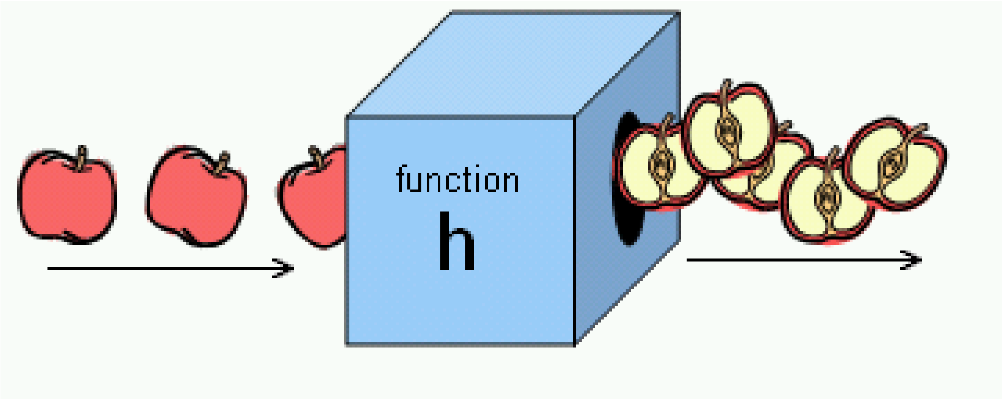
AI数学基石——高等数学篇

目录

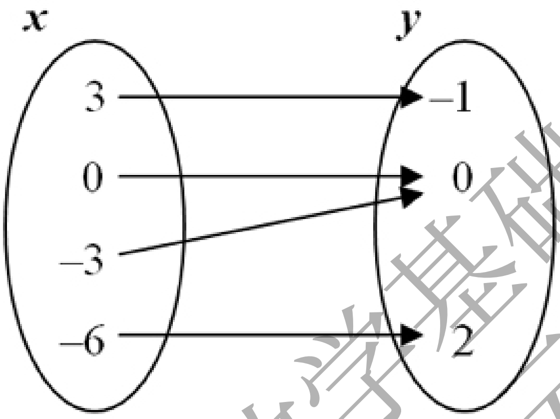
01 函数

1.1 函数的定义

形象地说，函数就是一个实现某种功能的盒子。



在数学里，函数是从实数集到实数集的映射：

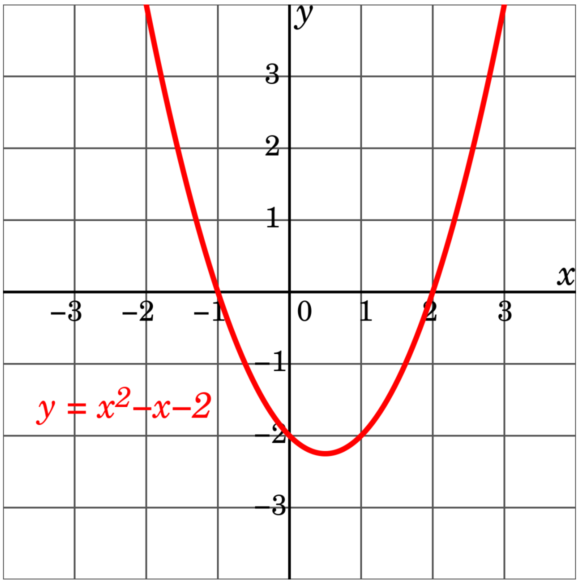


定义

> 设**数集**$D \subset \mathbb{R}$，则称映射$f: D \to \mathbb{R}$为定义在D上的**函数**，通常简记为
>
> $$y = f(x),\ x \in D,$$
>
> 其中x称为**自变量**，y称为**因变量**，D称为**定义域**，记作D_f，函数值$f(x)$所构成的集合称为函数的**值域**。

例子

- 一次函数：$y = ax + b$
- 二次函数：$y = ax^2 + bx + c$
- 指数函数：$y = a^x$
- 对数函数: $y = log_a(x)$



1.2 反函数

定义

> 若函数$f: D \to f(D)$是单射，它存在逆映射$f^{-1}: f(D) \to D$,则称此映射f^{-1}为函数f的**反函数**。

例子

$$y = x^3, \quad x \in \mathbb{R} \iff x = y^{1/3}, \quad y \in \mathbb{R}$$

1.3 复合函数

定义

> 设函数$y = f(u)$的定义域为D_f,函数$u = g(x)$定义域为D_x，且其值域$R_k \subset D_f$，则由下式确定的函数
>
> $$y = f[g(x)], \quad x \in D_x$$
>
> 称为由函数$u = g(x)$与函数$y = f(u)$构成的**复合函数**，它的定义域为D_x，变量u称为**中间变量**。函数f和g构成的复合函数通常记为$f \circ g$，即$(f \circ g)(x) = f[g(x)]$

$$x \xrightarrow{\quad g \quad} g(x) = u \xrightarrow{\quad f \quad} f(u) = f[g(x)]$$

例子

物体运动的动能为$E=mv^2/2$，而自由落体的速度为$v=gt$，所以自由落体的动能是时间t的复合函数：

$$E=\frac{1}{2}mg^2t^2$$

02 导数

2.1 引例

变速直线运动的速度

设某质点运动位置关于时间的函数为

$$s=f(t)$$

，则t_0到t的平均速度为

$$\overline{v}=\frac{f(t)-f(t_0)}{t-t_0}$$

在t_0时刻的瞬时速度为

$$v_{t_0}=\lim_{t\to t_0}\frac{f(t)-f(t_0)}{t-t_0}$$

2.2 导数的定义

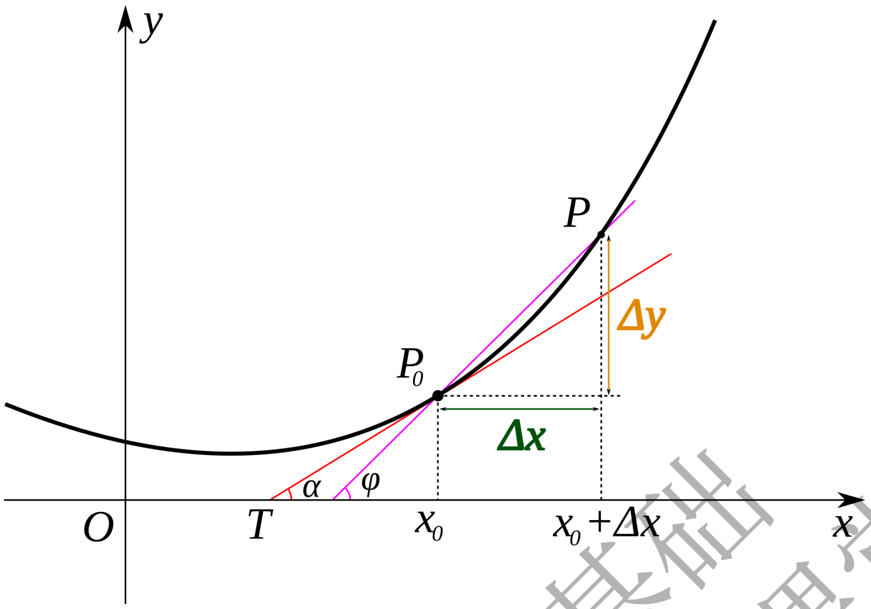
定义

> 设函数$y=f(x)$在点x_0的某个邻域内有定义，$\triangle x$为在x_0点附近取得的增量（点$x_0+\triangle x$仍在该邻域内）时，若
>
> $$\lim_{\triangle x\to 0}\frac{f(x_0+\triangle x)-f(x_0)}{\triangle x}=\lim_{\triangle x\to 0}\frac{\triangle y}{\triangle x}$$
>
> 存在，则称函数$y=f(x)$在点x_0处可导, 并称这个极限为函数$y=f(x)$在点x_0处的**导数**，记为$f'(x_0)$，
> 也可记作
>
> $$y'|_{x=x_0},\frac{dy}{dx}\ |_{x=x_0}\ \text{或}\frac{df(x)}{dx}\ |_{x=x_0}$$

简单地说，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。

导数的几何含义

根据导数的定义可知，图中P_0点的导数就是P_0点在函数曲线上的切线斜率。



2.3 函数的求导法则

常数和基本初等函数的导数公式

$$(C)' = 0, \qquad (x^{\mu})' = \mu x^{\mu-1},$$

$$(\sin x)' = \cos x, \qquad (\cos x)' = -\sin x,$$

$$(\tan x)' = \sec^2 x, \qquad (\cot x)' = -\csc^2 x,$$

$$(\sec x)' = \sec x \tan x, \qquad (\csc x)' = -\csc x \cot x,$$

$$(a^x)' = a^x \ln a, \qquad (\mathrm{e}^x)' = \mathrm{e}^x,$$

$$(\log_a x)' = \frac{1}{x \ln a}, \qquad (\ln x)' = \frac{1}{x},$$

$$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}, \qquad (\arccos x)' = -\frac{1}{\sqrt{1-x^2}},$$

$$(\arctan x)' = \frac{1}{1+x^2}, \qquad (\mathrm{arc\ cot}\ x)' = -\frac{1}{1+x^2}.$$

函数和差积商的求导法则

定理

> 如果函数$u=u(x)$及$v=v(x)$都在点x具有导数，那么它们的和, 差, 积, 商（除分母为零的点外）都在点x具有导数，且
>
> （1）
>
> $$[u(x) \pm v(x)]' = u'(x) \pm v'(x);$$

> （2）
>
> $$[u(x)v(x)]' = u'(x)v(x) + u(x)v'(x);$$
>
> （3）
>
> $$[\frac{u(x)}{v(x)}]' = \frac{u'(x)v(x) - u(x)v'(x)}{v^2(x)}(v(x) \neq 0)$$
>
> （4）
>
> $$(Cu)' = Cu' \ \ (C\text{为常数})$$

例子

已知$y = 2x^3 - 5x^2 + 3x - 7$，求y'。

解

$$\begin{aligned} y' &= (2x^3 - 5x^2 + 3x - 7)' \\ &= (2x^3)' - (5x^2)' + (3x)' - (7)' \\ &= 2 \cdot 3x^2 - 5 \cdot 2x + 3 - 0 \\ &= 6x^2 - 10x + 3 \end{aligned}$$

复合函数的求导法则

> 设$y = f(u)$, 而$u = g(x)$且$f(u)$及$g(x)$都可导，则复合函数$y = f[g(x)]$的导数为
>
> $$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$
>
> 或
>
> $$y'(x) = f'(u) \cdot g'(x)$$

例1

已知$y = e^{x^3}$，求$\dfrac{dy}{dx}$。

解

$$y = e^{x^3}\text{可以看作由}y = e^u, u = x^3\text{复合而成，因此}$$

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = e^u \cdot 3x^2 = 3x^2 e^{x^3}$$

例2

已知$y = ln\ sin\ x$，求$\dfrac{dy}{dx}$。

解

$$y = ln\ sin\ x\text{可以看作由}y = ln\ u, u = sin\ x\text{复合而成，因此}$$

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = \frac{1}{u} \cdot cos\ x = \frac{cos\ x}{sin\ x}$$

2.4 高阶导数

定义

> 若函数$y = f(x)$的导数$y' = f'(x)$可导，则称$f'(x)$的导数为$f(x)$的二阶导数，记作y''或$\dfrac{d^2y}{d\,x^2}$，即
>
> $$y'' = (y')'\text{或}\frac{d^2y}{d\,x^2} = \frac{d}{d\,x}(\frac{dy}{dx})$$
>
> 类似地，二阶导数的导数称为三阶导数，依次类推。二阶和二阶以上的导数统称为**高阶导数**

例子

已知$y = ax + b$，求y''。

解

$$y' = a,\ \ y'' = 0$$

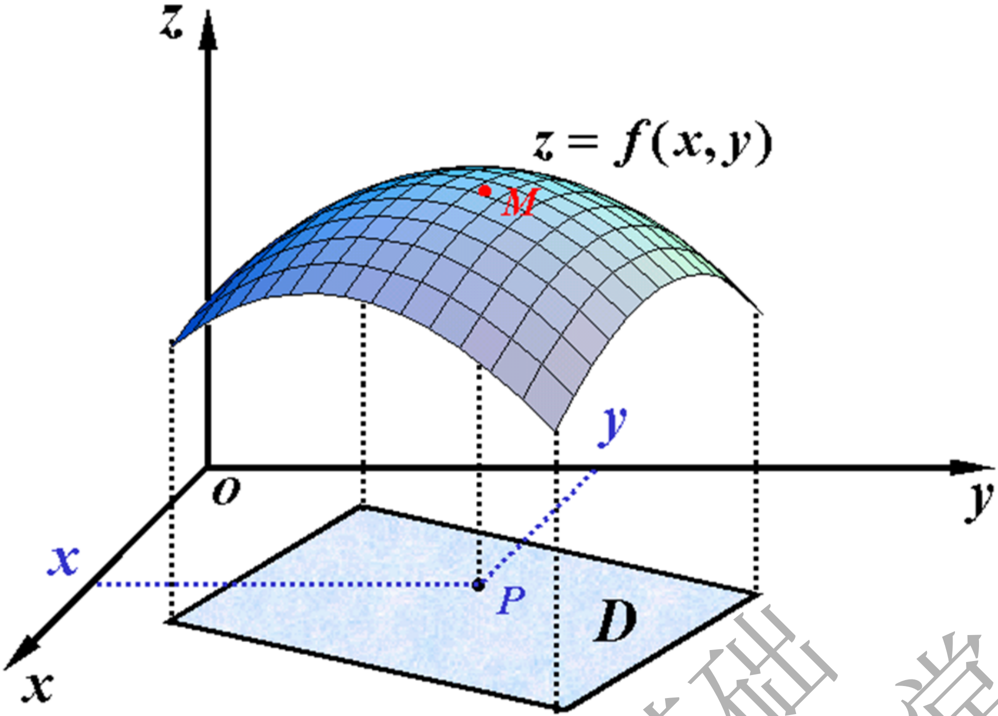
03 偏导数

3.1 二元函数

定义

> 设D是平面上的一个点集，如果对于每个点$P(x, y) \in D$，变量z按照一定的法则总有确定的值与它对应，则称z是变量(x, y)的二元函数，记为$z = f(x, y)$。

类似地可以定义三元以及三元以上函数。当$n \geq 2$时，n元函数统称为**多元函数**。
多元函数中同样有定义域，值域，自变量，因变量等概念。

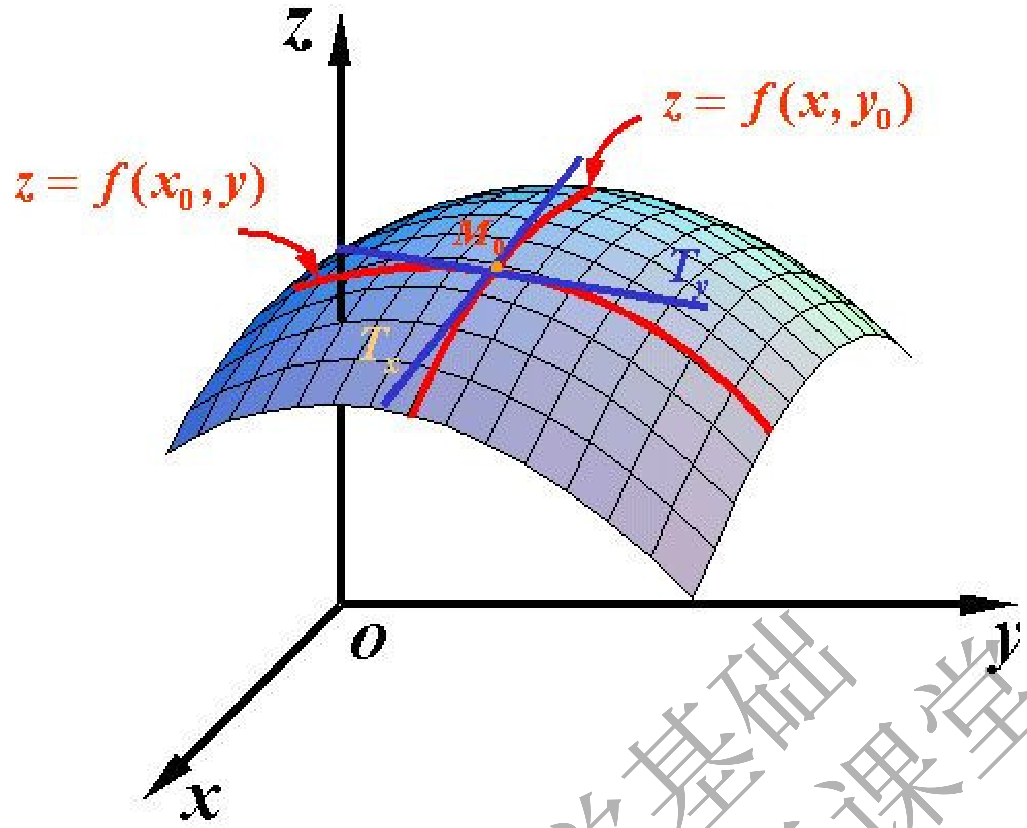


3.2 二元函数的偏导数

定义

> 假定二元函数$z=f(x,y)$在点(x_0,y_0)的某一邻域内有定义，当y固定在y_0而x在x_0上有增量$\triangle x$时，相应的函数有增量
>
> $$f(x_0+\triangle x,y_0)-f(x_0,y_0),$$
>
> 如果
>
> $$\lim_{\triangle x\to 0}\frac{f(x_0+\triangle x,\ y_0)-f(x_0,\ y_0)}{\triangle x}$$
>
> 存在，则称此极限值称为函数$z=f(x,y)$在(x_o,y_0)处对x的**偏导数**，记作$f_x(x_0,y_0)$，或对x的偏导数记作$\left.\frac{\partial f}{\partial x}\right|_{\substack{x=x_0\\ y=y_0}}$，$\left.\frac{\partial z}{\partial x}\right|_{\substack{x=x_0\\ y=y_0}}$，$\left.z_x\right|_{\substack{x=x_0\\ y=y_0}}$

几何含义



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

简单来说，**偏导数就是多元函数沿坐标轴方向的变化率**。

例子

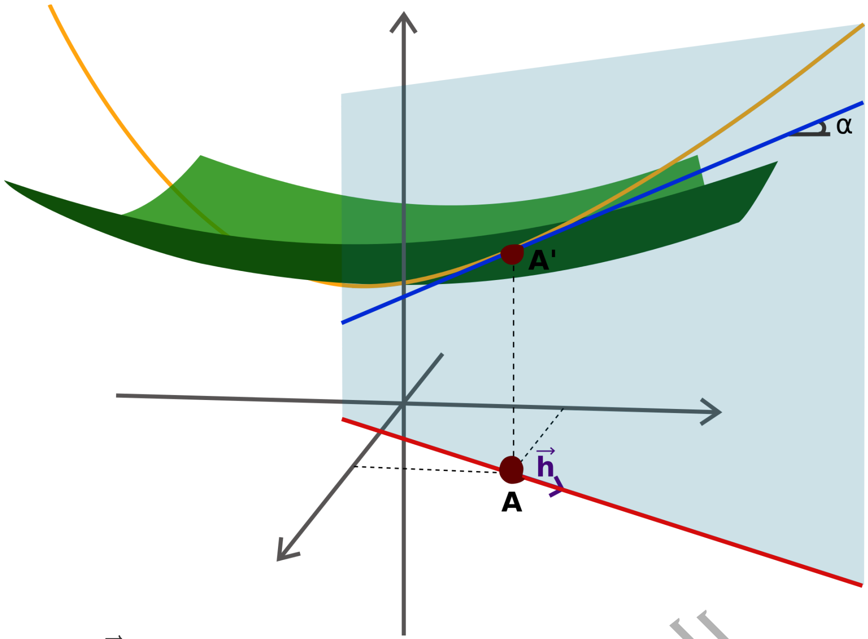
求$z = x^2 + 3xy + y^2$在点$(1, 2)$处的偏导数。
解

$$\frac{\partial z}{\partial x} = 2x + 3y, \quad \frac{\partial z}{\partial y} = 3x + 2y$$

$$\left.\frac{\partial z}{\partial x}\right|_{(\mathbf{1},\mathbf{2})} = 2 \cdot 1 + 3 \cdot 2 = 8, \quad \left.\frac{\partial z}{\partial y}\right|_{(\mathbf{1},\mathbf{2})} = 3 \cdot 1 + 2 \cdot 2 = 7$$

4.3 方向导数和梯度

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**。



那如何求出函数沿图中向量$\vec{h}$方向的变化率呢?

首先，假设A点的坐标为(x_0, y_0)，向量$\vec{h}$与x轴形成的夹角为θ，再假设我们现在从A点往向量方向移动t距离，那么到达的新的点的坐标为$(x_0+tcos\theta, y_0+tsin\theta)$，根据导数的定义，

$$\lim_{t\to 0}\frac{f(x_0+tcos\theta, y_0+tsin\theta)-f(x_0, y_0)}{t}$$

则为函数沿着$\vec{h}$方向的变化率，也叫函数沿着$\vec{h}$的方向导数。
在上例中，如果改变θ，我们就可以求出函数在任意方向上的方向导数（变化率）。

方向导数也可以用偏微分来表示：

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

接下来的问题是：一个平面上无数个方向，函数沿哪个方向变化率最大呢？

首先，我们把方向导数的式子

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

写成向量之间乘积的形式：

$$D_u f(x,y)=A\cdot I=||A||*||I||*cos\alpha$$

$$其中A=(f_x(x,y),f_y(x,y)),I=(cos\theta,sin\theta),\alpha为向量A和I之间的夹角$$

从上式可以看出，当α为0度的时候，$D_u f(x,y)$会取得最大值。换句话说，当向量I与向量A平行的时候，方向导数最大，即函数值在这个方向变化最快。向量A的另一个名字叫**梯度**，这也是为什么我们说**在梯度的方向上，函数值变化最大**。

3.4 雅可比（Jacobian）矩阵

定义

> 假设 $F:\mathbb{R}_n \rightarrow \mathbb{R}_m$ 是一个从 n维欧氏空间映射到到 m 维欧氏空间的函数。这个函数由 m个实函数组成:
>
> $$y_1(x_1,\cdots,x_n),\cdots,y_m(x_1,\cdots,x_n)。$$
>
> 这些函数的偏导数(如果存在)可以组成一个 m行 n列的矩阵，這個矩陣就是所谓的雅可比矩阵：
>
> $$\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$
>
> 此矩阵用符号表示为：
>
> $J_F(x_1,\ldots,x_n)$，或者$\frac{\partial(y_1,\ldots,y_m)}{\partial(x_1,\ldots,x_n)}$.

例子

设有映射 $f(x,y)$其中函数为

$$\mathbf{f}(x,y)=\begin{bmatrix} x^2y \\ 5x+\sin y \end{bmatrix}.$$

则 f的两个变量为

$$f_1(x,y)=x^2y$$
$$f_2(x,y)=5x+\sin y$$

其雅可比矩阵为

$$\mathbf{J_f}(x,y)=\begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix}=\begin{bmatrix} 2xy & x^2 \\ 5 & \cos y \end{bmatrix}$$

<本例来源于知乎作者快来签到小助手>

3.4 海森（Hessian）矩阵

定义

在数学中，海森矩阵（Hessian matrix 或 Hessian）是一个自变量为向量的实值函数的二阶偏导数组成的方块矩阵，假設有一实数函数

$$f(x_1, x_2, \ldots, x_n),$$

如果f 所有的二阶偏导数都存在，那么 f 的海森矩阵的第 ij项，即：

$$H(f)_{ij}(x) = D_i D_j f(x)$$

其中 $x = (x_1, x_2, \ldots, x_n)$，即

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

<以上定义来自于维基百科>

例子

求函数$f(x,y) = x^3 - 2xy - y^6$的海森矩阵。

首先求一阶偏导

$$f_x(x,y) = \frac{\partial}{\partial x}(x^3 - 2xy - y^6) = 3x^2 - 2y$$

$$f_y(x,y) = \frac{\partial}{\partial y}(x^3 - 2xy - y^6) = -2x - 6y^5$$

再求二阶偏导

$$f_{xx}(x,y) = \frac{\partial}{\partial x}(3x^2 - 2y) = 6x$$

$$f_{xy}(x,y) = \frac{\partial}{\partial y}(3x^2 - 2y) = -2$$

$$f_{yx}(x,y) = \frac{\partial}{\partial x}(-2x - 6y^5) = -2$$

$$f_{yy}(x,y) = \frac{\partial}{\partial y}(-2x - 6y^5) = -30y^4$$

得到海森矩阵

$$\mathbf{H}f(x,y) = \begin{bmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{bmatrix} = \begin{bmatrix} 6x & -2 \\ -2 & -30y^4 \end{bmatrix}$$

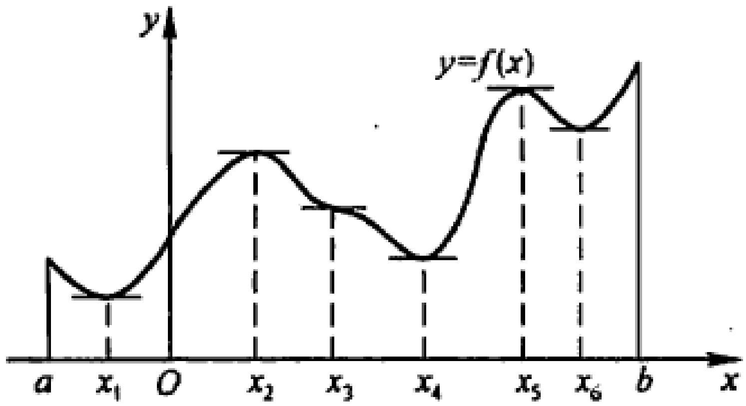
<例子来自于可汗学院>

04 函数的极值

4.1 定义

> 设函数$f(x)$在点x_0的某邻域$U(x_0)$内有定义，如果对于去心邻域$\mathring{U}(x_0)$内的任一x，有
>
> $$f(x) < f(x_0)或f(x) > f(x_0)$$
>
> 那么就称$f(x_0)$是函数的一个极大值或极小值。函数的极大值和极小值统称为函数的**极值**, 使函数取得极值的点被称作**极值点**。

最值是函数在其定义域内取得最大值（或最小值）的点的函数值。极值是局部性的概念，而最值是全局性的概念。



4.2 极值的定理

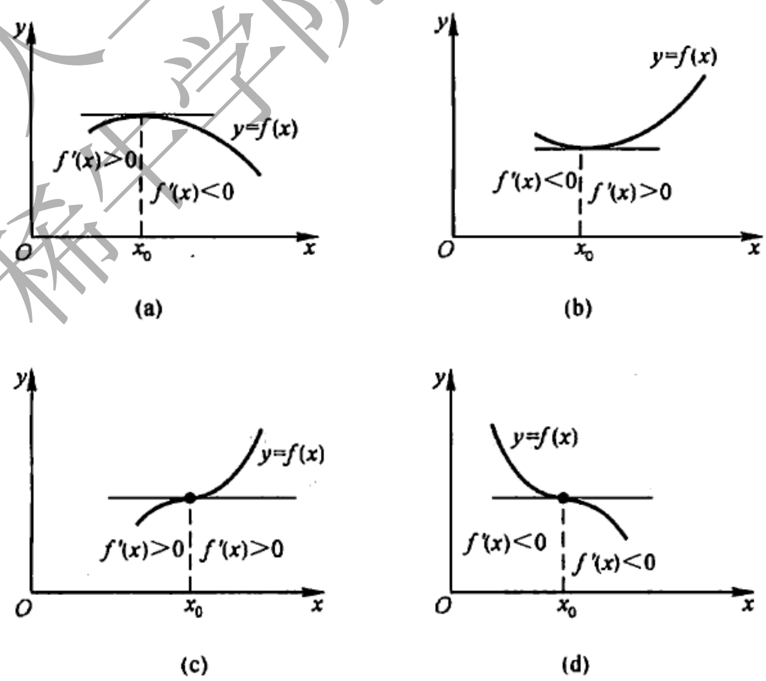
必要条件

> 设函数$f(x)$在x_0处可导，且在x_0处取得极值，那么$f'(x_0)=0$。

反过来并不成立，比如$f(x)=x^3$.

判定极值的第一充分条件

> 设函数$f(x)$在x_0处连续，且在x_0的某去心邻域$\mathring{U}(x_0,\delta)$内可导
> （1）若$x\in(x_0-\delta,x_0)$时，$f'(x)>0$,而$x\in(x_0,x_0+\delta)$时，$f'(x)<0$，则$f(x)$在x_0处取得极大值;
> （2）若 $x\in(x_0-\delta,x_0)$时，$f'(x)<0$,而 $x\in(x_0,x_0+\delta)$时，$f'(x)>0$,则$f(x)$在x_0处取得极小值
> （3）若$x\in\mathring{U}(x_0,\delta)$时，$f'(x)$的符号保持不变，则$f(x)$在$x_0$处没有极值



判定极值的第二充分条件

> 设函数$f(x)$在x_0处具有二阶导数且$f'(x_0)=0$，$f''(x_0)\neq 0$那么
> （1）当 $f''(x_0)<0$ 时，函数$f(x)$ 在x_0 处取得极大值;
> （2）当$f''(x_0)>0$ 时，函数$f(x)$ 在x_0 处取得极小值.

4.3 拉格朗日乘数法

引例

求表面积为a^2而体积最大的长方体的体积。

解

设长方体的三棱的长为x, y, z，则体积

$$V=xyz$$

又因假定表面积为a^2，所以自变量x，y，z还必须满足附加条件

$$2(xy+yz+xz)=a^2$$

像这种对自变量有附加条件的极值问题称为**条件极值**。而没有附加条件的问题就叫**无条件极值**。
有条件极值可以化为无条件极值，例如上述问题，可将条件$2(xy+yz+xz)=a^2$，用z表成x，y的函数

$$z=\frac{a^2-2xy}{2(x+y)}$$

再把它代人$V=xyz$中，于是问题就化为求

$$V=\frac{xy}{2}(\frac{a^2-2xy}{x+y})$$

概念

拉格朗日乘数法是一种寻找多元函数在其变量受到一个或多个条件的约束时的极值的方法。这种方法可以将一个有n个变量与k个约束条件的最优化问题转换为一个解有$n+k$个变量的方程组的解的问题。这种方法中引入的新未知数，即**拉格朗日乘数**，又称**拉格朗日乘子**。

定义

> 要找函数$z=f(x,y)$在附加条件$\varphi(x,y)=0$的可能极值点，可以先作拉格朗日函数
>
> $$L(x,y)=f(x,y)+\lambda\varphi(x,y),$$
>
> 其中λ为参数。求其对x与y的一阶偏导数，并使之为零，然后与附加条件联立起来：

> $$\begin{cases} f_x(x,y)+\lambda\varphi_x(x,y)=0 \\ f_y(x,y)+\lambda\varphi_y(x,y)=0 \\ \varphi(x,y)=0 \end{cases}$$
>
> 由这方程组解出x,y及λ, 这样得到的(x,y)就是函数$f(x,y)$在附加条件$\varphi(x,y)=0$下的可能极值点.

回到之前求体积的例子，已知

$$V=xyz$$
$$2(xy+yz+xz)=a^2$$

拉格朗日函数可以写成

$$L(x,y,z)=xyz+\lambda(2xy+2yz+2xz-a^2),$$

求其对x,y,z的偏导数，并使之为零，得到

$$yz+2\lambda(y+z)=0$$
$$xz+2\lambda(x+z)=0$$
$$xy+2\lambda(y+x)=0$$

上面三个式子在与条件等式联立，可得

$$\frac{x}{y}=\frac{x+z}{y+z},\frac{y}{z}=\frac{x+y}{x+z}$$

解得

$$x=y=z$$

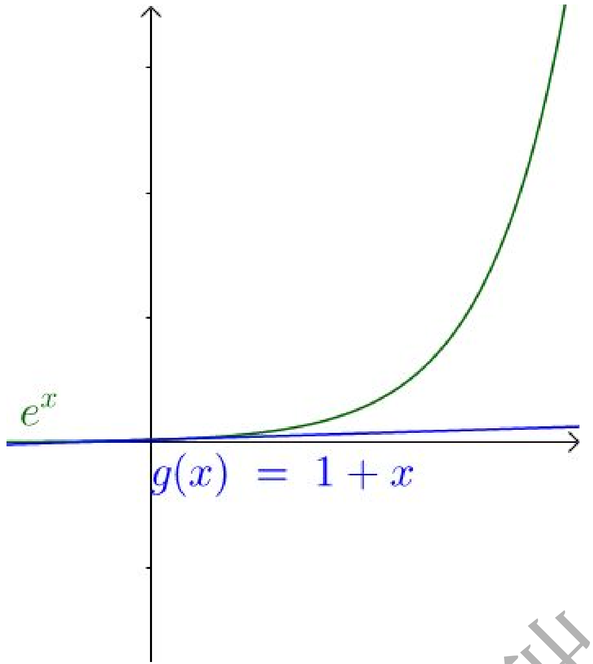
再将上式带入条件等式，最终得到

$$x=y=z=\frac{\sqrt{6}}{6}a$$

05 泰勒公式

对于一些较复杂的函数，我们经常会用一些简单的函数来近似表达。比如，在x很小的时候，

$$e^x\approx 1+x$$



但有没有什么函数可以近似整个曲线呢？答案就是泰勒公式的展开式。

泰勒公式，简单来说，就是**用多项式函数来逼近光滑函数**。

定义

> 设n是一个正整数。如果定义在一个包含a的区间上的函数f在a点处$n+1$次可导，那么对于这个区间上的任意x都有：
>
> $$f(x) = \frac{f(a)}{0!} + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x)$$
>
> $$= \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x)$$
>
> , 其中的多项式称为函数在a处的泰勒展开式， $R_n(x)$ 是泰勒公式的余项。

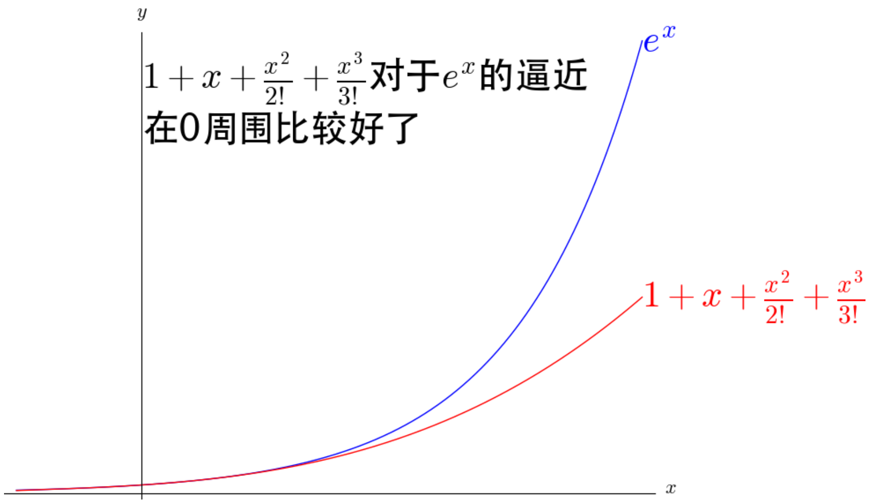
如果$a=0$的话，泰勒公式就叫麦克劳林公式。即：

$$f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(0)}{n!}(x-a)^n + R_n(x)$$

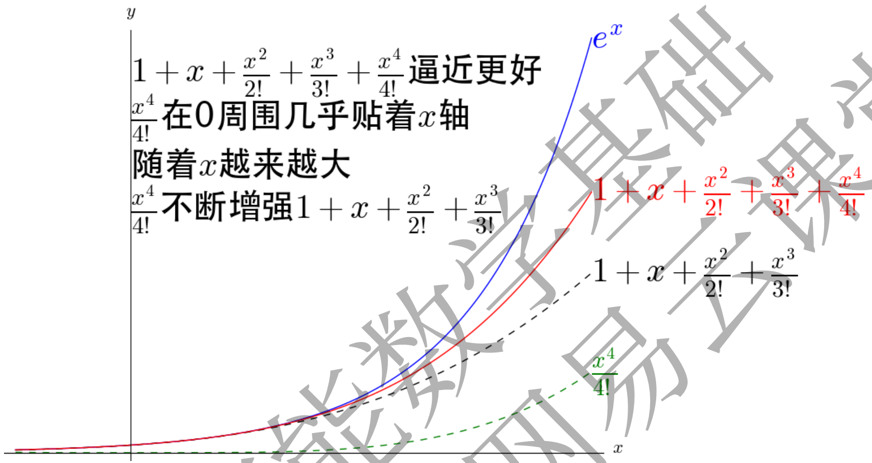
图例：用多项式对e^x进行逼近

根据泰勒公式，可以得到

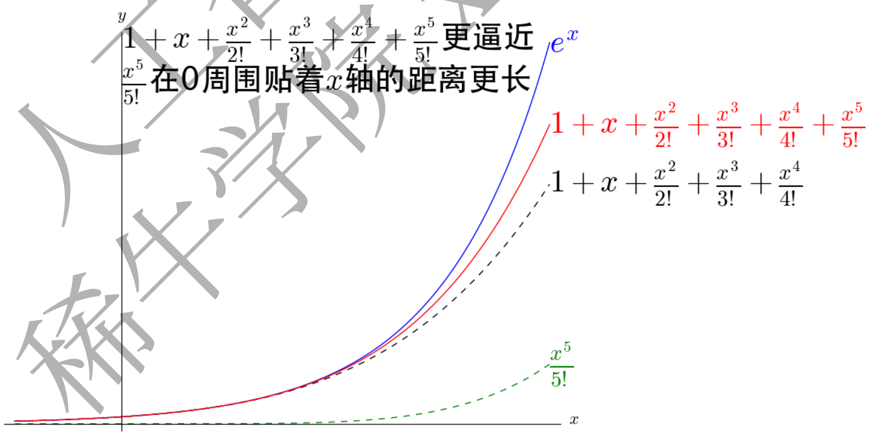
$$e^x = 1 + x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + R_n(x)$$



增加一个$\frac{1}{4!}x^4$



增加一个$\frac{1}{5!}x^5$



上图表示的就是泰勒公式在不同阶展开的效果，展开的多项式越多近似效果越好。

<图片来自于公众号《马同学高等数学》>

本章要点总结

- 导数就是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比值
- 导数的几何含义是函数曲线上的切线斜率
- 偏导数就是多元函数沿坐标轴方向的变化率

- 方向导数是多元函数沿任意方向的变化率，而梯度就是变化率最大的那个方向
- 判断函数极值的两个充分条件
- 拉格朗日乘数法是一种寻找多元函数在其变量受到一个或多个条件的约束时的极值的方法
- 泰勒公式就是用多项式函数来逼近函数

参考资料

- 《高等数学 第六版》上下册 by 高等教育出版社
- 《马同学高等数学》公众号文章
- 为什么梯度反方向是函数值局部下降最快的方向？
- 维基百科