忆臻
哈尔滨工业大学 计算机科学与技术博士在读

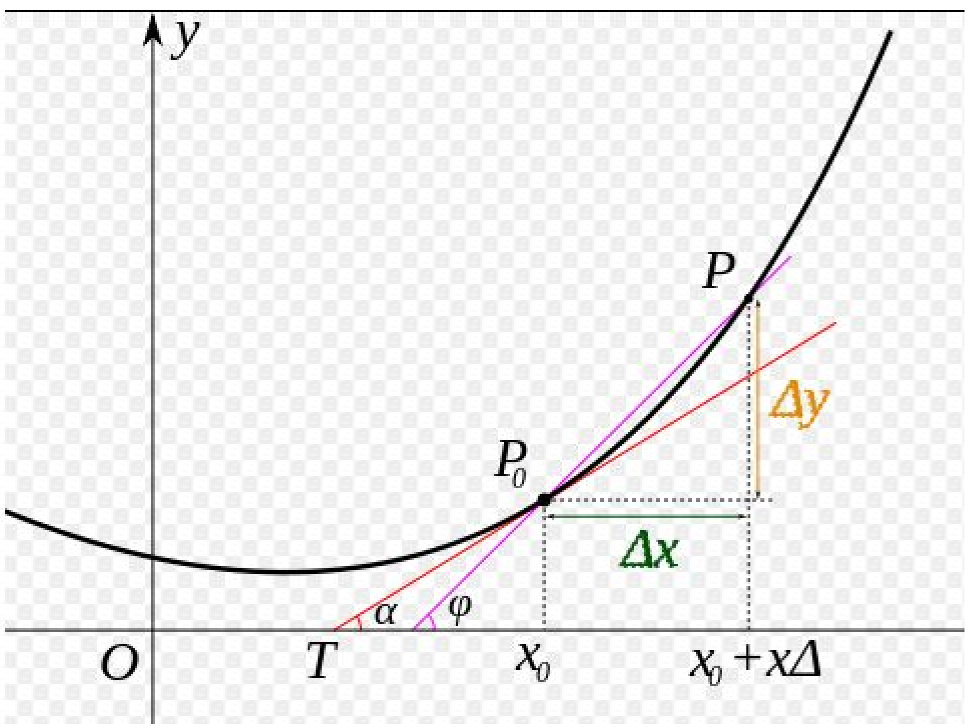
下面我一开始不提梯度的概念，完全根据自己的理解进行下文的梳理，一步一步推出梯度的来历：

导数

导数的几何意义可能很多人都比较熟悉: 当函数定义域和取值都在实数域中的时候，导数可以表示函数曲线上的切线斜率。 除了切线的斜率，导数还表示函数在该点的变化率。

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

将上面的公式转化为下面图像为：



（来自维基百科）

直白的来说，导数代表了在自变量变化趋于无穷小的时候，函数值的变化与自变量变化的比值代表了导数，**几何意义**有该点的切线。物**理意义**有该时刻的（瞬时）变化率...

注意在一元函数中，只有一个自变量变动，也就是说只存在一个方向的变化率，这也就是为什么一元函数没有偏导数的原因。

偏导数

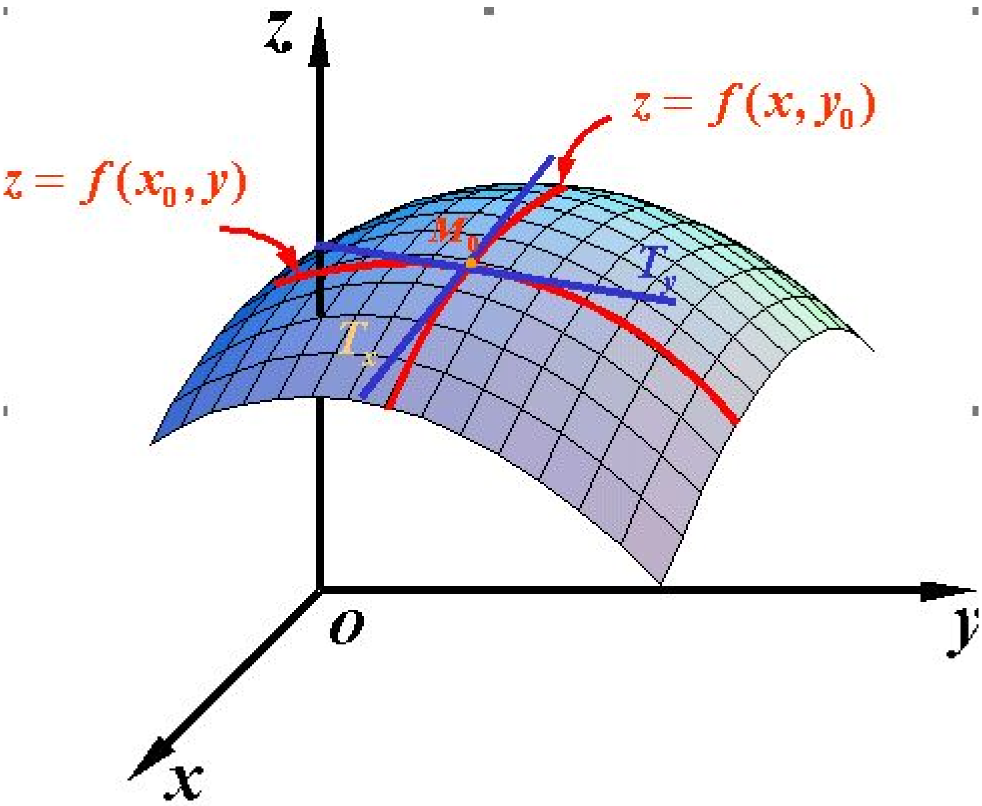
既然谈到偏导数，那就至少涉及到两个自变量，以两个自变量为例，z=f(x,y) . 从导数到偏导数，也就是从曲线来到了曲面. 曲线上的一点，其切线只有一条。但是曲面的一点，切线有无数条。

而我们所说的偏导数就是指的是多元函数沿坐标轴的变化率.

$f_x(x,y)$ 指的是函数在y方向不变，函数值沿着x轴方向的变化率

$f_y(x,y)$ 指的是函数在x方向不变，函数值沿着y轴方向的变化率

对应的图像形象表达如下：



那么偏导数对应的几何意义是是什么呢?

偏导数 $f_x(x_0,y_0)$ 就是曲面被平面 $y=y_0$ 所截得的曲面在点 M_0 处的切线 M_0T_x 对x轴的斜率
偏导数 $f_y(x_0,y_0)$ 就是曲面被平面 $x=x_0$ 所截得的曲面在点 M_0 处的切线 M_0T_y 对y轴的斜率

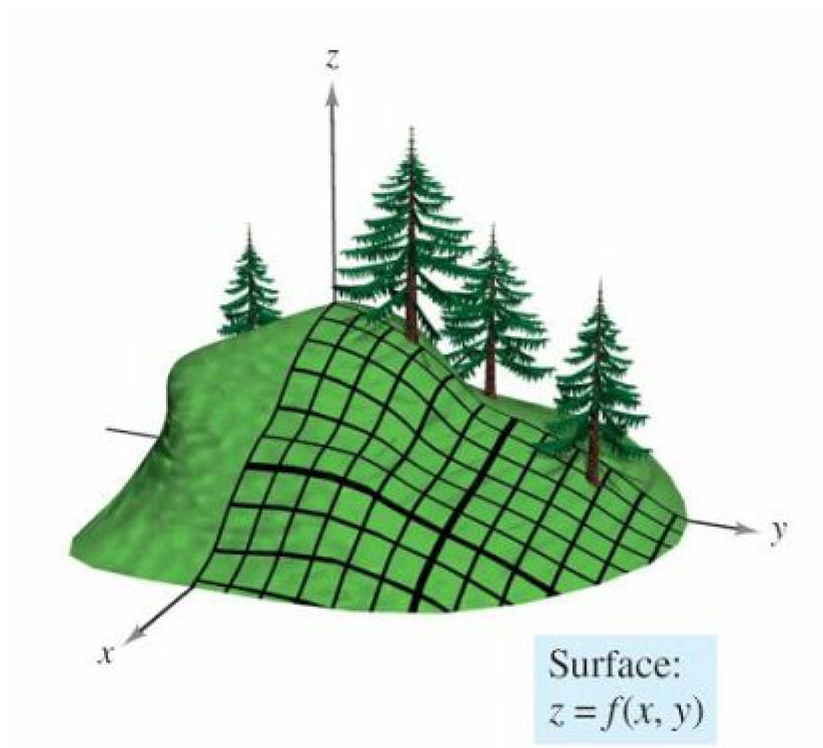
可能到这里，读者就已经发现偏导数的局限性了，原来我们学到的偏导数指的是多元函数沿坐标轴的变化率，但是我们往往很多时候要考虑多元函数沿任意方向的变化率，那么就引出了方向导数.

方向导数

终于引出我们的重头戏了，方向导数，下面我们慢慢来走进它

假设你站在山坡上，相知道山坡的坡度（倾斜度）

山坡图如下：



假设山坡表示为 $z=f(x,y)$,你应该已经会做主要俩个方向的斜率.

y方向的斜率可以对y偏微分得到.

同样的，x方向的斜率也可以对x偏微分得到

那么我们可以使用这俩个偏微分来求出任何方向的斜率（类似于一个平面的所有向量可以用俩个基向量来表示一样）

现在我们有这个需求，想求出 u 方向的斜率怎么办。假设 $z=f(x,y)$ 为一个曲面，$p(x_0,y_0)$ 为 f 定义域中一个点，单位向量 $u=cos\theta i+sin\theta j$ 的斜率，其中 θ 是此向量与 x 轴正向夹角。单位向量 u 可以表示对任何方向导数的方向。如下图：

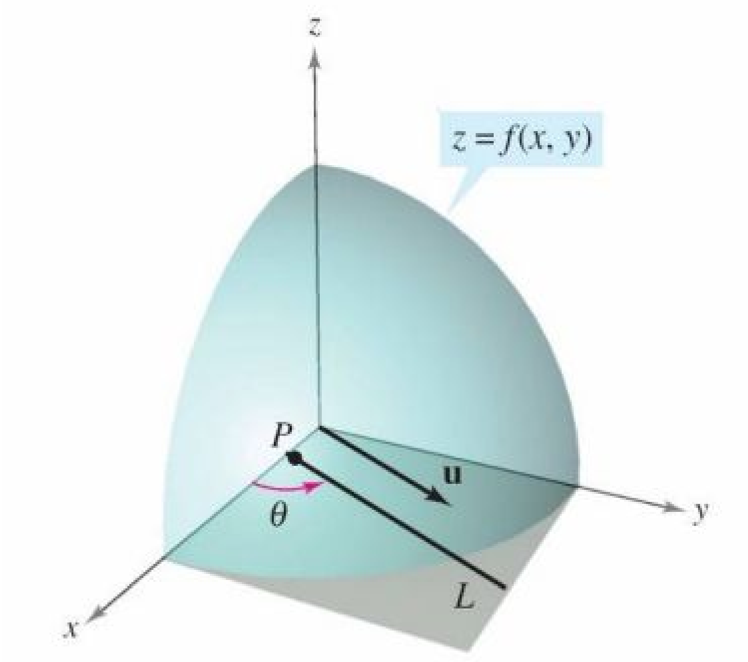


Figure 13.43

那么我们来考虑如何求出 u 方向的斜率，可以类比于前面导数定义，得出如下：

设 $f(x,y)$ 为一个二元函数，$u=cos\theta i+sin\theta j$ 为一个单位向量，如果下列的极限值存在

$\lim\limits_{t\to 0}\dfrac{f(x_0+tcos\theta,y_0+tsin\theta)-f(x_0,y_0)}{t}$ 此方向导数记为 $D_u f$

则称这个极限值是 f 沿着 u 方向的方向导数，那么随着 θ 的不同，我们可以求出任意方向的方向导数。这也表明了方向导数的用处，是为了给我们考虑函数对任意方向的变化率。

在求方向导数的时候，除了用上面的定义法求之外，我们还可以用偏微分来简化我们的计算。

表达式是： $D_u f(x,y)=f_x(x,y)cos\theta+f_y(x,y)sin\theta$ （至于为什么成立，很多资料有，不是这里讨论的重点）

那么一个平面上无数个方向，函数沿哪个方向变化率最大呢？

目前我不管梯度的事，我先把表达式写出来：

$$D_u f(x,y)=f_x(x,y)cos\theta+f_y(x,y)sin\theta$$

设 $A = (f_x(x,y), f_y(x,y))$ ŋ̂$I = (cos\theta, sin\theta)$

那么我们可以得到：

$D_u f(x,y) = A \bullet I = |A| * |I|\, cos\alpha$ Hα 为向量 A 与向量 I 之间的夹角ɪ

那么此时如果 $D_u f(x,y)$ 要取得最大值，也就是当 α 为0 度的时候，也就是向量 I（这个方向是一直在变，在寻找一个函数变化最快的方向）与向量 A（这个方向当点固定下来的时候，它就是固定的）平行的时候，方向导数最大ŋ̂方向导数最大，也就是单位步伐，函数值朝这个反向变化最快ŋ̂

好了，现在我们已经找到函数值下降最快的方向了，这个方向就是和 A 向量相同的方向ŋ̂那么此时我把2 向量命名为梯度（当一个点确定后，梯度方向是确定的），也就是说明了为什么梯度方向是函数变化率最大的方向了！！！（因为本来就是把这个函数变化最大的方向命名为梯度）

我的理解是，本来梯度就不是横空出世的，当我们有了这个需求（要求一个方向，此方向函数值变化最大），得到了一个方向，然后这个方向有了意义，我们给了它一个名称，叫做梯度（纯个人理解ȯ希望对大家理解有帮助）欢迎知友提出问题交流ȯ