最优化理论与方法

讲课人：吕茂斌
模式识别与智能系统研究所
北京理工大学自动化学院
lumaobin@bit.edu.cn

数学概念

数学基础知识

- 向量范数
- 矩阵范数
- 序列的极限
- 聚点
- Cauchy序列
- 开集、闭集、紧集
- Taylor展开式
- Jacobi矩阵
- 链式法则
- 连续
- 梯度
- Hesse矩阵
- 函数梯度

零、欧氏空间

欧氏空间: $\mathbf{R}^n$是全体 n 维实向量构成的集合

x 是空间$\mathbf{R}^n$维中的点，记为$x\in \mathbf{R}^n$

0 表示零向量，$(\ )^T$ 记为转置

令 $x,y\in \mathbf{R}^m$， $x+y$表示 x 和 y 所有分量的和

$x\leq y$ 表示 x 的每个分量均小于 y 的分量（$x_i \leq y_i$）

x,y的**内积**定义 $\langle x, y\rangle = x^\top y = \sum_{i=1}^{n} x_i y_i$

一、向量范数

若实值函数$\|.\|: \mathbf{R}^n \to \mathbf{R}$满足下列条件：

（1） $\|x\| \geq 0$， $\forall x \in \mathbf{R}^n$； $\|x\| = 0$ 当且仅当$x = 0$；

（2） $\|\alpha x\| = |\alpha| \|x\|$， $\forall \alpha \in \mathbf{R}, x \in \mathbf{R}^n$ ；

（3） $\|x + y\| \leq \|x\| + \|y\|$， $\forall x, y \in \mathbf{R}^n$.

则称$\|\cdot\|$为向量范数. 其中$\mathbf{R}^n$表示n维向量空间.

常用的向量范数

L_1范数 $$\|x\|_1 = \sum_{j=1}^{n} |x_j|$$

L_2范数 $$\|x\|_2 = \left(\sum_{j=1}^{n} {x_j}^2 \right)^{\frac{1}{2}}$$

L_∞范数 $$\|x\|_\infty = \max_j |x_j|$$

L_p范数 对于$1 \leq p < \infty$

$$\|x\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{\frac{1}{p}}$$

其中$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T \in \mathbf{R}^n$

一、向量范数

范数的概念：非负、齐次、三角不等式

范数定义了更一般的“距离”，更易推广

可以用简单的符号表达复杂的特性

直觉上正确的性质范数不一定满足

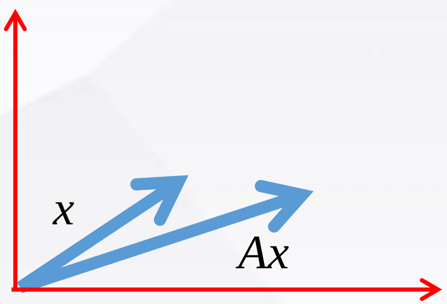
例子：“单位圆”是圆吗？

单位圆：给定范数$\|\ \|$，集合$\{x:\|x\|\le 1\}$

二、矩阵范数

回顾：向量的拉伸

给定二维向量 x：横坐标增加一倍，纵坐标不变



实现方法：左乘 $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$

给定 x，左乘A最大有可能拉伸几倍？

$$\|A\| = \max_{x\in\mathbb{R}^2, x\neq 0} \frac{\|Ax\|}{\|x\|}$$

二、矩阵范数

定义：设$\boldsymbol{A}$为$m\times n$矩阵，$\|\cdot\|_\alpha$ 是$\mathbf{R}^m$上向量范数，$\|\cdot\|_\beta$是$\mathbf{R}^n$上向量范数，定义矩阵范数

$$\|\boldsymbol{A}\|=\max_{\|\boldsymbol{x}\|_\beta=1}\|\boldsymbol{A}\boldsymbol{x}\|_\alpha$$

矩阵范数具有以下性质：

（1）$\|\boldsymbol{A}\boldsymbol{x}\|_\alpha\le\|\boldsymbol{A}\|\,\|\boldsymbol{x}\|_\beta$；

（2）$\|\lambda\boldsymbol{A}\|=|\lambda|\,\|\boldsymbol{A}\|$；

（3）$\|\boldsymbol{A}+\boldsymbol{B}\|\le\|\boldsymbol{A}\|+\|\boldsymbol{B}\|$；

（4）$\|\boldsymbol{A}\boldsymbol{D}\|\le\|\boldsymbol{A}\|\|\boldsymbol{D}\|$.

其中$\boldsymbol{A},\boldsymbol{B}$为$m\times n$矩阵，$\boldsymbol{D}$是$n\times p$矩阵，$\lambda$为实数，$x\in\mathbf{R}^n$.

常用矩阵范数

（1） $$\|\boldsymbol{A}\|_1=\max_j \sum_{i=1}^{m}|a_{ij}|$$

（2） $$\|\boldsymbol{A}\|_2=\sqrt{\lambda_{A^{\mathrm{T}}A}}$$ 谱范数

（3） $$\|\boldsymbol{A}\|_\infty=\max_i \sum_{j=1}^{n}|a_{ij}|$$

$\lambda_{A^{\mathrm{T}}A}$为$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$的最大特征值

三、序列的极限

定义：设 $\{x^{(k)}\}$ 是$\mathbf{R}^n$中一个向量序列，$\bar{x}\in\mathbf{R}^n$，如果对每个任给的$\varepsilon>0$存在正整数K_ε，使得当$k>K_\varepsilon$时就有$\|x^{(k)}-\bar{x}\|<\varepsilon$，则称序列收敛到$\bar{x}$或者称序列以$\bar{x}$为极限(limit)，记作

$$\lim_{k\to\infty} x^{(k)}=\bar{x}.$$

某个序列的子序列是从最初序列通过去除某些元素但不破坏余下元素的相对位置（在前或在后）而形成的新序列。

序列若存在极限，则任何子序列有相同的极限，即序列的极限是惟一的.

四、聚点

定义：设 $\{x^{(k)}\}$ 是$\mathbf{R}^n$中一个向量序列，如存在一个子序列 $\{x^{(k_j)}\}$ 使

$$\lim_{k_j \to \infty} x^{(k_j)} = \hat{x}.$$

则称$\hat{x}$是序列 $\{x^{(k)}\}$ 的一个聚点(accumulation point).

如无穷序列有界，即存在正数M，使得对所有k均有$\|x^{(k)}\| \le M$，则该序列必有聚点.

五、Cauchy序列

定义: 设 $\{x^{(k)}\}$ 是$\mathbf{R}^n$中一个向量序列,如对任意给定的$\varepsilon>0$，总存在正整数K_ε,使得当$m,l>K_\varepsilon$时,就有$\|x^{(m)}-x^{(l)}\|<\varepsilon$,则 $\{x^{(k)}\}$ 称为Cauchy序列 .

定理: 设 $\{x^{(j)}\}\subset\mathbf{R}^n$为Cauchy序列, 则 $\{x^{(j)}\}$ 的聚点必为极限点.

思考题： 若 $\{x^{(j)}\}\subset\mathbf{R}^n$收敛, 它一定是Cauchy序列吗？

六、开集、闭集、紧集

A是B的子集，记为$A \subset B$: A中所有元素都属于B

A和B的交集记为$A \cap B$ ：同时属于A和B的元素组成集合

A和B的并集记为$A \cup B$ ：同时属于A或B的元素组成集合

设$S \subset \mathbf{R}^n$, S的补集是$\{x \in \mathbf{R}^n | x \notin S\}$，记为$S^C$

如果S中每个收敛序列的极限均属于S, 则称S为闭集.

如果对每一点 $\hat{x} \in S$, 存在正数ε, 使得$\hat{x}$的ε邻域$N(\hat{x}, \varepsilon) =$

$\{x | \| x - \hat{x} \| < \varepsilon\} \subset S$, 则称$S$为开集.

如果S是有界闭集, 则称S为紧集.

六、开集、闭集、紧集

$\mathbf{R}^n$中开集和闭集的关键：是否包含边界点。

例子：

$\mathbf{R}^n$既是开集也是闭集；

$\{x \in \mathbf{R}^n \mid \| x - \hat{x} \| < \varepsilon\}$是开集；

$\{x \in \mathbf{R}^n \mid \| x - \hat{x} \| \leq \varepsilon\}$是闭集也是紧集；

$\{x \in \mathbf{R}^n \mid Ax = b\}$是闭集。

思考题：$\{x \in \mathbf{R}^n \mid Ax = b\}$有可能是开集吗？有可能是紧集吗？

六、开集、闭集、紧集

$\mathbf{R}^n$中的一些性质：

开集的补集闭集，反之亦然；

有限个开集的交集是开集，任意个开集的并集是开集；

有限个闭集的并集是闭集，任意个闭集的交集是闭集。

七、连续

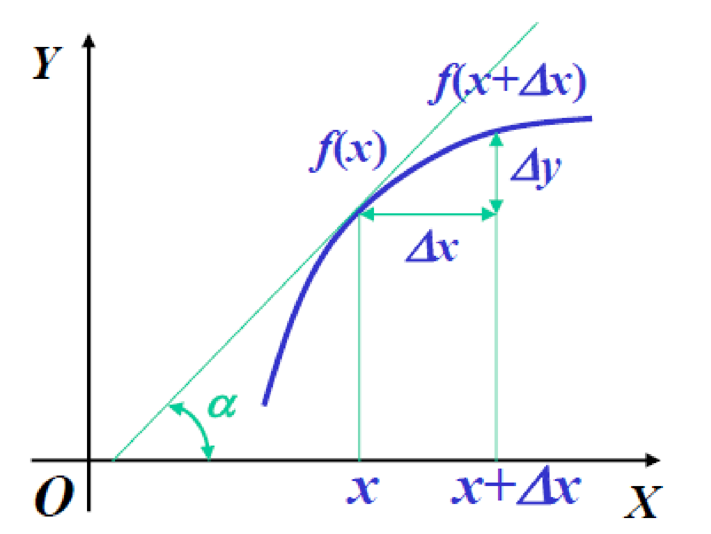
设$S \subset \mathbf{R}^n$非空,$f(x)$为定义在S上的实函数

- 在c点连续：对任意的$\varepsilon > 0$，存在一个$\delta > 0$使得只要$\|x-c\| \le \delta$, 就有$\|f(x)-f(c)\| \le \varepsilon$.
- 连续函数：如果f在每一点$x \in S$连续,则称f在S上连续,记作$f \in C(S)$.
- 连续可微：再设S为开集,如果在每一点$x \in S$，对所有$j=1,\cdots,n$, 各偏导数存在且连续,则称f在开集S上连续可微,记作$f \in C^1(S)$.
- 二次连续可微：如果在每一点$x \in \mathrm{S}$对所有$i=1,\cdots,n$和$j=1,\cdots,n$, 各二阶偏导数存在且连续,则称f在开集S上二次连续可微,记作$f \in C^2(S)$.

八、梯度

定义：考虑(单变量)函数$y = f(x)$。当自变量x 在点x 有一改变量Δx时，函数y 相应地有一改变量$\Delta y = f(x+\Delta x) - f(x)$，那么当$\Delta x$ 趋于零时，若比值$\Delta y/\Delta x$ 的极限存在，则称这个极限为函数$f(x)$ 在点x 的**导数**，记

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



从几何直观的角度看，函数 $f(x)$ 的导数是函数$y = f(x)$ 表示的曲线在点x 的切线的斜率，即 $f'(x) = \mathrm{tg}\alpha$，这里$\alpha$是曲线在点$x$ 处的切线与X-轴的夹角。

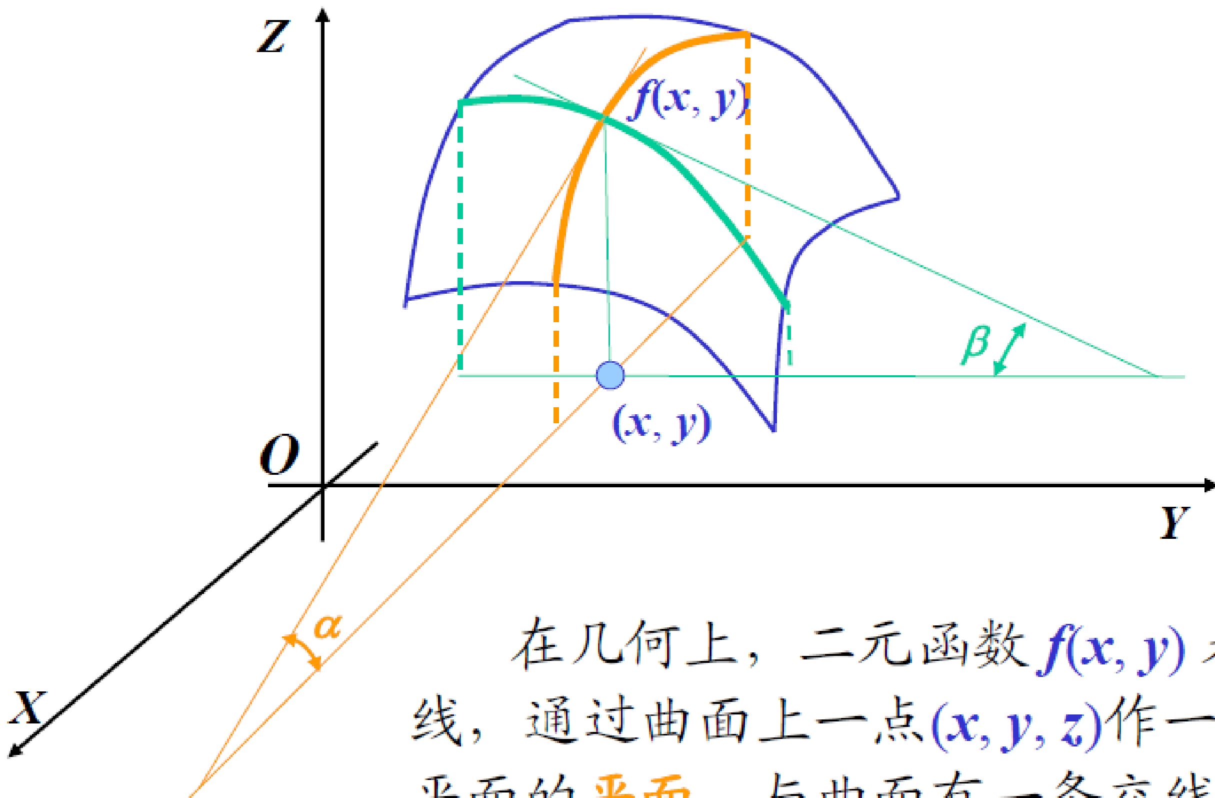
八、梯度

考虑多变量函数 $z = f(x, y)$。当变量x 在点x 有一改变量Δx，而变量y 保持不变时，函数z 相应地有一改变量$\Delta z = f(x+\Delta x, y)-f(x,y)$，那么当$\Delta x$ 趋于零时，若比值$\Delta z/\Delta x$ 的极限存在，则称这个极限为函数 $f(x,y)$ 在点(x, y) 关于变量x 的偏导数，记

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

多变量函数的偏导数可以按照单变量函数的微分法则求出，只需要对所讨论变量求导数，其余的变量都看作常数即可。

八、梯度

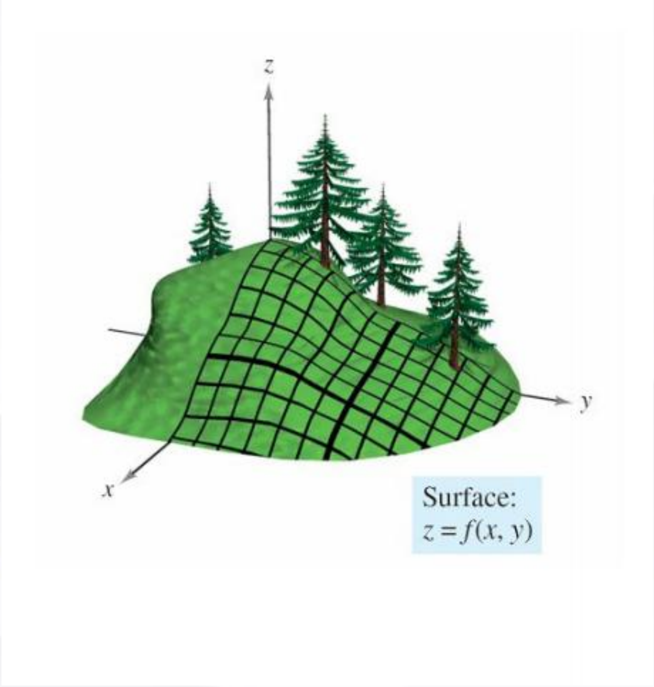


在几何上，二元函数 $f(x, y)$ 表示一个曲线，通过曲面上一点 (x, y, z) 作一平行于 **OXZ** 平面的**平面**，与曲面有一条交线，$\partial z/\partial x$ 就是这条曲线在该点的切线与 X-轴正向夹角 α 的正切，即 $\partial z/\partial x = \mathrm{tg}\,\alpha$。

八、梯度

函数f在x处的梯度为n维列向量：

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



九、Hesse矩阵

函数f在x处的Hesse矩阵： 向量变量函数的曲率

$$[\nabla^2 f(x)]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, 1 \le i, j \le n$$

$n \times n$ 矩阵

十、二次函数的梯度和Hesse阵

$$f(x) = \frac{1}{2} x^{\mathrm{T}} A x + b^{\mathrm{T}} x + c \, .$$

- 梯度：$\nabla f(x) = Ax + b$
- Hesse矩阵： $\nabla^2 f(x) = A$

$A: n \times n$矩阵

$b: n \times 1$

$c: 1 \times 1$

十一、*Taylor*展开式

- 开集$S \subset \mathbf{R}^n$上，$f \in C^1(S)$，给定点$\bar{x} \in S$，则f在点$\bar{x}$的一阶*Taylor*展开式为

$$f(x)=f(\bar{x})+\nabla f(\bar{x})^{\mathrm{T}}(x-\bar{x})+o(\|x-\bar{x}\|),$$

其中$o(\|x-\bar{x}\|)$当$\|x-\bar{x}\| \to 0$时,关于$\|x-\bar{x}\|$是高阶无穷小量．

- 开集$S \subset \mathbf{R}^n$上，$f \in C^2(S)$，给定点$\bar{x} \in S$，则f在点$\bar{x}$的二阶*Taylor*展开式为

$$f(x)=f(\bar{x})+\nabla f(\bar{x})^{T}(x-\bar{x})+\frac{1}{2}(x-\bar{x})^{\mathrm{T}}\nabla^2 f(\bar{x})(x-\bar{x})+o(\|x-\bar{x}\|^2),$$

其中$o(\|x-\bar{x}\|^2)$当$\|x-\bar{x}\|^2 \to 0$时,关于$\|x-\bar{x}\|^2$是高阶无穷小量．

十二、Jacobi矩阵

向量值函数，$h(x)=\left(h_1(x),h_2(x),\dots,h_m(x)\right)^{\mathrm{T}}$，每个分量$h_i(x)$为$n$元实函数，如偏导数存在, h在点x的Jacobi矩阵为：

$$h'(x)\text{或}\nabla h(x)^{\mathrm{T}}=\begin{bmatrix}\frac{\partial h_1(x)}{\partial x_1} & \frac{\partial h_1(x)}{\partial x_2} & \cdots & \frac{\partial h_1(x)}{\partial x_n}\\ \frac{\partial h_2(x)}{\partial x_1} & \frac{\partial h_2(x)}{\partial x_2} & \cdots & \frac{\partial h_2(x)}{\partial x_n}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial h_m(x)}{\partial x_1} & \frac{\partial h_m(x)}{\partial x_2} & \cdots & \frac{\partial h_m(x)}{\partial x_n}\end{bmatrix}$$

Q1.

十三、链式法则

设有复合函数$h(x)=f\big(g(x)\big)$,其中向量值函数 $f(g)$和 $g(x)$均可微,$x\in \mathrm{D}^n\subset \mathbf{R}^n$, $g:\mathrm{D}^n\to {\mathrm{D}_1}^m$, f: ${\mathrm{D}_2}^m\to \mathbf{R}^k$,其中${\mathrm{D}_1}^m\subset {\mathrm{D}_2}^m$, $h:\mathrm{D}^n\to \mathbf{R}^k$. 根据复合函数求导数的链式法则必有:

$$h'(x)=f'\big(g(x)\big)g'(x) \qquad x\in \mathrm{D}^n$$

其中f'和g'分别为$k\times m$和$m\times n$矩阵,h'为$k\times n$矩阵

$$\nabla f=(\nabla f_1,\nabla f_2,\dots,\nabla f_k),\nabla g=(\nabla g_1,\nabla g_2,\dots,\nabla g_m)$$

$$\nabla h(x)=\nabla g(x)\nabla f\big(g(x)\big) \quad n\times k$$