

Производная и градиентный спуск



Функции. Производная. Экстремумы функции. Выпуклость функции.
Правила дифференцирования. Правила дифференцирования сложной
функции. Chain-rule. Функция нескольких аргументов. Градиент.
Градиентный спуск как метод оптимизации.

Даниил Корбут

Специалист по Анализу Данных



НЕТОЛОГИЯ



Даниил Корбут
DL Researcher
Insilico Medicine, Inc

Окончил бакалавриат ФИВТ
МФТИ (Анализ данных) в 2018г
Учусь на 2-м курсе
магистратуры ФИВТ МФТИ
Работал в Statsbot и Яндекс.
Алиса.
Сейчас в Insilico Medicine, Inc,
занимаюсь генерацией
активных молекул и
исследованиями старения с
помощью DL.

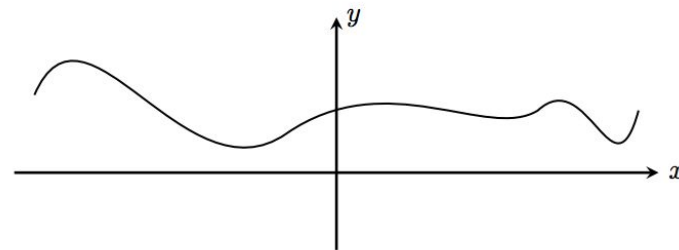
Функции и их свойства

Функция - это некоторое соответствие $x \rightarrow f(x)$, причём для каждого x определено единственное значение $f(x)$.

$D(f)$ - область определения функции

$E(f)$ - область значений функции

Будем работать только с функциями, у которых $D(f)$ и $E(f)$ - подмножество \mathbb{R} .



Функции и их свойства

Каковы область определения и область значений следующих функций?

1) $f(x) = 1 / (x-1)$

2) $f(x) = 2^x$

1) $D(f) = \mathbb{R} \setminus \{1\}, E(f) = \mathbb{R} \setminus \{0\}$

2) $D(f) = \mathbb{R}, E(f) = (0, +\infty)$

Функции и их свойства

Представление о функции, её свойствах и поведении можно получить, построив её график. Функции бывают непрерывными и разрывными.

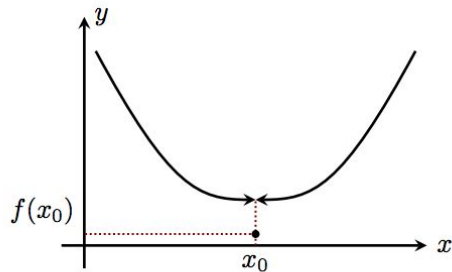


Рис. 2: Функция с устранимым разрывом

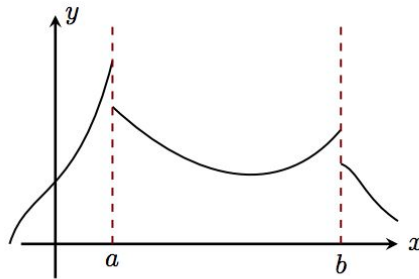


Рис. 3: Функция с разрывами в точках a и b .

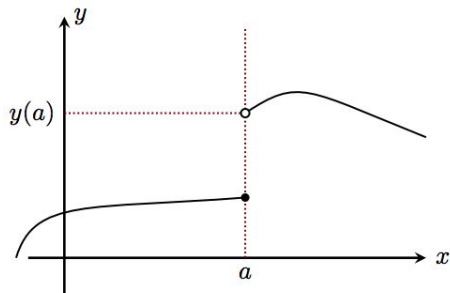


Рис. 4: Функция с разрывом типа «скачок».

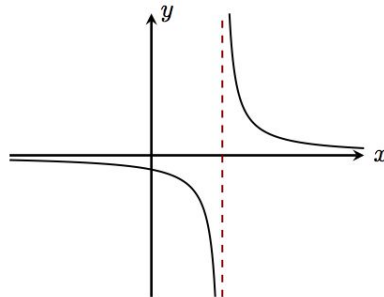


Рис. 5: Функция с бесконечным разрывом.

Предел функции

$$f(x) = (1 + x)^{\frac{1}{x}}$$

Функция не определена в $x=0$, но её значение может быть вычислено в точках сколь угодно близких к ней

x	0.1	0.01	0.001	0.0001	...
$f(x)$	2.593..	2.704..	2.716..	2.718..	...

$$\lim_{x \rightarrow 0} (1 + x)^{\frac{1}{x}} = e.$$

Не у всех функций есть конечный предел

$$g(x) = \frac{1}{x}$$

Функция неограниченно растёт при приближении к $x = 0$

x	0.1	0.01	0.001	0.0001	...
$1/x$	10	100	1000	10000	...

$$\lim_{x \rightarrow 0} \frac{1}{x} = \infty.$$

Предел функции

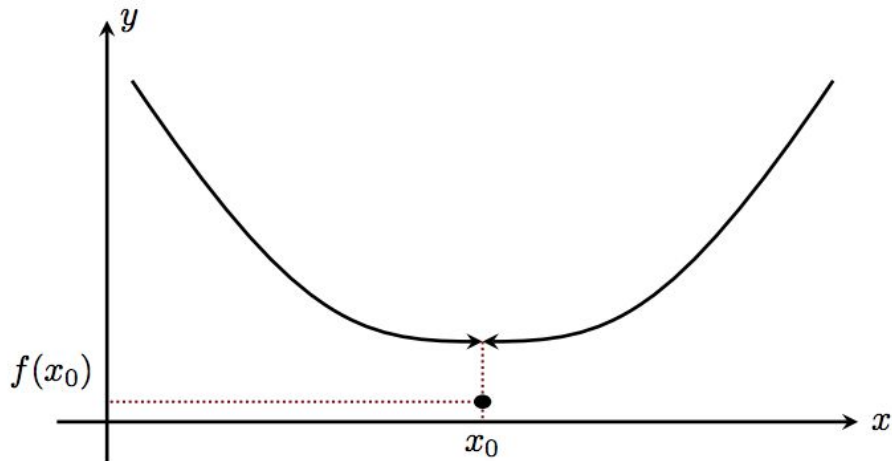
Понятие предела тесно связано с **понятием непрерывности** функции в точке.

Функция **непрерывна** в точке a , если:



$$\lim_{x \rightarrow a} f(x) = f(\lim_{x \rightarrow a} x) = f(a).$$

С помощью **понятия предела** определяется другое полезное понятие — **понятие производной**.



Производная функции

Производная - мгновенная скорость роста функции в заданной точке.

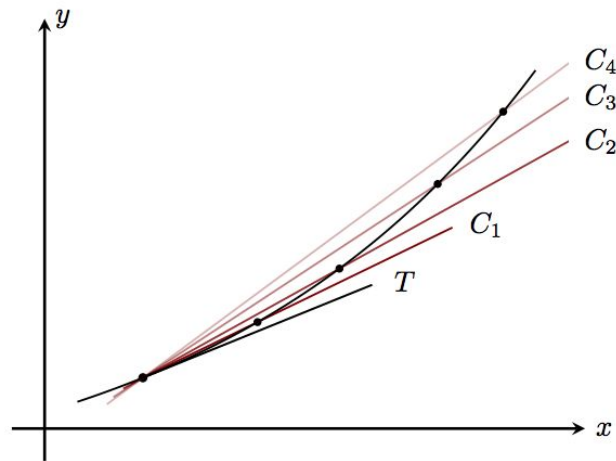
$$\frac{f(x + \Delta x) - f(x)}{\Delta x} = k.$$

Давайте посмотрим на линейную функцию $y=kx+b$

Как понять скорость роста для произвольной функции? **Предел!**

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

Гладкие функции - функции, производная которых непрерывна.



Производная сложной функции

Пусть имеются 2 функции $f(x)$ и $h(x)$, и область значений $f(x)$ принадлежит области определения $h(x)$. Тогда, $h(f(x))$ - применение одной функции к результату другой, называется **сложной функцией**.

Пример: $f(x) = x+1$, $h(x) = \ln(x)$, $g(x) = h(f(x)) = \ln(x+1)$

$$df = f'(x_0)dx, \quad dx = \Delta x.$$

Дифференциал - линейная часть приращения функции

$$f'(x_0) = \frac{df}{dx}(x_0).$$

Отсюда можно записать производную функцию через дифференциал

$$\frac{dg(h(x))}{dx} = \frac{dg(h(x))}{dh(x)} \frac{dh(x)}{dx} = \frac{dg(h)}{dh} \frac{dh(x)}{dx}$$

Производная сложной функции (пример)

$$\left(\frac{u(x)}{v(x)}\right)' = \frac{u(x)' \cdot v(x) - v(x)' \cdot u(x)}{v^2(x)}$$

$$(kf(x))' = kf'(x)$$

$$(u(x) \cdot v(x))' = u(x)' \cdot v(x) + v(x)' \cdot u(x)$$

$$(f + g)' = f' + g'$$

$$f(x) = \sin(\ln(x)+5x)$$

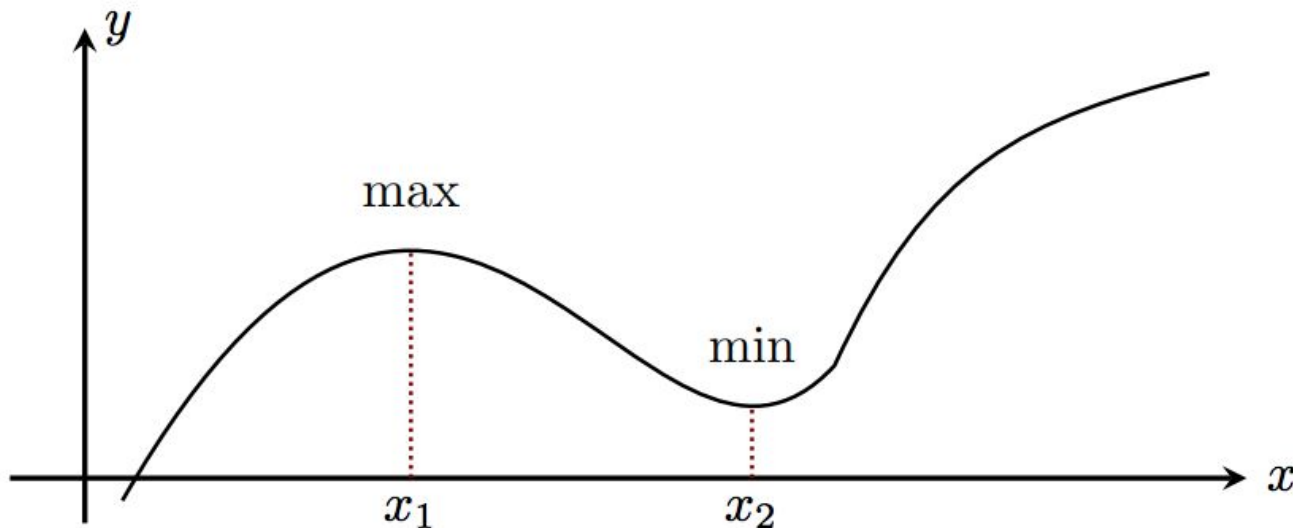
$$f'(x) = \cos(\ln(x)+5x) * (\ln(x)+5x)'$$

$$f'(x) = \cos(\ln(x)+5x) * (1/x + 5)$$

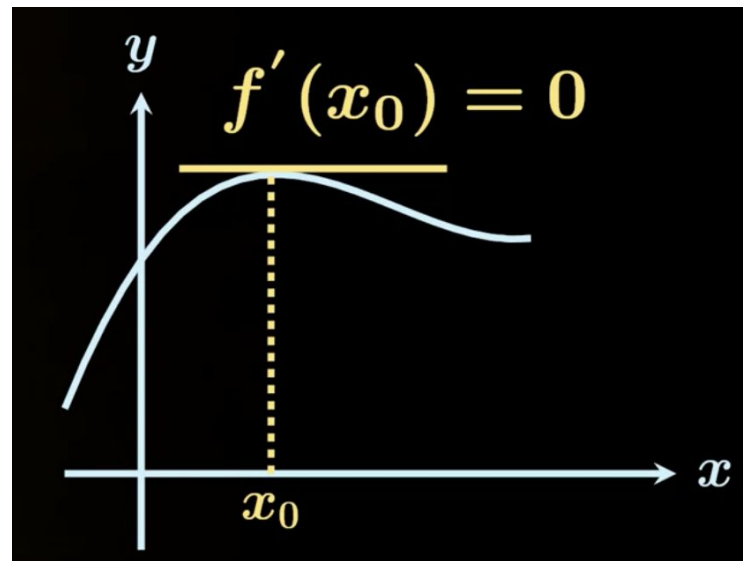
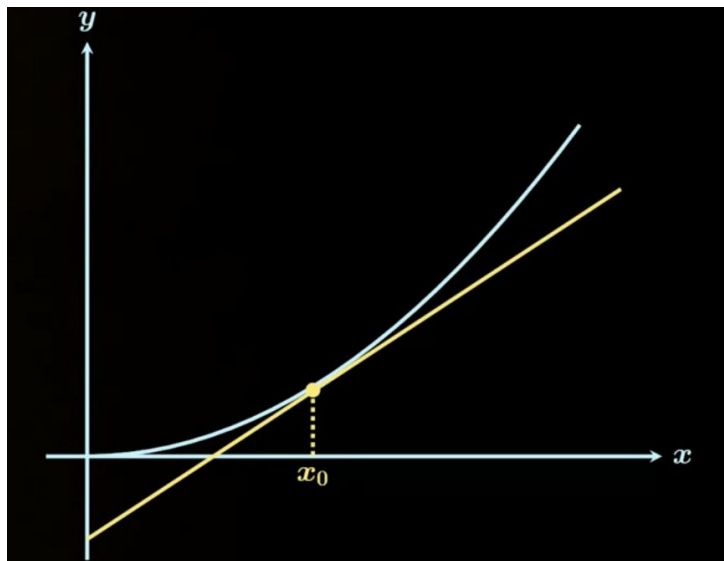
Экстремум функции



Точка x_0 - называется **локальным минимумом** функции $f(x)$, если существует такая окрестность $U(x_0)$, для которой $f(x) > f(x_0)$, x из $U(x_0)$. Аналогично для максимума. В случае глобального минимума $U(x_0) = D(f)$.

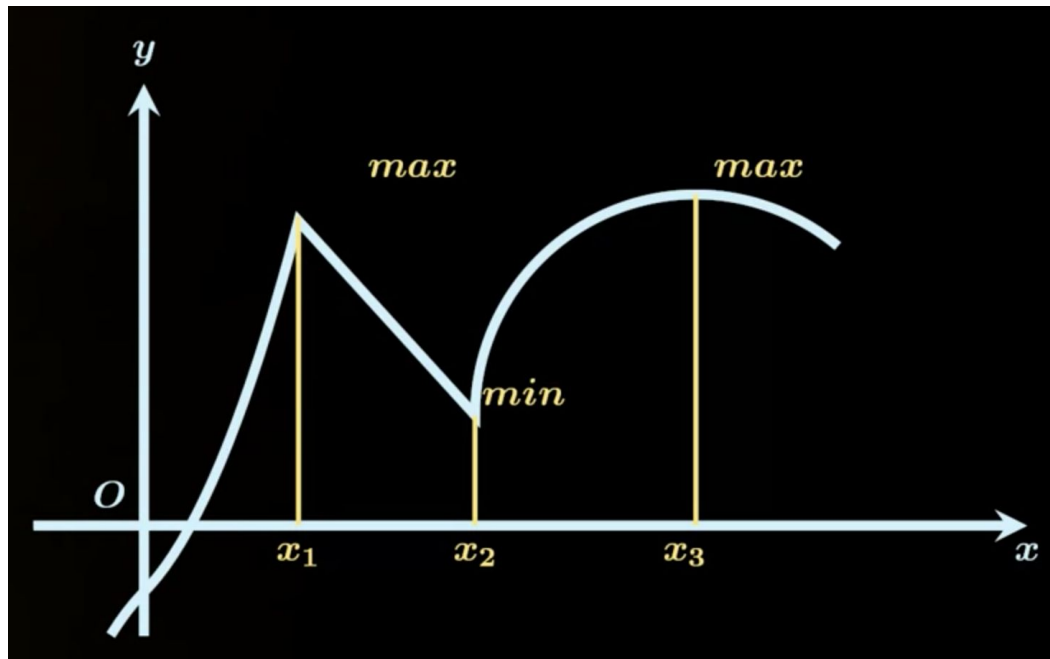
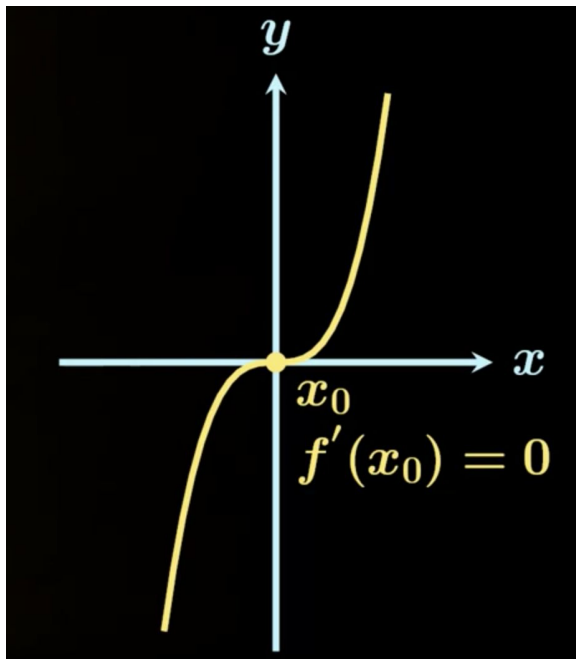


Экстремум функции и производная



В точках локальных экстремумов производная (если она определена (!), пример дальше) обязана равняться нулю.
Это **необходимое** условие.

Экстремум функции и производная

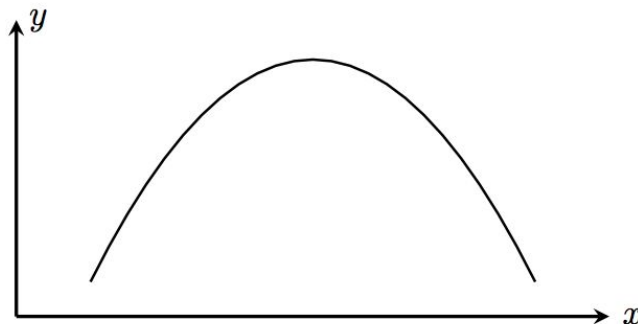
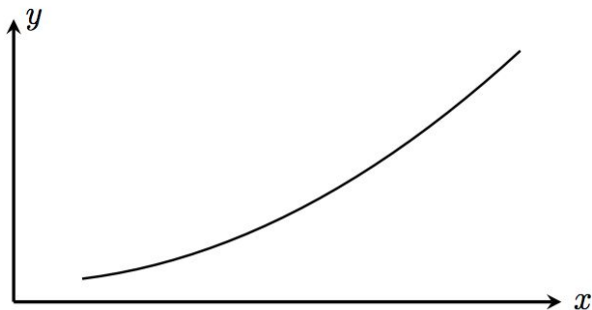


Однако равенство нулю производной **не является достаточным** условием локального экстремума. Также производная может быть вовсе не определена в точках локальных экстремумов.

Выпуклость функции и вторая производная

Как влияет знак производной на характер поведения функции?

1. $f'(x) \geq 0$ — функция возрастает,
2. $f'(x) > 0$ — функция строго возрастает,
3. $f'(x) \leq 0$ — функция убывает,
4. $f'(x) < 0$ — функция строго убывает.



Выпуклость функции и вторая производная

Давайте спустимся на уровень ниже: какому свойству функции соответствует монотонная производная?

1. $f''(x) \geq 0$ — функция $f(x)$ выпукла,
2. $f''(x) > 0$ — функция $f(x)$ строго выпукла,
3. $f''(x) \leq 0$ — функция $f(x)$ вогнута,
4. $f''(x) < 0$ — функция $f(x)$ строго вогнута.

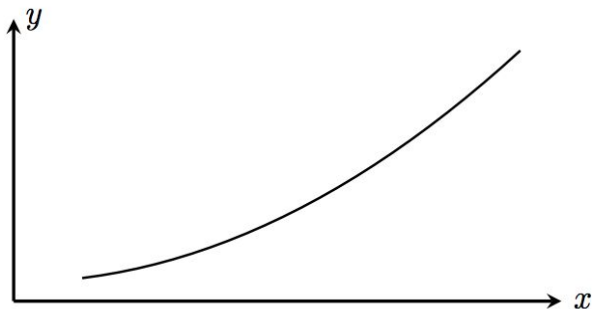


Рис. 11: Выпуклая функция

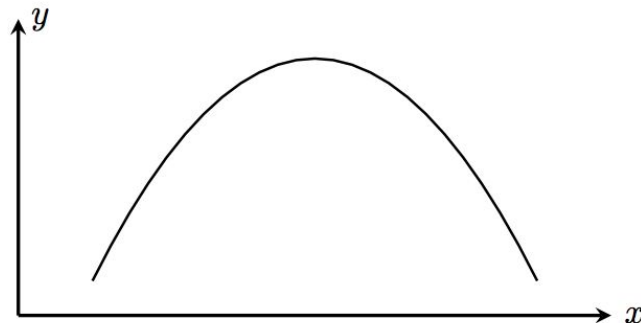


Рис. 12: Вогнутая функция

Выпуклость функции и вторая производная

Помните **необходимое** условие локального экстремума?



Наложив некоторые условия на вторую производную, можно сделать их **достаточными!**

Достаточное условие экстремума Пусть выполнено необходимое условие экстремума, то есть в некоторой точке x_0 значение $f'(x_0) = 0$. Если в таком случае

1. $f''(x) > 0$ — функция будет строго выпукла и реализуется строгий минимум.
2. $f''(x) < 0$ — функция будет строго вогнута и реализуется строгий максимум.

Функция нескольких переменных

Пусть теперь x - не число из R , а вектор (x_1, \dots, x_n) , где каждое x_i из R , а весь x из R^n .

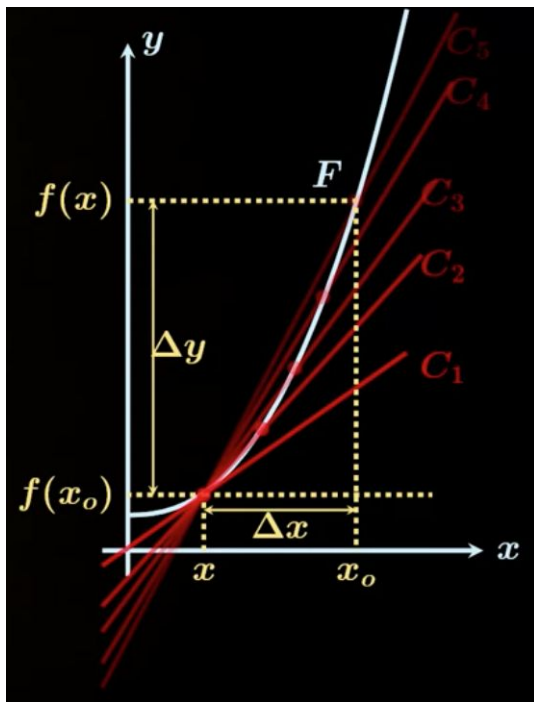
$D(f)$ - область определения функции (ничего не изменилось!)

$E(f)$ - область значений функции (ничего не изменилось!)

Снова будем работать только с функциями, у которых $D(f)$ и $E(f)$ - подмножество R^n .

Функция нескольких переменных

Из одномерного случая помним: геометрический смысл производной - угловой коэффициент касательной.



Как посчитать производную функции нескольких переменных?

$$\triangleright \frac{\Delta f}{\Delta x} \xrightarrow{\Delta x \rightarrow 0} f'_x, y \text{ — фиксирован}$$

$$\triangleright \frac{\Delta f}{\Delta y} \xrightarrow{\Delta y \rightarrow 0} f'_y, x \text{ — фиксирован}$$

Частная производная



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 .

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

Касательная плоскость

Пусть дана некоторая функция двух переменных $f: \mathbb{R}^2 \rightarrow \mathbb{R}$. 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 **касательную плоскость** к данной поверхности.

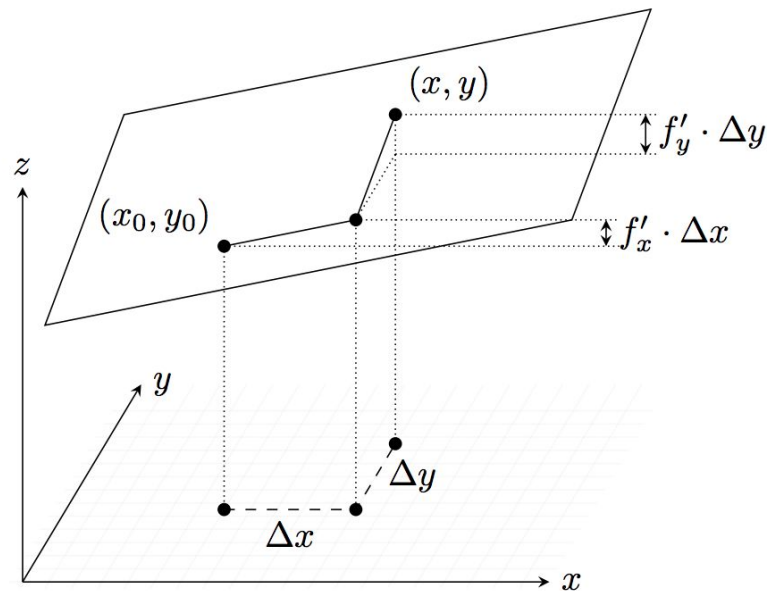
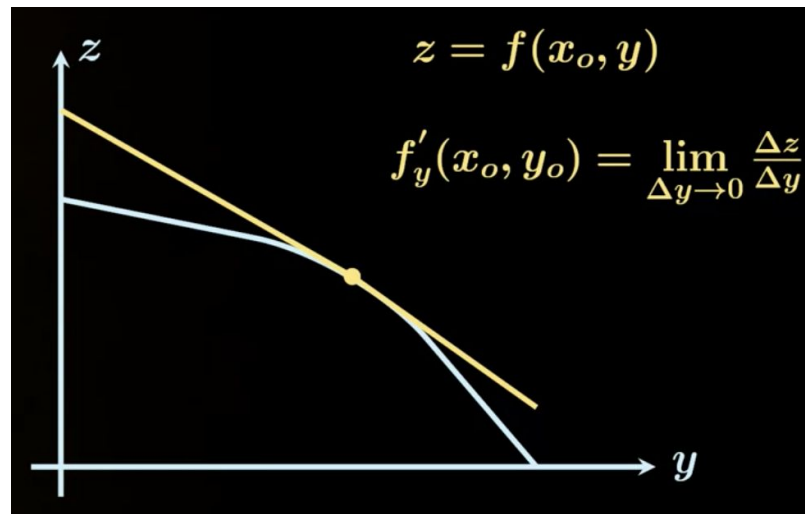
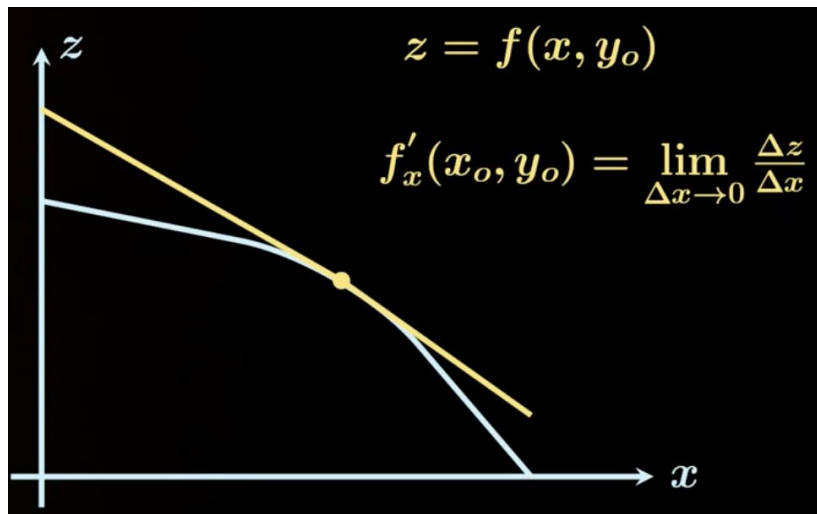


Рис. 1: Геометрический смысл частных производных.

Касательная плоскость

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 и линии уровня функции

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

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



называется **градиентом функции**.



Линией уровня функции называется множество точек, в которых функция принимает одно и то же фиксированное значение. Оказывается, что **градиент перпендикулярен линии уровня**.

Градиент в задачах оптимизации

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

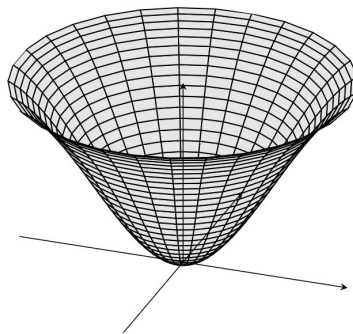
$$f(x_1, \dots, x_n) \rightarrow \min$$

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

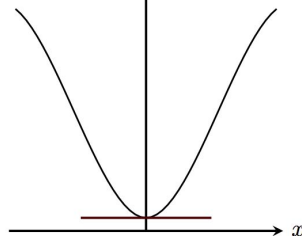
Вспомним необходимые условия экстремума из прошлой лекции!

Градиент в задачах оптимизации

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.



$$\left. \frac{\partial f}{\partial x} \right|_{(0,0)} = 0$$



$$\left. \frac{\partial f}{\partial y} \right|_{(0,0)} = 0$$

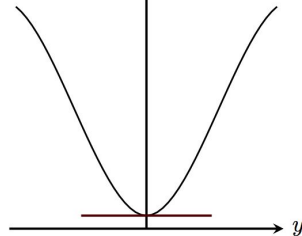


Рис. 2: Функция двух переменных достигает минимума в начале координат.

Градиентный спуск

Это итерационный метод. Решение задачи начинается с выбора начального приближения $\vec{x}^{[0]}$

После вычисляется приблизительное значение \vec{x}^1

Затем \vec{x}^2

и так далее...

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

Идея: идти в направлении наискорейшего спуска, а это направление задаётся антиградиентом $-\nabla F$.

Градиентный спуск

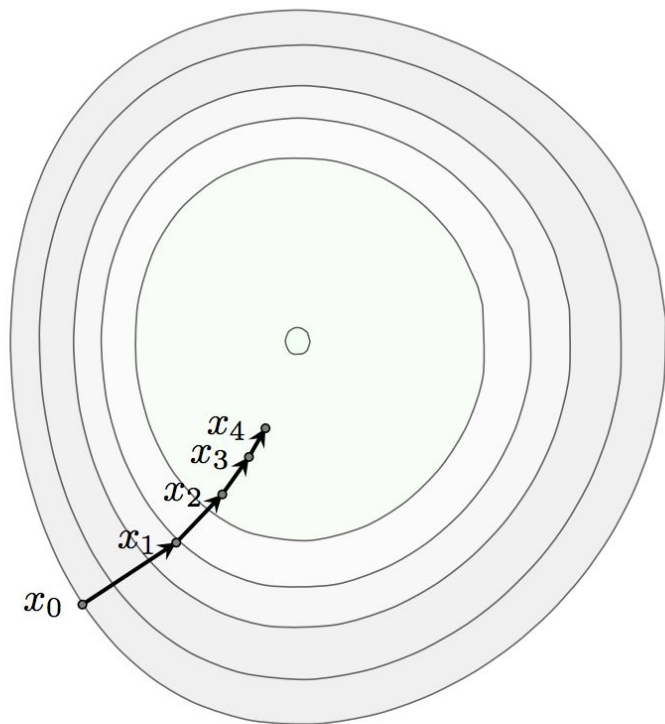


Рис. 3: Градиентный спуск

Аналогия: домик в низине

Если заблудились, то верным решением будет двигаться в направлении наискорейшего спуска

Спасибо за внимание!