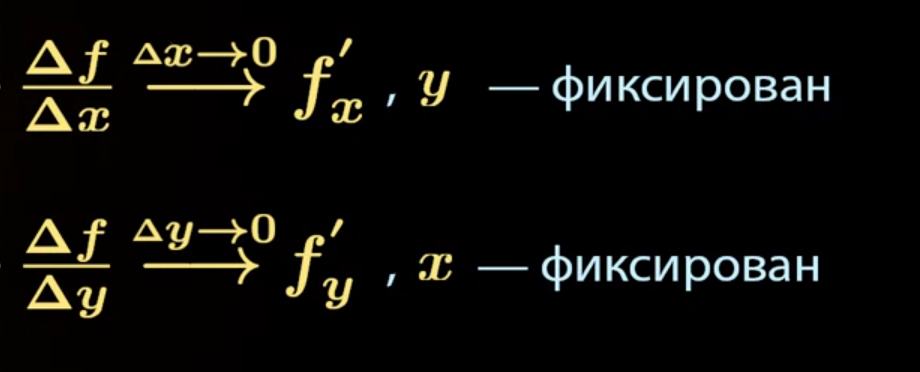
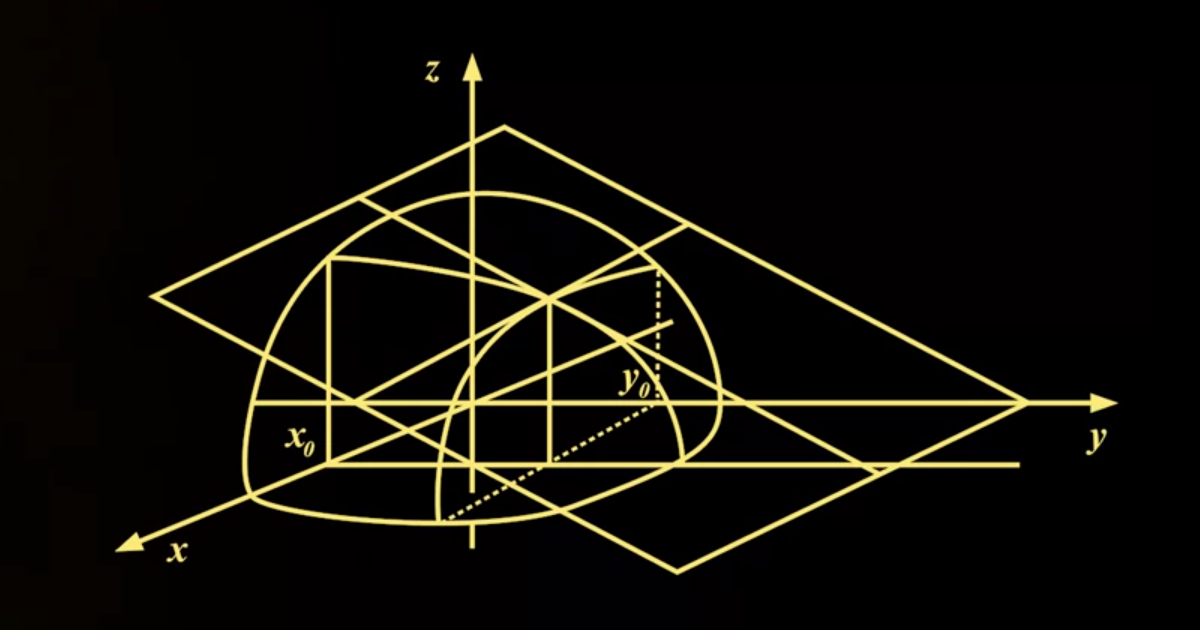
**Частные производные и градиент**

Зафиксируем y, продифференцируем по x, получим производную f'x. Зафиксируем x, продифференцируем по y и получим производную f'y.

Такие производные называются частными производными.



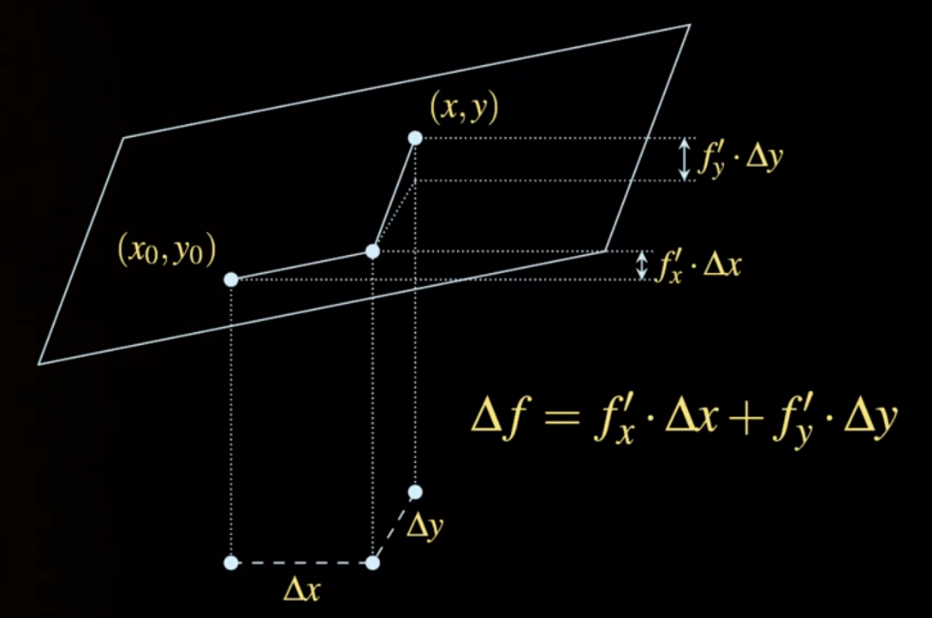
Если в случае функции одной переменной мы приближали график функции касательной, то в случае функции двух переменных, мы приближаем график целой касательной плоскостью.



Как же все это будет выглядеть в пространстве.  
Мы приближаем касательной плоскостью график функции от x и y. Будем считать, что при переходе от точки (x0, y0) к точке (x, y) у нас сначала x поменялся на дельта x, а затем y поменялся на дельта y.

Изменения координаты z вдоль касательной плоскости при

изменении x на дельта x будет равно f'x умножить на дельта x.

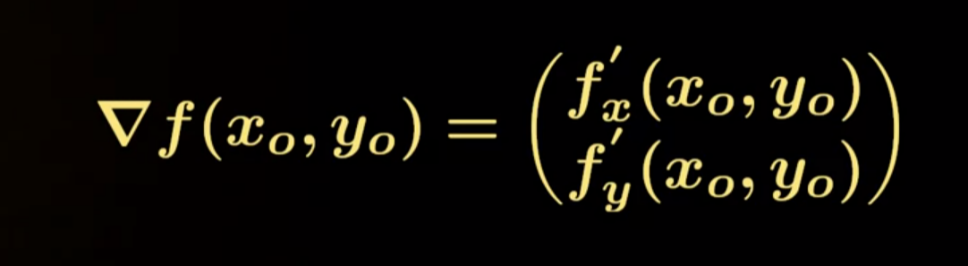


Изменение координаты z при изменении y на дельта y будет

равно f'y умножить на дельта y.

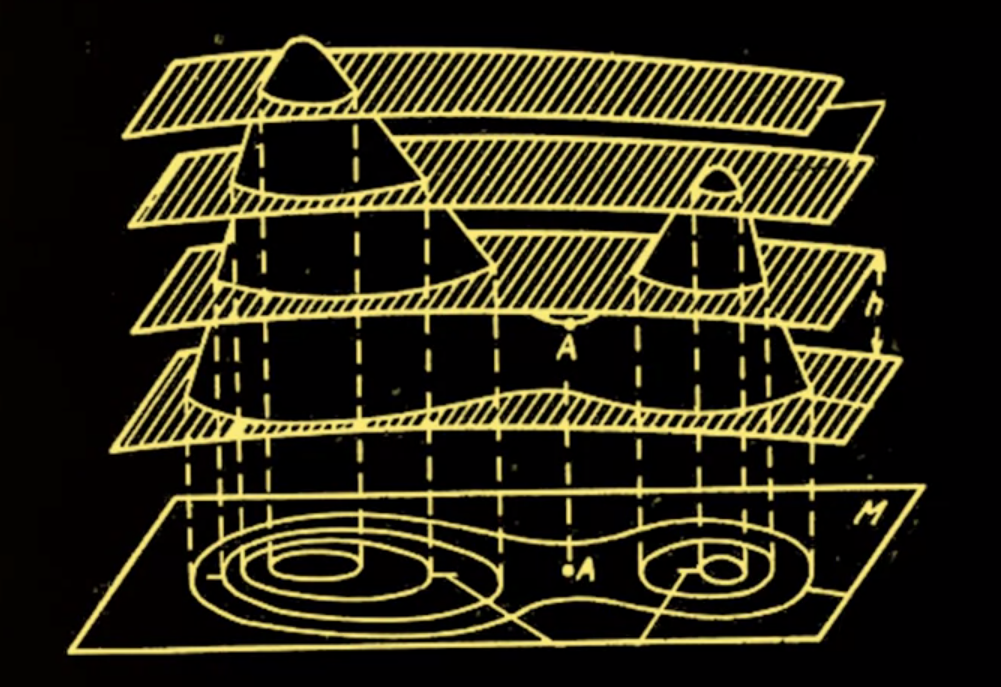
А сумма этих изменений даст итоговое изменение координаты z, которым мы и приближаем изменение функции дельта f.

**Градиент – градиентом** называют вектор построенный из частных производных функции. Такое вектор обладает интересными свойствами. Например, *он задает направление наискорейшего роста функции.*



Линии уровня — это линии, вдоль которых функция принимает какое-то фиксированное значение.

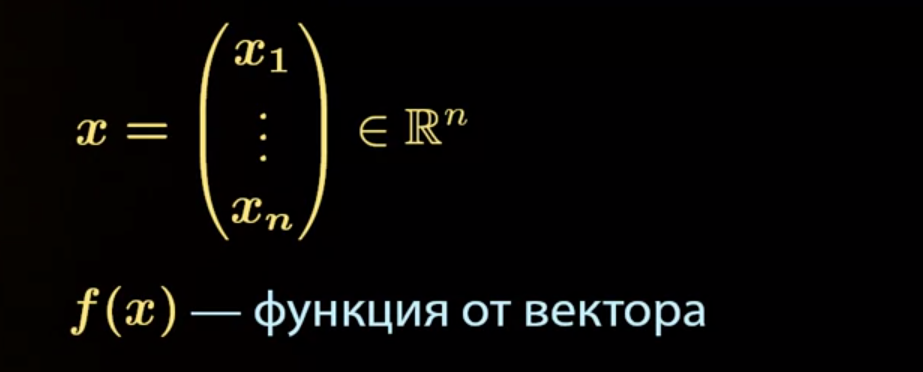
Таким образом, можно просто изображать эти линии на плоскости и получать какое-то представление о том, как меняется функция в разных точках.



Вот если изобразить градиент вместе с линиями уровня,

то градиент будет перпендикулярен линиям уровня.

В общем случае, когда мы работаем с функцией n переменных, мы можем считать, что функция просто зависит от вектора x из пространства Rn, то есть состоящего из координат x1, ..., xn. 



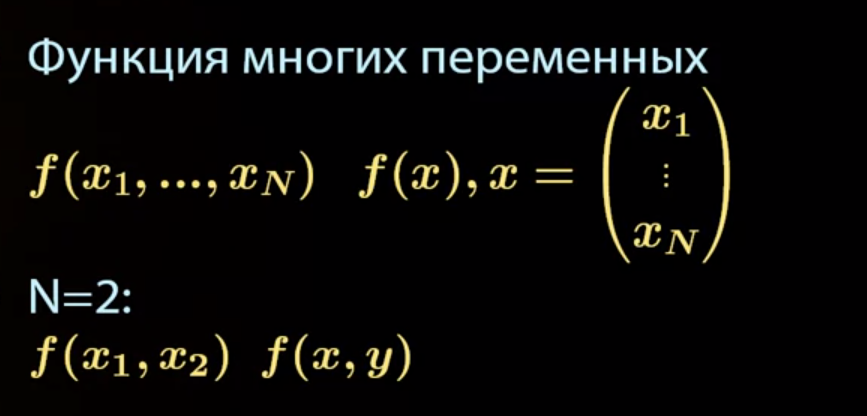
Мы можем обозначать функции многих переменных как f (x1,

..., xn), а можем обозначать как f(x), где x — это вектор.

С другой стороны, в случае, когда у нас переменные две,

мы можем обозначать функцию f(x1, x2) по аналогии с предыдущими обозначениями,

но есть достаточно привычный способ обозначать f(x, y).



**Применение градиента**

Представим, что мы решаем некоторую задачу оптимизации.

Ну, например, у нас есть функция от многих переменных, x1, ..., xn, и мы хотим её минимизировать.

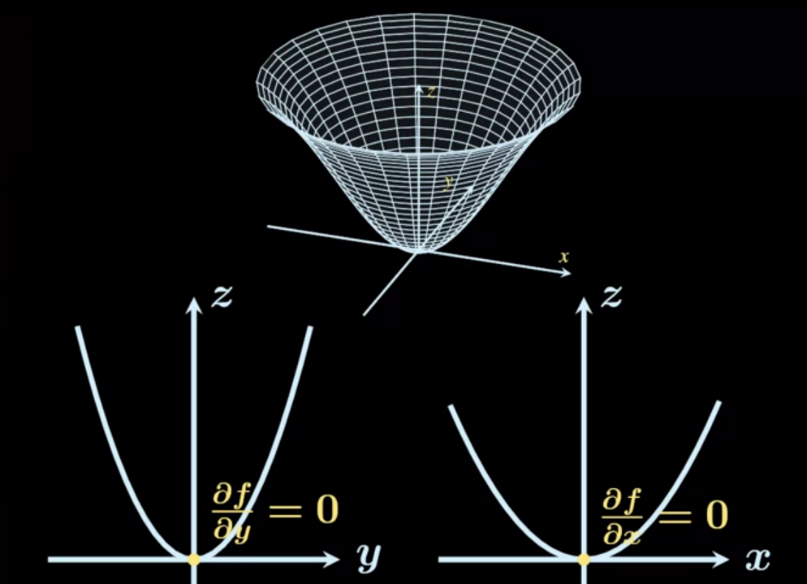
Ну откуда может взяться такая задача?

Например, мы можем выбирать оптимальные параметры рекламной кампании или подбирать веса классификатора текстов, минимизируя ошибку классификации, или настраивать параметры сложного технологического процесса, минимизируя вероятность неудачи. Но как же нам выбрать параметры таким образом, чтобы функция была минимальной?

Давайте посмотрим на ситуацию в простом двумерном случае,

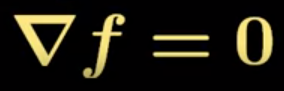
то есть в случае функции от двух переменных x и y.

Если зафиксировать значение x и посмотреть на всё в координатах z и y,



и если зафиксировать значение y и посмотреть на всё в координатах z и x, из графиков станет понятно,

что значения частных производных в точке экстремума должны быть нулевыми.



Так как все частные производные должны равняться 0, можно просто сказать, что градиент должен быть нулевым вектором.

**(необходимое условие экстремума)**Но бывают ситуации, когда функция не выписывается аналитически или когда полученные уравнения слишком сложны.

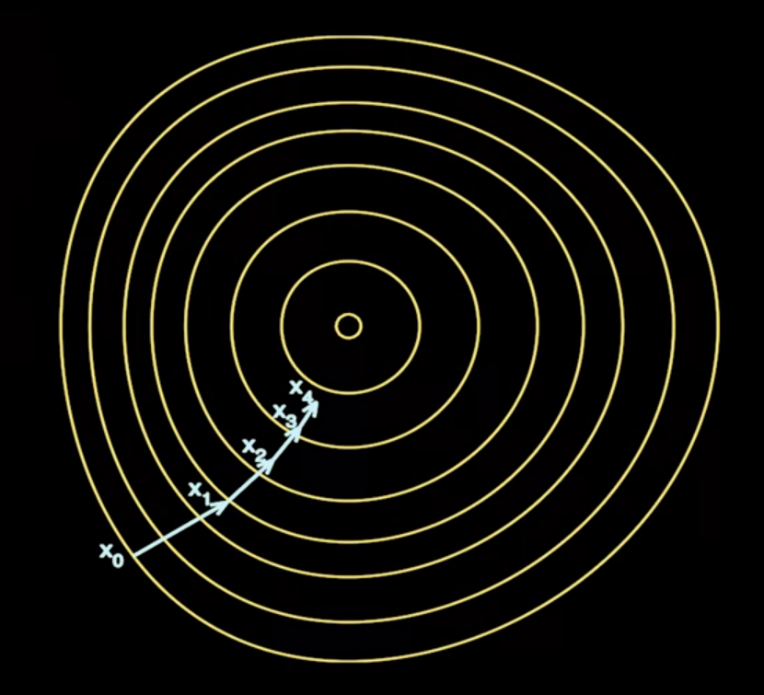
Что же делать в этом случае?

В этом случае нам поможет численная оптимизация.

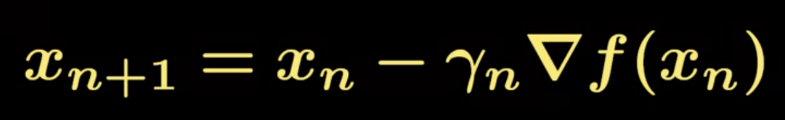
Так как *градиент* — это направление наискорейшего роста,

а *антиградиент* — это направление наискорейшего убывания,

то просто начнём в произвольной точке (x1, ..., xN), и на каждом шаге будем двигаться в сторону антиградиента.



Таким образом, значение x n+1 (нашего аргумента функции, векторного аргумента) будет получаться из прошлого значения xn просто вычитанием антиградиента, умноженного на некоторую величину, называемую *шагом градиентного спуска.*

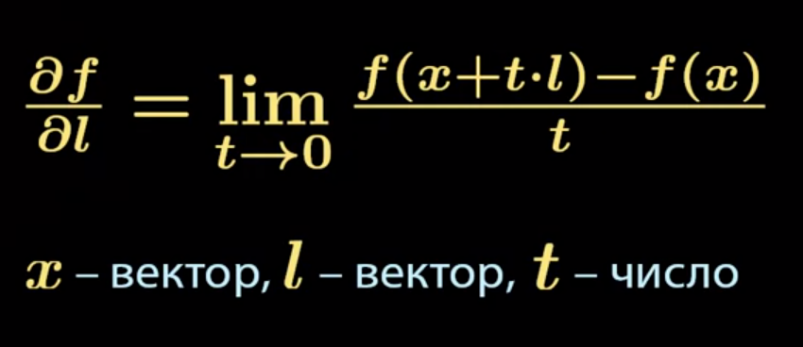


**Производная по направлению**

Представим, что мы хотим узнать, как меняется функция в определенной точке x в направлении l.

l — это некоторый вектор.

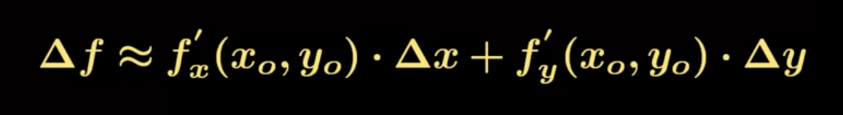
Для этого можно посмотреть на значение функции в точке x + t \* l, и сравнить его со значением функции в точке x.



Если мы будем смотреть на отношение разности этих значений к параметру t и будем устремлять параметр t к 0, то всё это выражение будет стремиться к некоторому числу, которое и называется **производной по направлению** **l**.

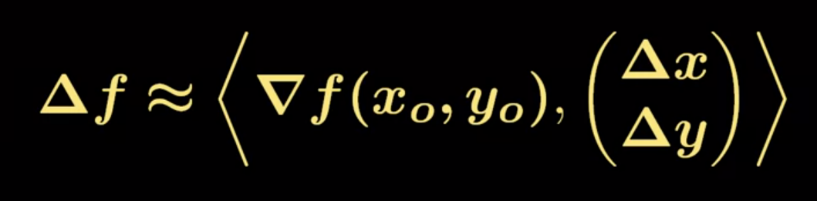
Так как нас интересует только направление вектора l, мы можем считать, что мы ограничились векторами, норма которых равна 1.

**Касательная плоскость и линейное приближение**



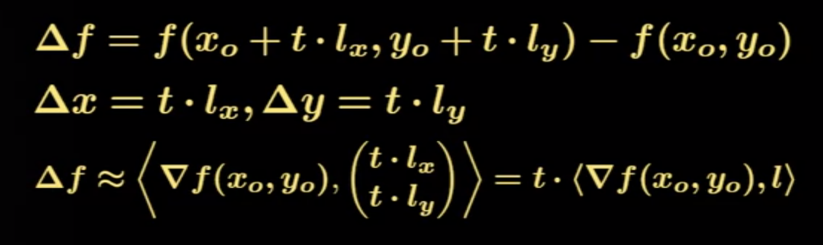
Давайте рассмотрим изменение значения функции f(x, y) при переходе от точки (x0, y0) к точке (x0 + Δx, y0 + Δy).

Как мы знаем, мы можем оценить изменение значения функции с помощью несложной формулы, используя частные производные. Для этого достаточно взять частную производную функции по x и умножить ее на Δx, взять частную производную по y и умножить ее на Δy и сложить.



Эту формулу можно записать компактней, используя скалярное произведение градиента на вектор изменения значений координат.

  
  
Чем ближе мы к точке (x0, y0), тем более точно выполняется наше приближенное равенство. Так как это выражения для Δf линейно по Δx = x − x0 и Δy = y − y0, оценка значений функции с помощью этого выражения называется линейным приближением. Линейное приближение часто используется для анализа сложных функций, в частности при построении алгоритмов анализа данных.   
  
**Направление наискорейшего роста**

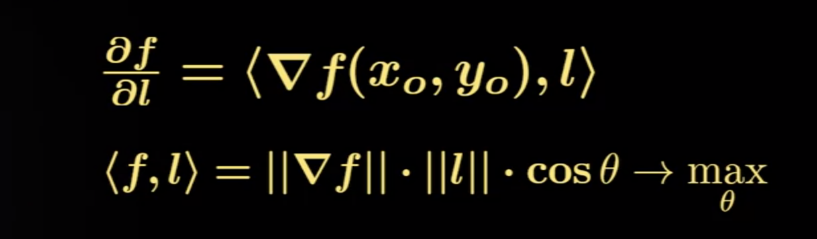


Давайте обозначим изменение значения функции Δf, и в наших прошлых обозначениях Δx теперь будет t \* lx, а Δy — t \* ly, то есть наши сдвиги по x и по y. Если мы распишем Δf по приближенной формуле, то мы увидим, что получается t умножить на скалярное произведение градиента, на вектор l. 

При стремлении t к 0 Δx и Δy тоже стремятся к 0. Ну то есть мы все ближе подбираемся к точке (x0, y0). А как вы помните, чем ближе мы к точке (x0, y0), тем точнее работает наша приближенная формула. Именно поэтому мы можем воспользоваться ей под знаком предела. Таким образом, мы видим, что Δf / t при t стремящемся к 0 будет стремиться к скалярному произведению градиента на вектор l.

И это есть производная по направлению в точке (x0, y0).

Когда производная по направлению будет максимальной, при каком направлении.



Скалярное произведение — вещь достаточно простая. Это просто произведение норм векторов на косинус угла между ними. На нормы вектора l мы влиять не можем, потому что мы им задаем просто направление. Мы можем считать, что его норма просто равна 1. На градиент мы тоже влиять не можем.

А вот косинус θ мы можем менять, выбирая разные направления. И косинус будет максимален тогда, когда равен 1, то есть когда угол θ между градиентом и вектором l равен 0, то есть вектор l направлен туда же, куда и градиент.