西安电子科技大学
XIDIAN UNIVERSITY

智能感知与图像理解

最优化理论

第二章 最优化问题的数学基础I

人工智能学院
智能感知与图像理解实验室

第二章 最优化问题的数学基础 I

向量的内积

设在n维线性空间R^n中， $\vec{a}=(a_1,a_2,\cdots,a_n)^T\in R^n$,

$\vec{b}=(b_1,b_2,\cdots,b_n)^T\in R^n$,则

$a_1b_1+a_2b_2+\cdots+a_nb_n$称为向量$\vec{a}$与$\vec{b}$的内积，

记作$\left[\vec{a},\vec{b}\right]=\vec{a}^T\vec{b}=(a_1,a_2,\cdots,a_n)\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$

内积的性质

i) $[\alpha,\beta]=[\beta,\alpha]$（对称性）；

ii) $[\alpha+\beta,\gamma]=[\alpha,\gamma]+[\beta,\gamma]$,

$[k\alpha,\beta]=k[\beta,\alpha]$（线性性）；

iii) $[\alpha,\alpha]\geq 0$当且仅当$\alpha=\vec{0}$时，

$[\alpha,\alpha]=0$（正定性）；

向量性质

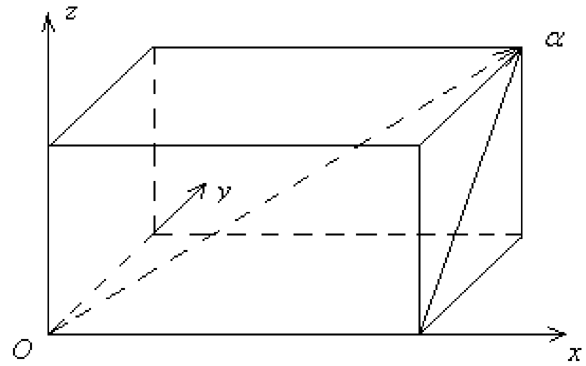
向量的长 $\|\alpha\| = \sqrt{[\alpha, \alpha]}$

单位向量 $\|\alpha\| = 1$

向量的夹角 $\langle \alpha, \beta \rangle = \arccos \dfrac{[\alpha, \beta]}{\|\alpha\| \cdot \|\beta\|}$

$$0 \le \langle \alpha, \beta \rangle \le \pi$$

向量的正交 $\langle \alpha, \beta \rangle = \dfrac{\pi}{2} \Leftrightarrow [\alpha, \beta] = 0$



范数

定义：若函数$\|\cdot\|: \mathrm{R}^n \to \mathrm{R}$满足：

(1)正定性：$\forall x \in R^n, \|x\| \ge 0, \|x\| = 0 \Leftrightarrow x = 0;$

(2)三角不等式：$\forall x, y \in R^n, \|x + y\| \le \|x\| + \|y\|;$

(3)齐次性：$\forall x \in R^n, \forall \alpha \in R, \|\alpha x\| = |\alpha| \|x\|.$

则$\|\cdot\|$称为R^n上的范数

范数

常见范数：$x=(x_1,x_2,\ldots,x_n)^T\in R^n$

$1-$范数 $\|x\|_1=\sum_{i=1}^{n}|x_i|;$ 表示向量x中非零元素的绝对值之和

$2-$范数 $\|x\|_2=(\sum_{i=1}^{n}x_i^2)^{1/2};$ 表示向量元素的平方和再开平方。

$\infty-$范数 $\|x\|_\infty=\max_i|x_i|;$ 用来度量向量元素的最大值

$p-$范数 $\|x\|_p=(\sum_{i=1}^{n}x_i^p)^{1/p},1\le p<+\infty;$ 向量元素绝对值的p次方和的1/p次幂

范数举例

例1:对于向量x=(1,-2,3)^T，尝试计算L1,L2,L3,L$_\infty$

$$\text{对向量}x=(1,-2,3)^T,\text{有}$$

$$\|x\|_1=|1|+|-2|+|3|=6,$$

$$\|x\|_2=\sqrt{1^2+(-2)^2+3^2}=\sqrt{14}\approx 3.74,$$

$$\|x\|_3=\sqrt[3]{1^3+(-2)^3+3^3}=\sqrt{36}\approx 3.30,$$

$$\|x\|_\infty=\max\{|1|,|-2|,|3|\}=3$$

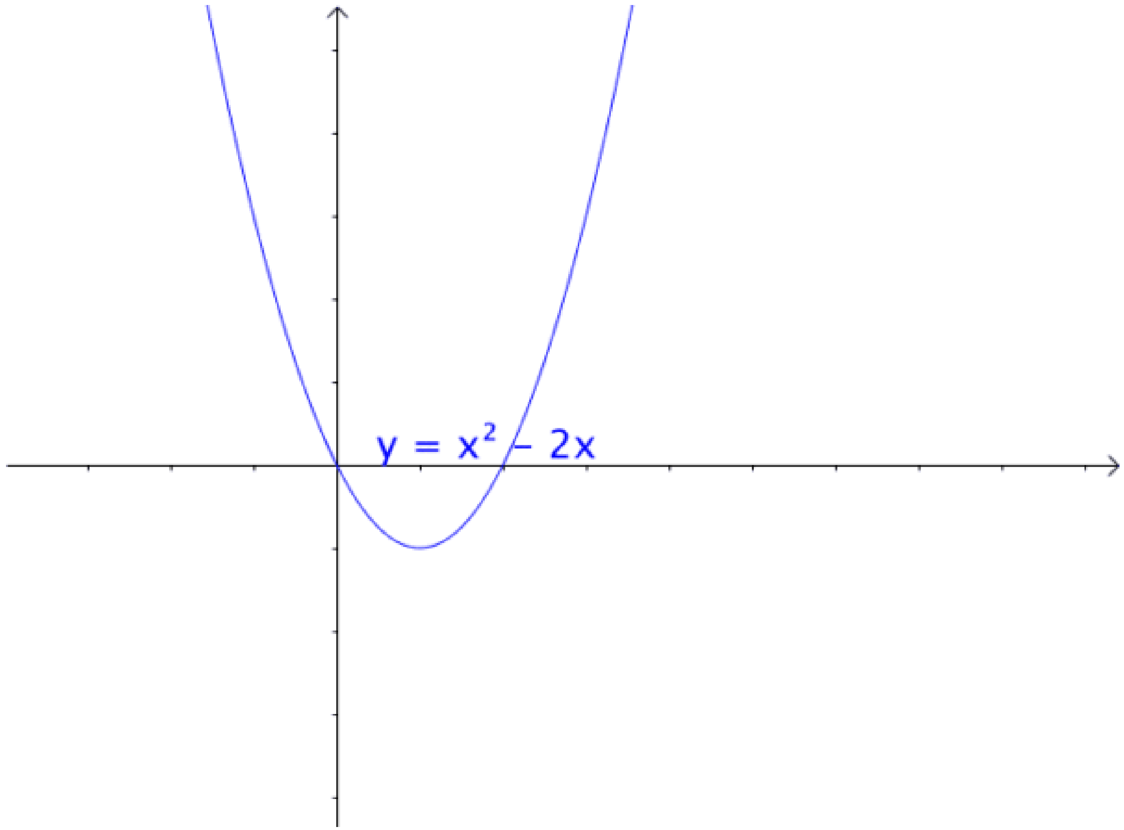


二次型

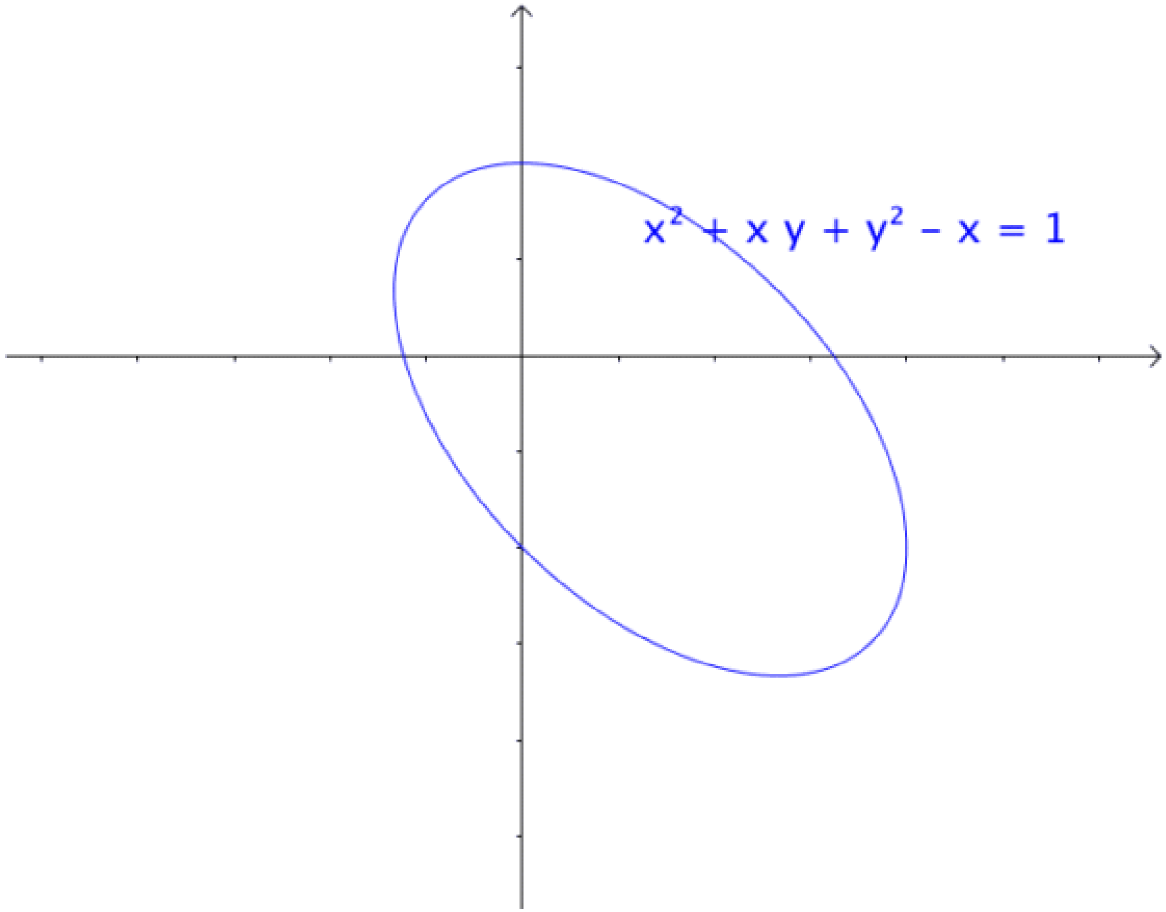
n维二次函数:

$$
\begin{aligned}
f(x_1, x_2, \cdots x_n) = {} & a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n + \\
& a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n + \\
& \cdots + \\
& a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2 \\
= {} & \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j
\end{aligned}
$$

为什么要通过矩阵来研究二次型？



为什么要通过矩阵来研究二次型？



二次型

用矩阵可表示为：

$$f(x_1,x_2,\cdots x_n)=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j=[x_1,x_2,\cdots x_n]A\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

$$=x^T Ax$$

其中二次型的系数矩阵A为对称针，$A=A^T$

其元素$a_{ij}=a_{ji}(i\neq j)$是二次型x_ix_j项系数的一半

a_{ii}是二次型x_i^2的系数

因此二次型和它的系数矩阵A是相互唯一决定的



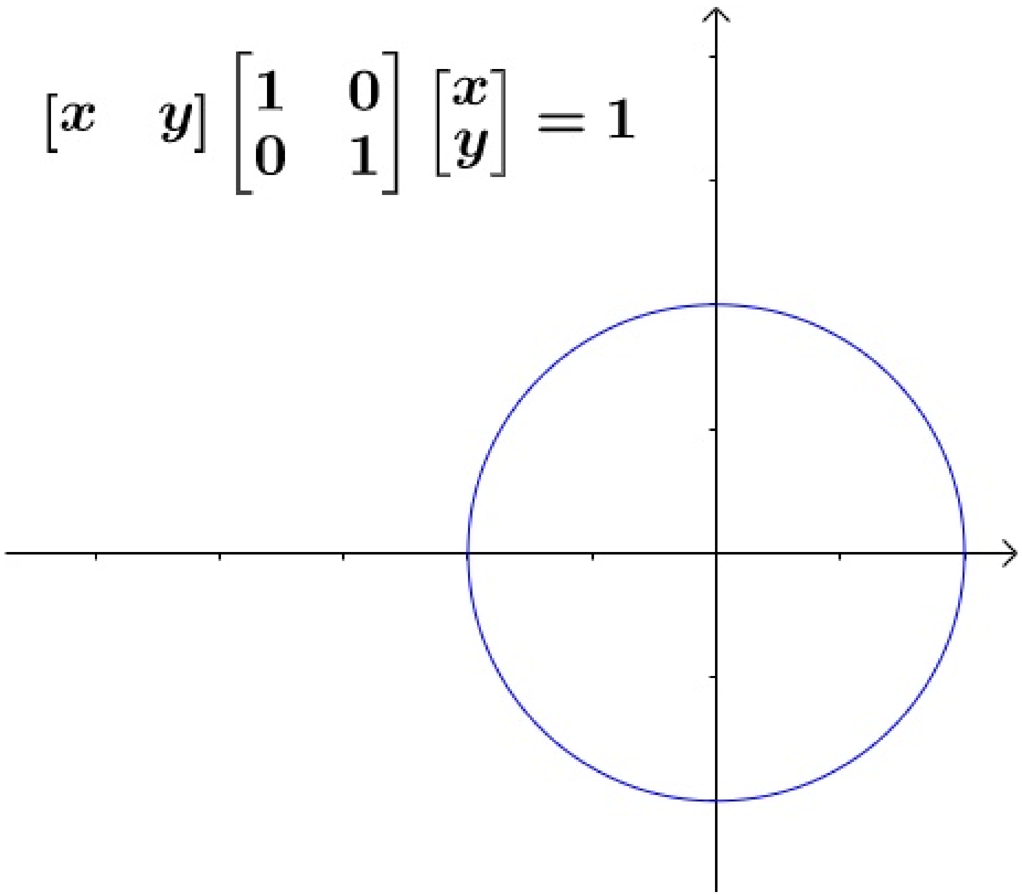
二次型及其系数矩阵

$$\left.\begin{aligned} \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1 \\ X = \begin{bmatrix} x \\ y \end{bmatrix} \\ A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \end{aligned}\right\} \Longrightarrow X^T A X$$

对称阵 $\Leftrightarrow$ 二次型

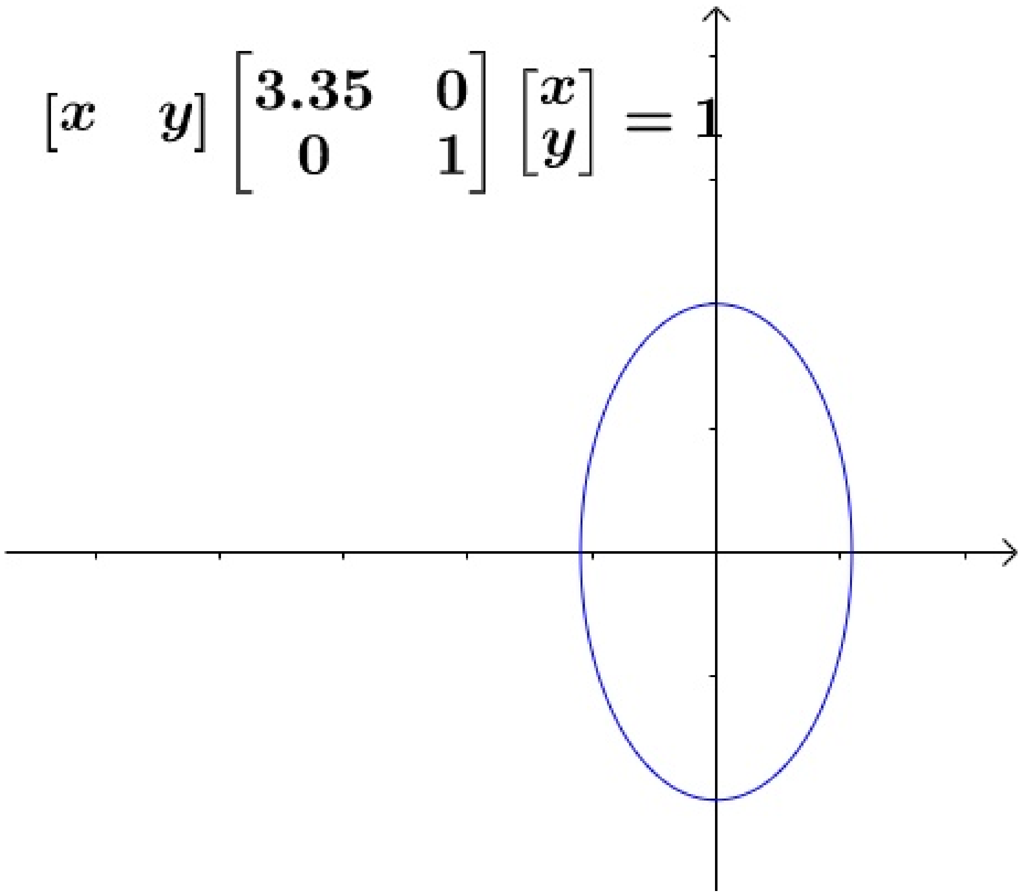
一个简单的二次型

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1$$



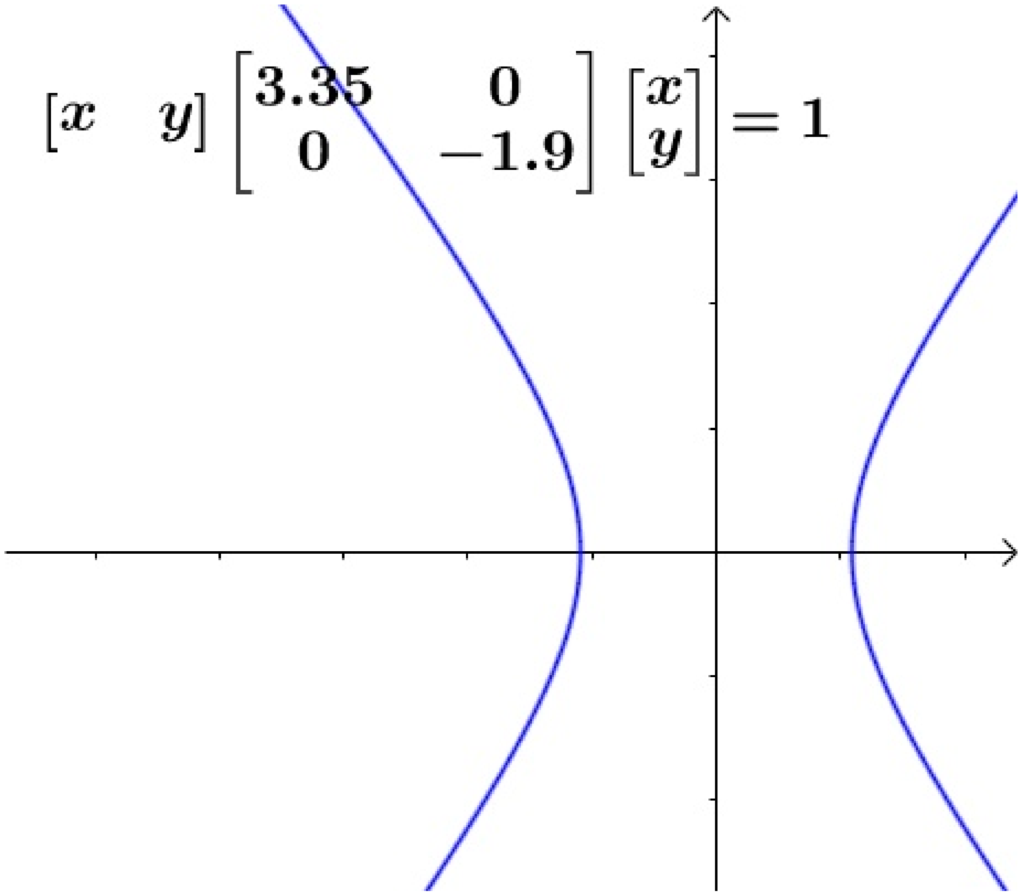
稍作改变

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 3.35 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1$$



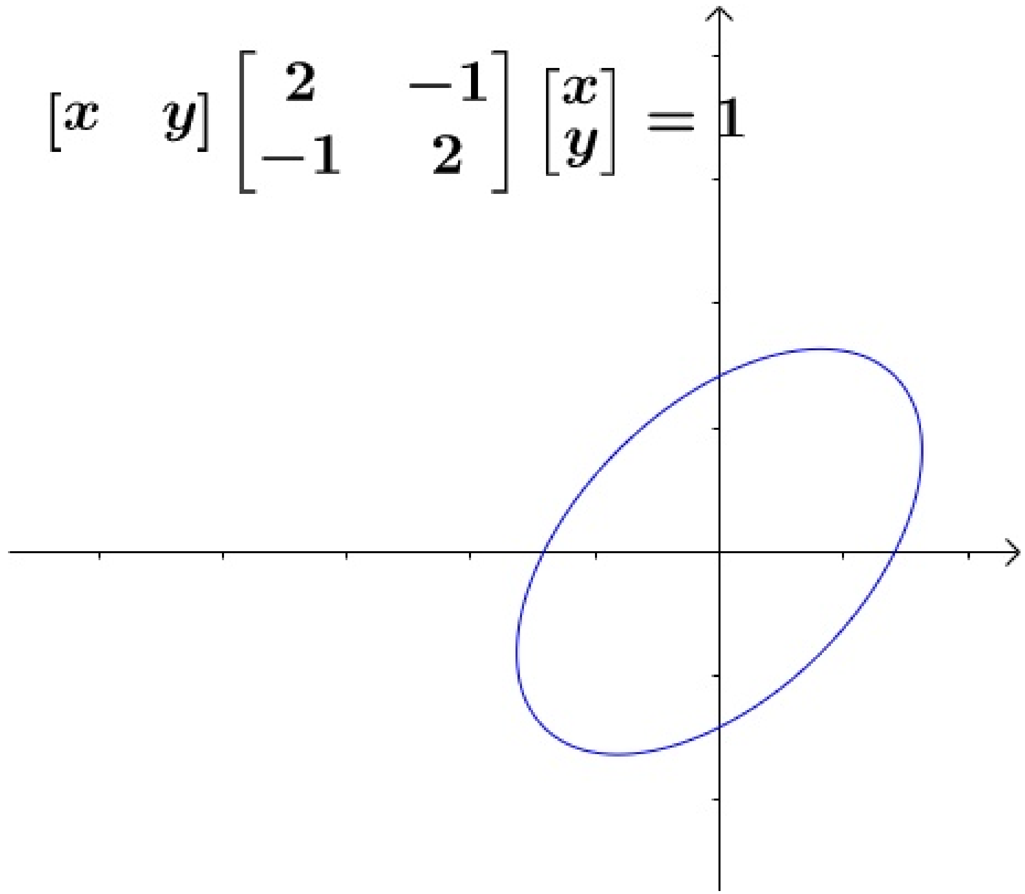
稍作改变

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 3.35 & 0 \\ 0 & -1.9 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1$$



改变副对角元素

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1$$



课后练习 Tutorial

例2 用矩阵符号表示下列二次型：

$$f\left(x_1,x_2,x_3\right)=x_1^2+4x_1x_2+4x_2^2+2x_1x_3+x_3^2+4x_2x_3$$

解：

$$f\left(x_1,x_2,x_3\right)=\left(x_1,x_2,x_3\right)\begin{bmatrix}1&2&1\\2&4&2\\1&2&1\end{bmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$$


例3. 用矩阵符号表示下列二次型：

$$f\left(x_1,x_2,x_3\right)=x_1^2+x_2^2-7x_3^2-2x_1x_2-4x_1x_3-4x_2x_3$$

解：

$$f\left(x_1,x_2,x_3\right)=\left(x_1,x_2,x_3\right)\begin{bmatrix}1 & -1 & -2\\ -1 & 1 & -2\\ -2 & -2 & -7\end{bmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}$$

正定矩阵

定义：如果二次型

$$f(x_1,x_2,\cdots x_n)=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j=X^TAX$$

对于任意一组不全为零的数$x_1,x_2,\cdots x_n$，恒有

$$f(x_1,x_2,\cdots x_n)=X^TAX>0$$

则称$f(x_1,x_2,\cdots x_n)$正定，其A为正定矩阵

类似的，

若有$f(x_1,x_2,\cdots x_n)=X^TAX\geq 0$,则A为半正定矩阵

若有$f(x_1,x_2,\cdots x_n)=X^TAX\leq 0$,则A为半负定矩阵



正定矩阵的判定

判定定理1：A正定的充要条件为A的特征值都大于零

判定定理2：A正定的充要条件为A的所有顺序主子式都大于零

设A为 $n\times n$ 阶矩阵，子式

$$D_i=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1i} \\ a_{21} & a_{22} & \cdots & a_{2i} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ii} \end{vmatrix}(i=1,2,\ldots,n)$$

称为A的i阶顺序主子式。



例4. 用顺序主子式判断矩阵是否正定

判断矩阵$A=\begin{bmatrix}4 & 2 & 1\\ 2 & 3 & 0\\ 1 & 0 & 2\end{bmatrix}$是否正定。

$$\text{解1}: 4>0,\begin{vmatrix}4 & 2\\ 2 & 3\end{vmatrix}=8>0,\begin{vmatrix}4 & 2 & 1\\ 2 & 3 & 0\\ 1 & 0 & 2\end{vmatrix}=13>0$$

A是正定的

例5. 用特征值判断矩阵是否正定

判断矩阵 $A=\begin{bmatrix}4 & 2 & 1\\ 2 & 3 & 0\\ 1 & 0 & 2\end{bmatrix}$ 是否正定。

解2: $|\lambda I - A| = 0$

$$\begin{vmatrix}\lambda-4 & 2 & 1\\ 2 & \lambda-3 & 0\\ 1 & 0 & \lambda-2\end{vmatrix} = (\lambda-4)(\lambda-3)(\lambda-2)-(\lambda-3)-4(\lambda-2)=0$$

$\lambda_1 = 1, \lambda_2 = 2.28, \lambda_3 = 5.73$

A 是正定的

例6. 判断下列二次型是否正定

1） $f(x_1,x_2,x_3,x_4)=x_1^2+3x_2^2+9x_3^2+19x_4^2-2x_1x_2+4x_1x_3-6x_2x_4-12x_3x_4+2x_1x_4$

解：

$$A=\begin{pmatrix} 1 & -1 & 2 & 1 \\ -1 & 3 & 0 & -3 \\ 2 & 0 & 9 & -6 \\ 1 & -3 & -6 & 19 \end{pmatrix} \qquad |A_1|=1>0 \qquad |A_2|=\begin{vmatrix} 1 & -1 \\ -1 & 3 \end{vmatrix}=3+1=4>0$$

$$|A_3|=\begin{vmatrix} 1 & -1 & 2 \\ -1 & 3 & 0 \\ 2 & 0 & 9 \end{vmatrix}=27+(-1)\times[0-(-9)]+2\times(0-6)=6>0$$


$$|A_4|=\begin{vmatrix}1 & -1 & 2 & 1\\ -1 & 3 & 0 & -3\\ 2 & 0 & 9 & -6\\ 1 & -3 & -6 & 19\end{vmatrix}=1\times\begin{vmatrix}3 & 0 & -3\\ 0 & 9 & -6\\ -3 & -6 & 19\end{vmatrix}+(-1)\times\begin{vmatrix}1 & 0 & -3\\ 2 & 9 & -6\\ 1 & -6 & 19\end{vmatrix}$$

$$+2\times\begin{vmatrix}-1 & 3 & -3\\ 2 & 0 & -6\\ 1 & -3 & 19\end{vmatrix}+1\times\begin{vmatrix}-1 & 3 & 0\\ 2 & 0 & 9\\ 1 & -3 & -6\end{vmatrix}=24>0$$

该二次型是正定的

例6. 判断下列二次型是否正定

$$f(x_1,x_2,x_3)=5x_1^2+x_2^2+5x_3^2+4x_1x_2-8x_1x_3-4x_2x_3$$

解：

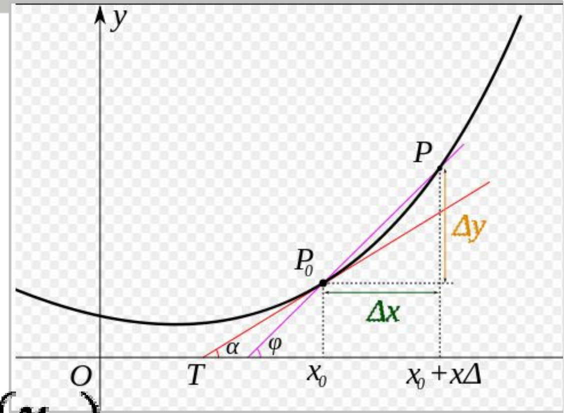
$$A=\begin{pmatrix}5 & 2 & -4\\ 2 & 1 & -2\\ -4 & -2 & 5\end{pmatrix} \qquad |A_1|=5>0 \qquad |A_2|=\begin{vmatrix}5 & 2\\ 2 & 1\end{vmatrix}=1>0$$

$$|A_3|=\begin{vmatrix}5 & 2 & -4\\ 2 & 1 & -2\\ -4 & -2 & 5\end{vmatrix}=5\times(1\times5-4)+2\left[(-2)\times(-4)-2\times5\right]$$

$$-4\left[2\times2-1\times(-4)\right]=1>0$$

该二次型是正定的

方向导数与梯度

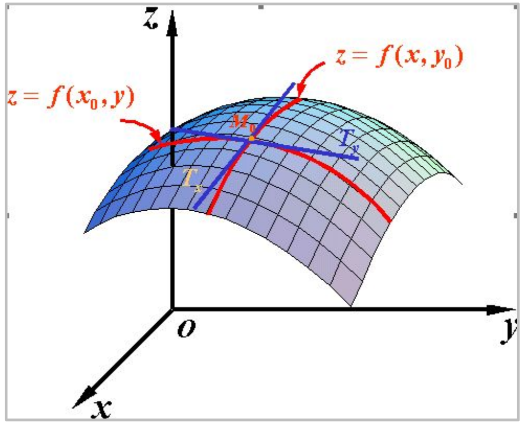


导数

函数曲线上的切线斜率；函数在某点的变化率。

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

偏导数



多元函数在坐标轴正方向上的变化率

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

- 偏导数$f_x(x,y)$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$f_y(x,y)$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

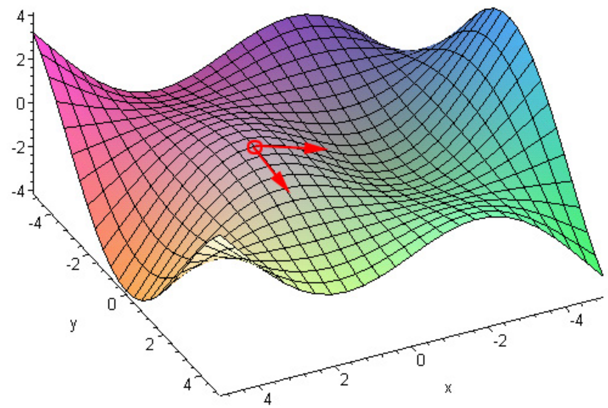
方向导数与梯度

Directional Derivatives : Along the Axes…

方向导数用于研究函数沿各个不同方向时函数的变化率，是标量。

$$\frac{\partial f(x,y)}{\partial y}$$

$$\frac{\partial f(x,y)}{\partial x}$$



$$\frac{\partial f}{\partial l}\Big|_{(x_0,y_0)} = \lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}} \frac{f(x_0+\Delta x, y_0+\Delta y) - f(x_0,y_0)}{\sqrt{\Delta x^2+\Delta y^2}}$$

$$\frac{\partial f}{\partial l}\Big|_{(x_0,y_0)} = f_x'(x_0,y_0)cos\alpha + f_y'(x_0,y_0)cos\beta$$

方向导数

定义：设$f:R^n \to R^1$在点x_0处可微，P是固定不变的非零向量，e是方向P上的单位向量，则称极限

$$\frac{\partial f(x_0)}{\partial p} = \lim_{t \to 0^+} \frac{f(x_0 + te) - f(x_0)}{t}$$

为函数$f(x)$在点x_0处沿P方向的方向导数，记为$\frac{\partial f(x_0)}{\partial p}$

上升方向与下降方向

定义：设$f:R^n \to R^1$是连续函数，$x_0 \in R^n, P \in R^n$，且$P \neq 0$，若存在$\delta > 0$，当$t \in (0,\delta)$时都有$f(x_0 + tP) < f(x_0)$，则称P为f在x_0处的下降方向。若$f(x_0 + tP) > f(x_0)$，则称P为f在x_0处的上升方向。

由上两个定义可以得出如下结论：

若$\dfrac{\partial f(x_0)}{\partial \mathrm{P}} < 0$，则$f(x)$在$x_0$处沿P方向是下降的；

若$\dfrac{\partial f(x_0)}{\partial \mathrm{P}} > 0$，则$f(x)$在$x_0$处沿P方向是上升的；

梯度

定义：以函数$f(x)$的n个偏导数为分量的向量称为$f(x)$在点x 处的梯度

记为

$$\nabla f(x)=\left[\frac{\partial f(x)}{\partial x_1},\frac{\partial f(x)}{\partial x_2},\cdots,\frac{\partial f(x)}{\partial x_n}\right]^T$$

梯度也可以称为$f(x)$关于向量x的一阶导数。



梯度与方向导数之间的关系

定理：设$f:R^n \to R^1$在x_0处可微，则

$$\frac{\partial f(x_0)}{\partial P} = \nabla f(\mathrm{x}_0)^T e$$

其中e是P方向上的单位向量。

$$\left|\frac{\partial f(x_0)}{\partial P}\right| = \left|\nabla f(\mathrm{x}_0)^T e\right|$$

$$= \left|\nabla f(\mathrm{x}_0)\right| \cdot \left|\cos(\nabla f(\mathrm{x}_0), e)\right| \le \left|\nabla f(\mathrm{x}_0)^T\right|$$

梯度与方向导数之间的关系

推论：

1. 若$\nabla f(x_0)^T P<0$,则P的方向是函数$f(x)$在x_0处的下降方向；

2. 若$\nabla f(x_0)^T P>0$,则P的方向是函数$f(x)$在x_0处的上升方向；

证明1： $\nabla f(x_0)^T P<0$， $\nabla f(\mathrm{x}_0)^T e=\nabla f(x_0)^T \dfrac{P}{\|\mathrm{P}\|}$

$$\Rightarrow \nabla f(\mathrm{x}_0)^T e<0$$

由定理$\dfrac{\partial f(x_0)}{\partial \mathrm{P}}=\nabla f(\mathrm{x}_0)^T e$

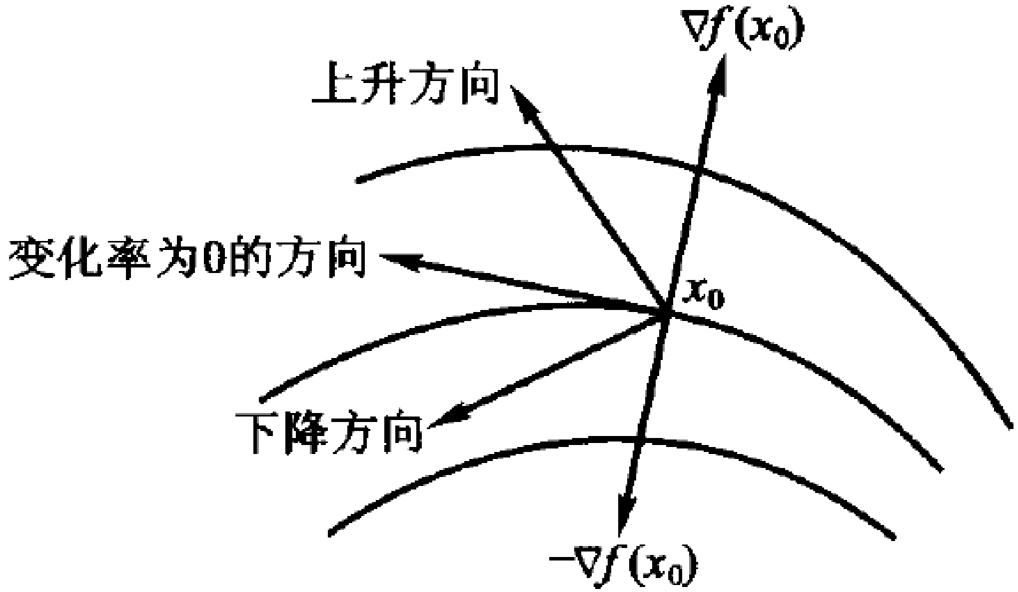
$$\Rightarrow \frac{\partial f(x_0)}{\partial \mathrm{P}}<0$$

得证



关于梯度与方向导数的结论

1. 梯度方向是函数值的最速上升方向；
2. 函数在与其梯度正交的方向上变化率为零；
3. 函数在与其梯度成锐角的方向上是上升的，而在与其梯度成钝角的方向上是下降的；
4. 梯度反方向是函数值的最速下降方向。



例7.

试求目标函数$f(\mathrm{x})=\mathrm{x}_1^2+x_2^2+1$在店$\mathrm{x}_0=[0,3]^T$处的最速下降方向，并求沿这个方向移动一个单位后新点的目标函数值。

解：$\dfrac{\partial \mathrm{f}}{\partial x_1}=2x_1,\dfrac{\partial \mathrm{f}}{\partial x_2}=2x_2$

最速下降方向是

$$-\nabla f(\mathrm{x}_0)=-\begin{bmatrix}2x\\2x_2\end{bmatrix}_{x_1=0,x_2=3}=\begin{bmatrix}0\\-6\end{bmatrix}$$

此方向的单位向量为

$$\mathrm{e}=\frac{-\nabla f(\mathrm{x}_0)}{\left\|-\nabla f(\mathrm{x}_0)\right\|}=\begin{bmatrix}0\\-1\end{bmatrix}$$

（接下页）



例题

试求目标函数$f(\mathrm{x})=\mathrm{x}_1^2+x_2^2+1$在店$\mathrm{x}_0=[0,3]^T$处的最速下降方向，并求沿这个方向移动一个单位后新点的目标函数值。

（接上页）

故新点是

$$x_1=x_0+e=\begin{bmatrix}0\\3\end{bmatrix}+\begin{bmatrix}0\\-1\end{bmatrix}=\begin{bmatrix}0\\2\end{bmatrix}$$

对应目标函数值为$f(\mathrm{x}_1)=0^2+2^2+1=5$

原函数值为$f(\mathrm{x}_0)=0^2+3^2+1=10$

Hesse矩阵与泰勒展开

Hesse矩阵

定义：$f:R^n \to R^1$，$x_0 \in R^n$，如果f在点x_0处对于自变量x的各分量的二阶偏导数

$$\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}, (i, j = 1, 2, \cdots n)$$

都存在，则称函数f在点x_0处二阶可到，并且称矩阵

$$\nabla^2 f(x_0) = \begin{bmatrix} \frac{\partial^2 f(x_0)}{\partial x_1^2} & \frac{\partial^2 f(x_0)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x_0)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x_0)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x_0)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x_0)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(x_0)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x_0)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x_0)}{\partial x_n^2} \end{bmatrix}$$

是f在店x_0处的$Hesse$矩阵。

Hesse矩阵与泰勒展开

例8. 求目标函数f(x)=$x_1^4+2x_2^3+3x_3^2-x_1^2x_2+4x_2x_3-x_1x_3^2$的梯度和Hesse矩阵

解：

$$\frac{\partial f}{\partial x_1}=4x_1^3-2x_1x_2-x_3^2$$

$$\frac{\partial f}{\partial x_2}=6x_2^2-x_1^2-4x_3$$

$$\frac{\partial f}{\partial x_3}=6x_3+4x_2-2x_1x_3$$

所以$\nabla f(\mathrm{x})=\begin{bmatrix}4x_1^3-2x_1x_2-x_3^2\\6x_2^2-x_1^2-4x_3\\6x_3+4x_2-2x_1x_3\end{bmatrix}$

Hesse矩阵与泰勒展开

例题

求目标函数f(x)=$x_1^4+2x_2^3+3x_3^2-x_1^2x_2+4x_2x_3-x_1x_3^2$的梯度和Hesse矩阵

解：

$$\frac{\partial^2 f}{\partial x_1^2}=12x_1^2-2x_2,\frac{\partial^2 f}{\partial x_1 x_2}=-2x_1,\frac{\partial^2 f}{\partial x_1 x_3}=-2x_3$$

$$\frac{\partial f}{\partial x_2^2}=12x_2,\frac{\partial^2 f}{\partial x_2 x_3}=4,\frac{\partial^2 f}{\partial x_3^2}=6-2x_1$$

所以$\nabla^2 f(\mathrm{x})=\begin{bmatrix} 12x_1^2-2x_2 & -2x_1 & -2x_3 \\ -2x_1 & 12x_2 & 4 \\ -2x_3 & 4 & 6-2x_1 \end{bmatrix}$



泰勒展开

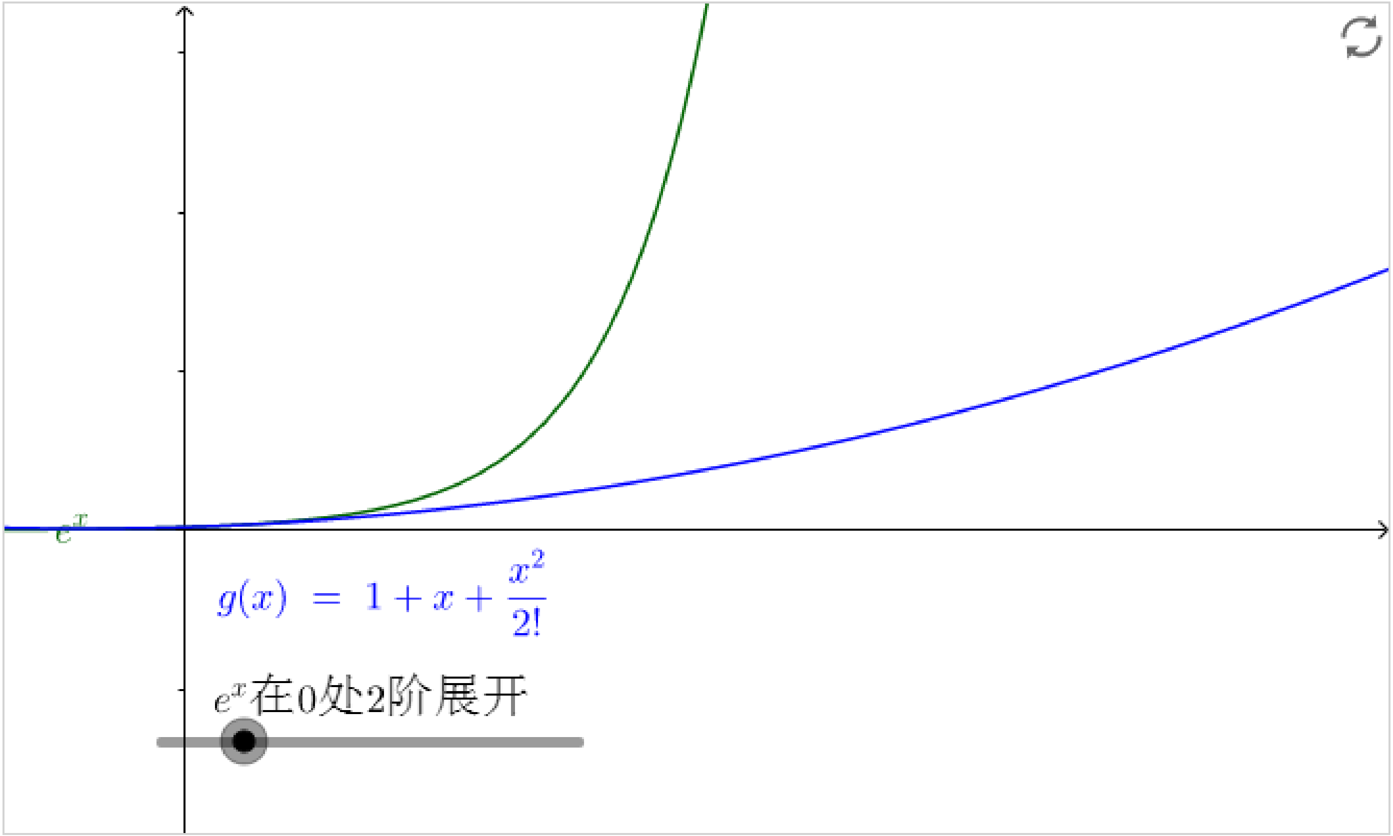
二阶泰勒展开式：

设函数f：$R^n \to R$具有二阶连续偏导数，则

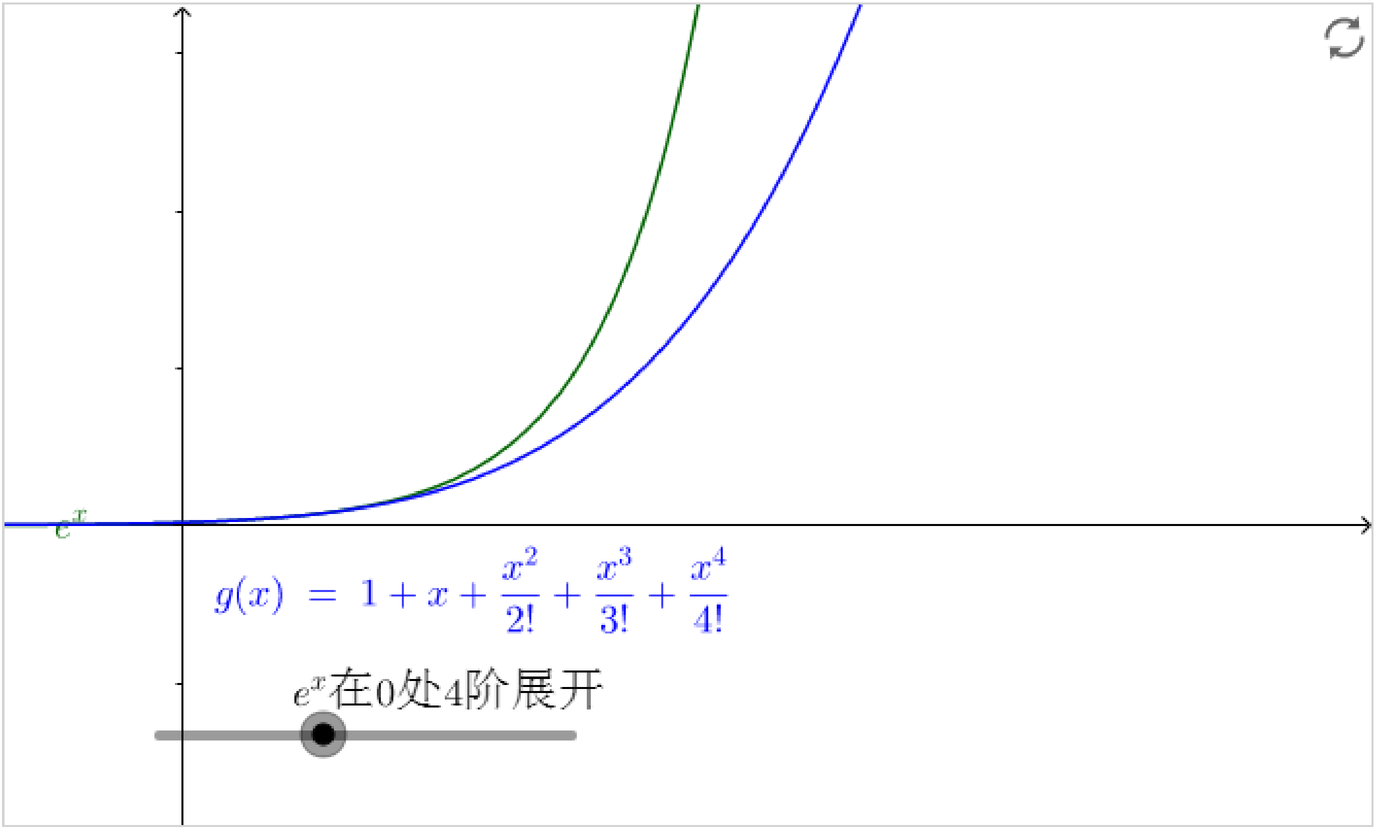
$$f(X+p)=f(X)+\nabla f(X)^T p+\frac{1}{2}p^T\nabla^2 f(X)^T p+o(\|p\|^2)$$

其中$o(\|p\|^2)$ 当$\|p\|^2\to 0$时，是关于$\|p\|^2$的高阶无穷小量。

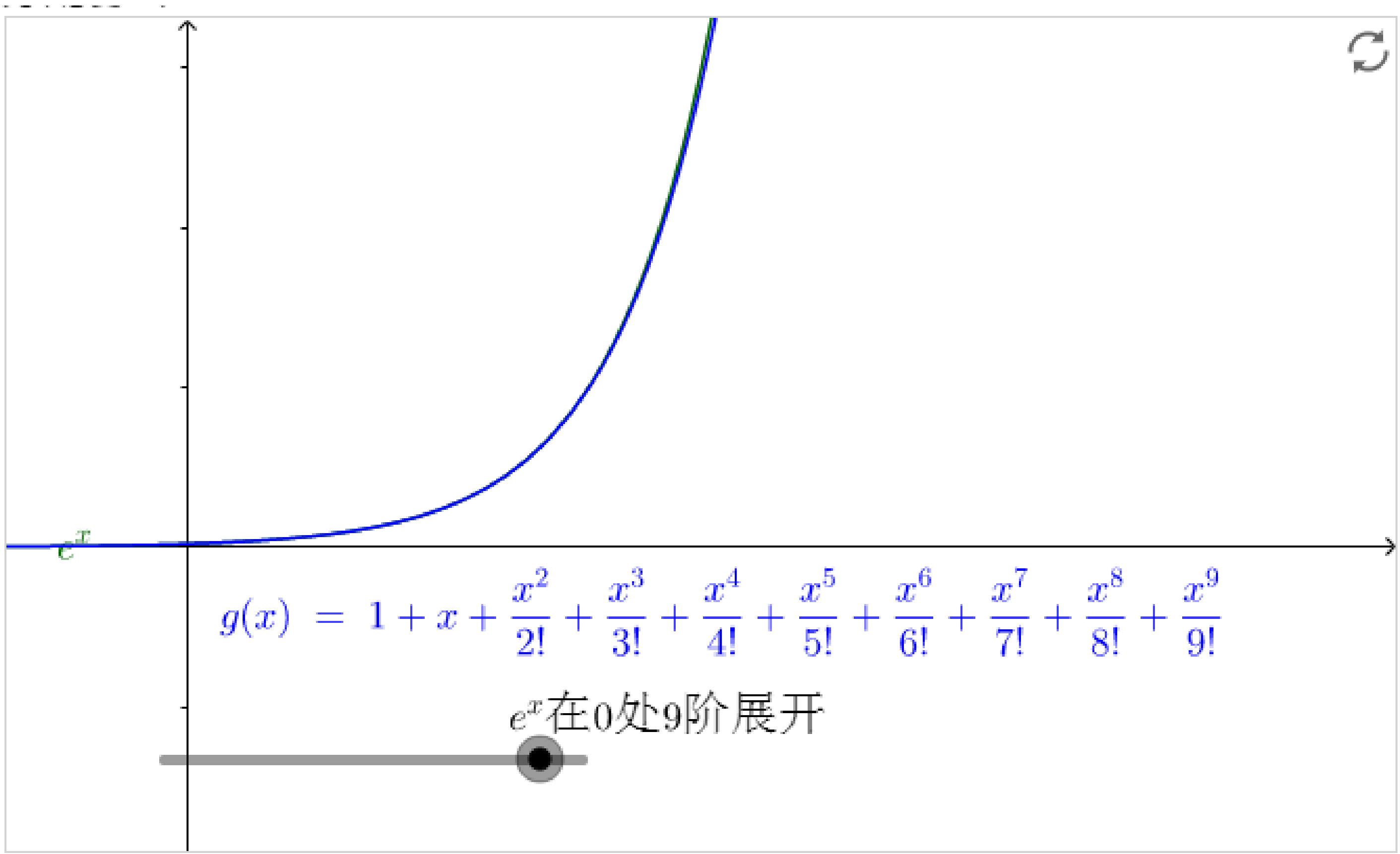
泰勒展开举例



泰勒展开举例



泰勒展开举例


e^x
$g(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \frac{x^6}{6!} + \frac{x^7}{7!} + \frac{x^8}{8!} + \frac{x^9}{9!}$
e^x在0处9阶展开

证明：梯度方向是函数的最速上升方向

证明：设函数f：$R^n \rightarrow R$在x_0处连续可导，其任意方向导数

$$\frac{\partial f(x_0)}{\partial p} = \lim_{t\rightarrow 0^+} \frac{f(x_0+te)-f(x_0)}{t}$$

可将论题等价于以下优化问题：

$$\max \frac{\partial f(x_0)}{\partial p}$$

$$s.t. \|p\| = 1$$

证明：梯度方向是函数的最速上升方向

证明：设函数f：$R^n \to R$在x_0处连续可导，其任意方向导数

$$\frac{\partial f(x_0)}{\partial p} = \lim_{t \to 0^+} \frac{f(x_0 + te) - f(x_0)}{t}$$

（接上页）

对$f(x_0 + te)$在x_0处进行泰勒展开：

$$f(x_0 + te) = f(x_0) + \nabla f(x_0)^T te + \frac{t^2}{2} e^T \nabla^2 f(x_0)^T e + o(\|e\|^2)$$

证明：梯度方向是函数的最速上升方向

证明：设函数f：$R^n \to R$在x_0处连续可导，其任意方向导数

$$\frac{\partial f(x_0)}{\partial p} = \lim_{t\to 0^+}\frac{f(x_0+te)-f(x_0)}{t}$$

（接上页）

将展开带入方向导数$\frac{\partial f(x_0)}{\partial p}$表达式：

$$\frac{f(x_0+te)-f(x_0)}{t} = \nabla f(x_0)^T e \ + \frac{t}{2} e^T \nabla^2 f(x_0)^T e \ + o(\|\mathrm{e}\|^2)$$

$$\lim_{t\to 0^+}\frac{f(x_0+te)-f(x_0)}{t} = \nabla f(x_0)^T e = \nabla f(x_0)^T \frac{\mathrm{p}}{\|p\|} = \nabla f(x_0)^T p$$

证明：梯度方向是函数的最速上升方向

证明：设函数f：$R^n \to R$在x_0处连续可导，其任意方向导数

$$\frac{\partial f(x_0)}{\partial p} = \lim_{t \to 0^+} \frac{f(x_0 + te) - f(x_0)}{t}$$

（接上页）

可以将优化问题等价写为：

$$\max \nabla f(x_0)^T p$$

$$s.t. \|p\| = 1$$

证明：梯度方向是函数的最速上升方向

证明：设函数f：$R^n \to R$在x_0处连续可导，其任意方向导数

$$\frac{\partial f(x_0)}{\partial p} = \lim_{t \to 0^+} \frac{f(x_0 + te) - f(x_0)}{t}$$

（接上页）

向量的内积可以写为如下形式：

$\left\|\nabla f(x_0)^T\right\| \|p\| \cos(\theta)$,其中$\theta$为$\nabla f(x_0)^T$与p的夹角

前两项为常数，优化问题的最大值在$\cos(\theta)=1$即$\theta=0$处取得

因此$p=\dfrac{\nabla f(x_0)^T}{\left\|\nabla f(x_0)^T\right\|}$，为梯度方向。

课后练习 Tutorial #3

方向导数和梯度的课外阅读材料

https://opentextbc.ca/calculusv3openstax/chapter/directional-derivatives-and-the-gradient/

西安電子科技大學
XIDIAN UNIVERSITY

IPIU
智能感知与图像理解

谢谢！