o(\|(x, y)\|)},$$

но только в предположении $f'_x(x, y + \triangle y) \to f'_x(x, y)$ при $\triangle y \to 0$, что обеспечивается, например, *непрерывностью* производной $\dfrac{\partial f}{\partial x}$ в окрестности (x, y).

Проделанное указывает на справедливость следующего результата.

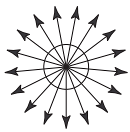
• ***Теорема.*** *Если функция $u = f(x)$ имеет непрерывные в окрестности точки $x = \{x_1, \ldots, x_n\}$ частные производные, то справедлива формула приращения*

$$\triangle z = \sum_{i=1}^{n} f'_{x_i}(x) \triangle x_i + o(\|\triangle x\|). \tag{1}$$

Непрерывность частных производных здесь не обязательна[1]. Однако просто наличия производных, которой хватает[2] в одномерном случае, тут недостаточно.

Как и в скалярном случае, линейную часть (1) записывают в форме

$$dz = \sum_{i=1}^{n} f'_{x_i}(\boldsymbol{x}) dx_i, \tag{2}$$



и называют ***полным дифференциалом*** функции $f(\boldsymbol{x})$. При этом $f'_{x_i}(x)dx_i$ называют *частными дифференциалами.*

[1]Но вполне приемлема, ибо на практике чаще всего имеет место.
[2]Для справедливости формулы приращения.

§2. Производные и дифференциалы высших порядков

Снова дифференцируя частные производные, имеем частные производные более высокого порядка. Например, в случае $z = xy^2$

$$\frac{\partial^2 z}{\partial x^2} = \frac{\partial}{\partial x}\frac{\partial z}{\partial x} = 0, \qquad \frac{\partial^2 z}{\partial y^2} = \frac{\partial}{\partial y}\frac{\partial z}{\partial y} = 2x.$$

Аналогично

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial}{\partial x}\frac{\partial z}{\partial y} = 2y,$$

причём в данном случае оказывается

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x}. \tag{3}$$

Совпадение смешанных производных (3) — явление характерное, хотя и необязательное. Равенство (3) по своей природе является *равенством повторных пределов*. Справедлив следующий общий результат.

• *Если функция $z = f(x_1, \ldots, x_n)$ имеет непрерывные производные до m-го порядка включительно, то любая смешанная производная*

$$\frac{\partial^m z}{\partial x_1^{m_1} \ldots \partial x_n^{m_n}},$$

не зависит от порядка дифференцирования.

Формальное обоснование этого результата может оказаться громоздким и неуклюжим. Здесь проще всего ограничиться доказательством (3), в качестве упражнения, из чего обычно становится ясна справедливость общего утверждения.

Время от времени полезно задумываться о сути используемых и наблюдаемых доказательств. Строго говоря, они являются лишь эмоциональной аргументацией. Степень подробности и траектория рассуждения всегда зависит от того, с чем «противная сторона» готова согласиться. Доказывая, мы оперируем смысловыми категориями. При этом логические прорехи очень трудно контролировать. «Имея семейство множеств, в каждом множестве можно выбрать по элементу» — не самоочевидный ли тезис? Вышло — нет[3]. Так что до синтаксического уровня аргументации человеку никогда не добраться. Хотя некоторые авторы пытаются увлечь население в эту «чёрную дыру».

---

[3]См. *аксиому выбора.*

Дифференциалы высших порядков идут «по тому же следу». Дифференциал от дифференциала даёт второй дифференциал:

$$d^2 f = d(df) = d\left(\sum_i \frac{\partial f}{\partial x_i} dx_i\right) = \sum_i d\left(\frac{\partial f}{\partial x_i}\right) dx_i,$$

т. е.

$$d^2 f = \sum_i \frac{\partial^2 f}{\partial x_i^2} dx_i^2 + 2\sum_{i \neq j} \frac{\partial^2 f}{\partial x_i \partial x_j} dx_i dx_j.$$

Продолжая процесс, получаем

$$d^m f = \sum \frac{\partial^m f}{\partial x_1^{m_1} \dots \partial x_n^{m_n}} dx_1^{m_1} \dots dx_n^{m_n}, \tag{4}$$

где суммирование ведется по всем $m_1 + \dots + m_n = m$.

Что касается разложения Тейлора

$$\triangle f = \sum_{m=1}^{k} \frac{1}{m!} \sum_\sigma \frac{\partial^m f}{\partial x_1^{m_1} \dots \partial x_n^{m_n}} \triangle x_1^{m_1} \dots \triangle x_n^{m_n} + o(\|\triangle x\|^k),$$

то оно естественным образом перекликается с (4), и может быть получено *разложением Тейлора* в нуле функции одной переменной $\varphi(\tau) = f(x + \tau\triangle v)$, где $\triangle x = \tau\triangle v$.

§3. Градиент

В многомерном анализе важную роль играют понятия *градиента* и *производной по направлению*.

Производная функции $z = f(x, y)$ по направлению *единичного вектора* $\boldsymbol{s} = \{s_x, s_y\}$ определяется как обычная производная по θ:

$$z'_s = z'_\theta = \frac{d}{d\theta} f(x + \theta s_x, y + \theta s_y) = \frac{\partial f}{\partial x} s_x + \frac{\partial f}{\partial y} s_y. \tag{5}$$

Правую часть (5) естественно интерпретировать как скалярное произведение вектора $\boldsymbol{s} = \{s_x, s_y\}$ на вектор

$$\mathrm{grad} f = \left\{\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right\},$$

который называют *градиентом* функции f. Для градиента используют также обозначение ∇f (читается «набла эф»). Таким образом

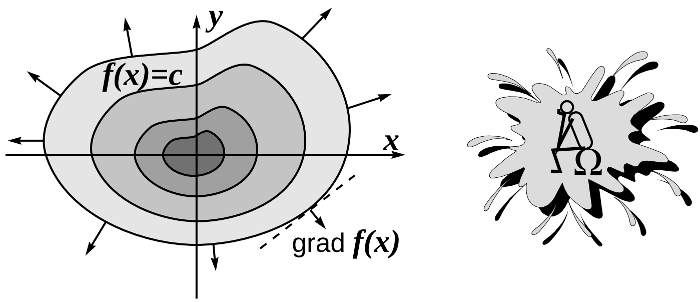
$$z'_s = \boldsymbol{s} \cdot \nabla f. \tag{6}$$

В силу

$$\boldsymbol{s} \cdot \operatorname{grad} f = \|\operatorname{grad} f\| \|\boldsymbol{s}\| \cdot \cos\varphi$$

максимум в (6) достигается при $\varphi = 0$, т. е. когда вектор $\boldsymbol{s}$ совпадает по направлению с градиентом. При этом значение максимума равно модулю градиента $\|\nabla f\|$, поскольку $\|\boldsymbol{s}\| = 1$. Следовательно, градиент ∇f — это ***вектор скорости максимального роста функции*** f.

Производная (6) обращается в нуль, когда $\boldsymbol{s}$ и $\|\nabla f\|$ ортогональны друг другу, что имеет место, если $\boldsymbol{s}$ направлен вдоль касательной к линии постоянного уровня[4] $f(x, y)$. Стало быть, в каждой точке поверхности $f(x, y) = \text{const}$ градиент ∇f перпендикулярен этой поверхности, направлен по нормали.



(7)

Таким образом, касательная плоскость в точке $\{x_0, y_0\}$ к линии постоянного уровня функции $z = f(x, y)$, описывается уравнением[5]

$$\frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0) = 0, \tag{8}$$

[4]Или лежит в касательной плоскости к поверхности постоянного уровня $f(\boldsymbol{x})$, если размерность аргумента $\geqslant 3$.

[5]На рис. (7) это пунктирная прямая.

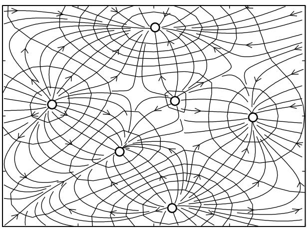
а *касательная плоскость* к поверхности $z = f(x, y)$ в точке $\{z_0, x_0, y_0\}$ определяется уравнением

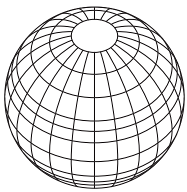
$$z - z_0 = \frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0). \tag{9}$$

Уравнение (9) является непосредственным аналогом (8), поскольку касательная плоскость к графику (9) — это касательная плоскость к поверхности постоянного уровня функции трёх переменных $u = z - f(x, y)$, градиент которой

$$\left\{1, -\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}\right\}.$$

Двумерный случай (аргумент лежит в плоскости), конечно, в силу своей наглядности должен служить первоисточником при изучении многомерного случая. Визуальное представление линий постоянного уровня и перпендикулярных им линий, вдоль которых направлены градиенты, помогает лучше усвоить геометрию внутреннего устройства функций $z = f(x, y)$, а там и до $z = f(x_1, \ldots, x_n)$ недалеко.





При этом благотворно для воображения поразмышлять над траекториями движения по градиенту

$$\dot{x} = \nabla f(x), \tag{10}$$

где

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

Один из источников интереса к системам типа (10) — численные процедуры поиска локальных максимумов функции $f(x)$.

Отталкиваясь от двумерного случая, легко записать касательные плоскости в общем случае $z = f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^n$. Уравнение касательной плоскости в $\mathbb{R}^n$ к поверхности постоянного уровня в точке x_0:

$$\nabla f(\boldsymbol{x}) \cdot (\boldsymbol{x} - \boldsymbol{x}_0) = 0.$$

Уравнение касательной плоскости в $\mathbb{R}^{n+1}$ к поверхности графика $z = f(x)$ в точке $\{z_0, \boldsymbol{x}_0\}$, $z_0 = f(\boldsymbol{x}_0)$:

$$z - z_0 = \nabla f(\boldsymbol{x}) \cdot (\boldsymbol{x} - \boldsymbol{x}_0).$$

Формула конечного приращения в векторных обозначениях приобретает вид

$$\triangle z = \nabla f(\boldsymbol{x}) \cdot \triangle \boldsymbol{x} + o(\|\triangle \boldsymbol{x}\|).$$

Применение *теоремы Лагранжа* к функции скалярного аргумента

$$\varphi(\tau) = f(\boldsymbol{x} + \tau\boldsymbol{(y - x)})$$

даёт

$$\varphi(1) - \varphi(0) = \varphi'(\theta) = \sum_i \frac{\partial f(z)}{\partial x_i}(y_i - x_i) = \nabla f(\boldsymbol{w}) \cdot \boldsymbol{(y - x)},$$

где $\boldsymbol{w = x} + \theta\boldsymbol{(y - x)}$ при некотором $\theta \in (0, 1)$. Поэтому

$$\boxed{f(\boldsymbol{y}) - f(\boldsymbol{x}) = \nabla f(\boldsymbol{w}) \cdot \boldsymbol{(y - x)}},$$

что называют *теоремой о среднем*.