第二章 基础知识

目录

§2.1 范数

§2.1 范数

- 向量范数
- 矩阵范数
- 矩阵内积

§2.1 向量范数

定义： 称一个从向量空间$\mathbb{R}^n$到实数域$\mathbb{R}$的**非负**函数$\|\cdot\|$为范数

- **范数的满足条件：**

正定性：

对于所有的$v \in \mathbb{R}^n$，有$\|v\| \geq 0$，且$\|v\| = 0$当且仅当$v = 0$

① 非负性： 向量的长度（范数）总是**大于等于0**，不可能为负数。

② 零向量判别： 只有当 v是零向量（所有分量都为 0）时，它的长度才为 0。反过来，**只要向量不是零向量，它的长度就严格大于0。**

§2.1 向量范数

- **范数的满足条件：**

齐次性：

对于所有的$v \in \mathbb{R}^n$和$\alpha \in \mathbb{R}$，有$\|\alpha v\| = |\alpha|\|v\|$

① 数乘作用： 当向量v被放大或缩小α倍时，它的长度也会相应被放大或缩小$|\alpha|$倍；

- 如果$\alpha > 0$，长度成比例变化；
- 如果$\alpha < 0$，方向翻转，但长度依然取绝对值，所以依然为正值；

$$\|2v\| = 2\|v\|$$

$$\|-3v\| = 3\|v\|$$

§2.1 向量范数

- **范数的满足条件：**

三角不等式：

对于所有的$v,w\in\mathbb{R}^n$，有$\|v+w\|\le\|v\|+\|w\|$

- 两向量相加后的“长度”，不会超过他们各自长度之和；
- 走**“折线路径”**的总长度一定不小于**“直线路径”**的长度。

$$v=(3,4),\ w=(1,2),\ v+w=(4,6)$$

$$\|v+w\|\approx 7.21\le\|v\|+\|w\|\approx 7.24$$

§2.1 向量范数

- 最常用的向量范数为l_p范数($\boldsymbol{p \geq 1}$)

$$\|v\|_p = (|v_1|^p + |v_2|^p + \cdots + |v_n|^p)^{\frac{1}{p}}$$

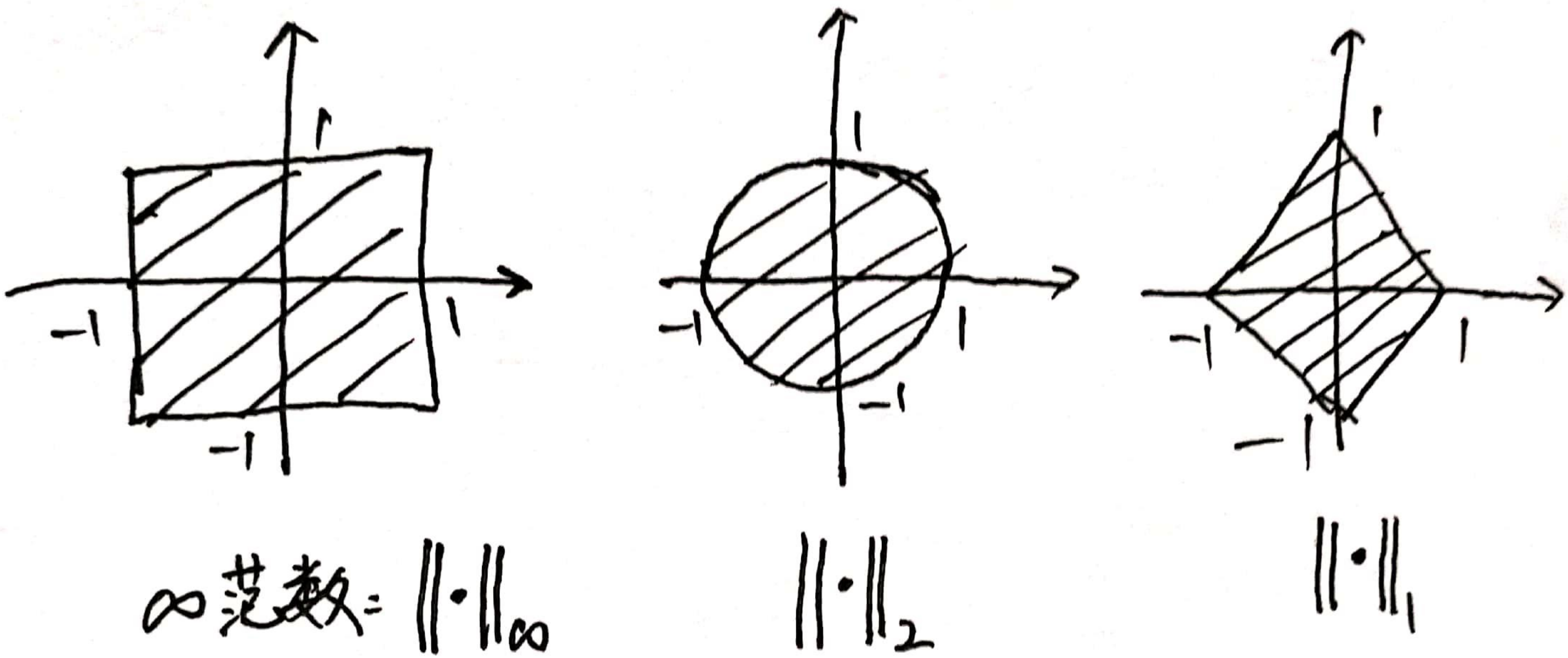
- 举例:

$\boldsymbol{l_1}$范数: $\|v\|_1 = \sum_{i=1}^{n} |v_i|$ ——曼哈顿范数

$\boldsymbol{l_2}$范数: $\|v\|_2 = (\sum_{i=1}^{n} |v_i|^2)^{\frac{1}{2}}$ ——欧几里得范数

$\boldsymbol{l_\infty}$范数: $\|v\|_\infty = \max_i |v_i|$ ——无穷/最大范数

§2.1 向量范数



§2.1 向量范数

- 设向量：

$$v = (3, -4, 1) \in \mathbb{R}^3$$

- 分别计算：

1. $\|v\|_1$
2. $\|v\|_2$
3. $\|v\|_\infty$

习题1

§2.1 向量范数

- 设$\boldsymbol{v} \in \mathbb{R}^{\boldsymbol{n}}$，证明或验证下列范数之间的关系：

① 对任意 $\boldsymbol{v} \in \mathbb{R}^{\boldsymbol{n}}$，有

$$\|v\|_\infty \leq \|v\|_2 \leq \|v\|_1$$

② 进一步证明存在常数$\boldsymbol{C_1}$，$\boldsymbol{C_2}$，使得

$$\|v\|_1 \leq \sqrt{C_1}\|v\|_2, \quad \|v\|_2 \leq \sqrt{C_2}\|v\|_\infty$$

习题2

§2.1 向量范数

- **柯西不等式定义：** 设$\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{\boldsymbol{n}}$，则

$$|a^T b| \leq ||a||_2 ||b||_2$$

等号成立当且仅当 a 与 b 线性相关

§2.1 向量范数

- 考虑优化问题：

$$\min_{x\in\mathbb{R}^2} \ f(x)=c^Tx \quad \text{s.t. } \|x\|_1\le 1$$

其中$c=(2,-3)^T$

① 写出可行域的几何性状；

② 找到最优解$\boldsymbol{x}^*$以及最优值$\boldsymbol{f}(\boldsymbol{x}^*)$

习题3

§2.1 范数

- 向量范数
- **矩阵范数**
- 矩阵内积

§2.1 矩阵范数

定义：

和向量范数类似，矩阵范数是定义在**矩阵空间**上的**非负函数**，并且满足**正定性、齐次性和三角不等式。**

l_1范数： $$\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|$$

每一列元素绝对值之和的最大值，表示矩阵“列最大权重”

l_2范数： $$\|A\|_2 = \max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2}$$

描述矩阵对向量的最大拉伸倍数

§2.1 矩阵范数

定义：

和向量范数类似，矩阵范数是定义在**矩阵空间**上的**非负函数**，并且满足**正定性、齐次性和三角不等式。**

l_∞范数： $$\|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|$$

每一行元素绝对值之和的最大值，表示矩阵"行最大权重"

l_F范数： $$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}$$

直观上就是所有元素平方和的平方根

§2.1 矩阵范数

给定矩阵A:

$$A = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix}$$

计算四种范数 $||A||_1, ||A||_2, ||A||_\infty, ||A||_F$

解答:

$$||A||_1 = \max\{|1| + |3|,\ |-2| + |4|\} = \max\{4, 6\} = 6$$

$$||A||_2 \approx 5.117$$

$$||A||_\infty = \max\{|1| + |-2|,\ |3| + |4|\} = \max\{3, 7\} = 7$$

$$||A||_F = \sqrt{1^2 + (-2)^2 + 3^2 + 4^2} = \sqrt{30} \approx 5.477$$

习题4

§2.1 矩阵范数

给定矩阵A:

$$A = \begin{bmatrix} 2 & -1 & 0 \\ 3 & 4 & -2 \end{bmatrix}$$

① 计算 $\|A\|_F$

② 验证 $\|A\|_F^2 = \mathrm{trace}(A^T A)$

习题5

§2.1 矩阵范数

给定矩阵A:

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$$

① 计算 $\|A\|_1, \|A\|_\infty$

② 给出 $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$ 的表达式，并求出近似值;

③ 解释 $\|A\|_2$ 的几何意义

习题6

§2.1 范数

- 向量范数
- 矩阵范数
- **矩阵内积**

§2.1 矩阵内积

定义：

范数一般用来衡量**矩阵的模的大小**，而内积一般用来表征两个矩阵（或其张成的空间）之间的**夹角**。

- **Frobenius 内积：**

$$< A, B >= Tr(AB^T) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij}$$

其为**两个矩阵逐分量相乘的和**，因而满足内积的定义。此外，当A=B时，该内积等于**F范数的平方。**

§2.1 矩阵内积

给定矩阵$\boldsymbol{A}$，$\boldsymbol{B}$：

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 0 \\ 2 & 1 \end{bmatrix}$$

① **计算矩阵内积** $\langle A, B\rangle = \mathrm{tr}(A^\top B)$

② **验证是否满足** $\langle A, B\rangle = \sum_{i,j} a_{ij} b_{ij}$

习题7

§2.1 矩阵内积

给定矩阵$\boldsymbol{C}$：

$$C = \begin{bmatrix} 1 & -2 & 0 \\ 2 & 1 & -1 \end{bmatrix}$$

① 计算 $\langle C, C \rangle$

② 验证 $\|C\|_F^2 = \langle C, C \rangle$

习题8

§2.2 导数

§2.2 导数

- 梯度与海瑟矩阵
- 矩阵变量函数的导数
- 自动微分

§2.2 导数

定义：

给定函数$f:\mathbb{R}^n \longrightarrow \mathbb{R}$，且$f$在点$x$的一个邻域内有意义，若存在向量$g \in \mathbb{R}^n$满足

$$\lim_{p \to 0} \frac{f(x+p) - f(x) - g^T p}{||p||} = 0$$

- 其中$\|\cdot\|$是任意的向量范数，就称f在点x处可微
- 此时g称为f在点x处的梯度，记作$\nabla f(x)$
- 如果对区域D上的每一个点x都有$\nabla f(x)$存在，则称f在D上可微

§2.2 导数

已知函数：

$$f(x,y) = x^2 + y^2$$

① 验证 $f(x,y)$ 在点 $(1,2)$ 处是否可微；

② 计算其梯度 $\nabla f(x,y)$

习题9

§2.2 导数

已知函数：

$$f(x) = x_1^2 + 3x_1x_2 + 2x_2^2$$

① 验证 $\nabla f(x)$ 存在

② 写出在点$(1,1)$的梯度值

习题10

§2.2 导数

已知函数：

$$f(x,y) = x^2 y$$

① 在点(1,1)，沿方向 $p = (1,1)^T$ 的方向导数是多少？

习题11

§2.2 导数

定义:

若f在点x处的梯度存在，令$p = \varepsilon e_i$，e_i是第i个分量为1的单位向量，可知$\nabla f(x)$的第i个分量即为$\frac{\partial f(x)}{\partial x_i}$.

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

§2.2 导数

定义——海瑟矩阵:

如果函数$f(x)$: $\mathbb{R}^n \to \mathbb{R}$在点$x$处的二阶导数$\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$都存在，则:

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

当$\nabla^2 f(x)$在区域D上的每个点x处都存在时，称f在D上二阶可微。若$\nabla^2 f(x)$在D上还连续，则称f在D上二阶连续可微。

§2.2 导数

已知函数：

$$f(x,y)=x^4+x^2y+y^2$$

① 计算 $\nabla f(x,y)$

② 写出 $\nabla^2 f(x,y)$

③ 在点(1,2)处求Hessian矩阵，并判断它是否正定

习题12

§2.2 导数

已知函数：

$$f(x) = x^T A x, \quad A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad x \in \mathbb{R}^2$$

① 计算 $\nabla f(x)$

② 写出 $\nabla^2 f(x)$

③ 判定A是否正定，从而说明f(x)是否为凸函数

习题13

§2.2 导数

定义（梯度利普希茨连续）：

给定可微函数 f，若存在 $L>0$，对任意的$x,y\in\mathbf{dom}\ f$有：

$$\|\nabla f(x)-\nabla f(y)\|\leq L\|x-y\|$$

则称f是梯度利普希茨连的，相应利普希茨常数为L，有时也记作梯度L-利普希茨连续或L-光滑。

梯度利普希茨连续表明$\nabla f(x)$的变化可以被自变量x的变化所控制，满足该性质的函数具有很多好的性质，一个重要的性质是其具有二次上界。

§2.2 导数

证明：

给定可微函数 $f(x)$ 的定义域 $\mathbf{dom}\ f = \mathbb{R}^n$，且为梯度 L-利普希茨连续的，则函数 $f(x)$ 有二次上界：

$$f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|^2, \quad \forall x, y \in \operatorname{dom} f$$

习题14

§2.2 导数

- 梯度与海瑟矩阵
- **矩阵变量函数的导数**
- 自动微分

§2.2 矩阵变量函数的导数

定义：

多元函数梯度的定义可以推广到变量矩阵的情形，对于以$m \times n$矩阵X为自变量的函数$f(X)$，若存在矩阵$G \in \mathbb{R}^{m \times n}$满足：

$$\lim_{V \to 0} \frac{f(X+V) - f(X) - \langle G, V \rangle}{\|V\|} = 0$$

类似于向量情形，矩阵变量函数$f(X)$的梯度可以用其偏导数表示为：

$$\nabla f(X) = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}$$

§2.2 导数

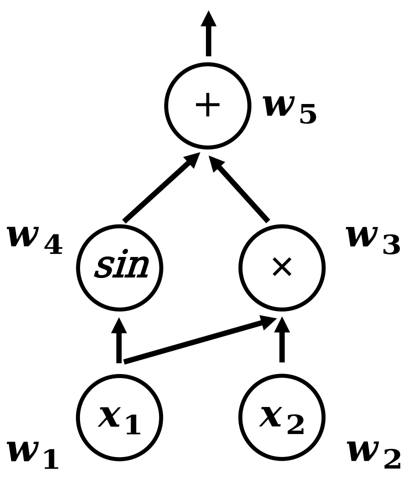
- 梯度与海瑟矩阵
- 矩阵变量函数的导数
- 自动微分

§2.2 自动微分

定义：

自动微分是使用计算机计算导数的算法，对于一个由很多简单函数复合而成的函数，**根据复合函数的链式法则**，可以通过每个简单函数的导数的乘积来计算对于各层变量的导数。

$$f(x_1, x_2) = x_1 x_2 + \sin x_1$$



$$\frac{\partial f}{\partial w_5} = 1,$$

$$\frac{\partial f}{\partial w_4} = \frac{\partial f}{\partial w_5}\frac{\partial w_5}{\partial w_4} = 1,$$

$$\frac{\partial f}{\partial w_3} = \frac{\partial f}{\partial w_5}\frac{\partial w_5}{\partial w_3} = 1,$$

$$\frac{\partial f}{\partial w_2} = \frac{\partial f}{\partial w_3}\frac{\partial w_3}{\partial w_2} = w_1 = x_1,$$

$$\frac{\partial f}{\partial w_1} = \frac{\partial f}{\partial w_3}\frac{\partial w_3}{\partial w_1} + \frac{\partial f}{\partial w_4}\frac{\partial w_4}{\partial w_1} = w_2 + \cos w_1 = x_2 + \cos x_1$$

§2.2 自动微分

已知函数：

$$f(x_1, x_2) = x_1^2 x_2 + \cos(x_1)$$

① 请写出计算图中各个中间变量的定义；

② 利用链式法则，逐步计算出 $\frac{\partial f}{\partial x_1}$, $\frac{\partial f}{\partial x_2}$

习题15

§2.2 自动微分

已知函数:

$$f(x_1, x_2, x_3) = (x_1 + x_2)x_3 + e^{x_1}$$

① 构造其计算图;

② 给出前向传播计算各节点数值的步骤;

③ 给出反向传播的过程，计算在该点处的梯度向量 $\nabla f(x_1, x_2, x_3)$

习题16

§2.3 凸集

§2.3 凸集

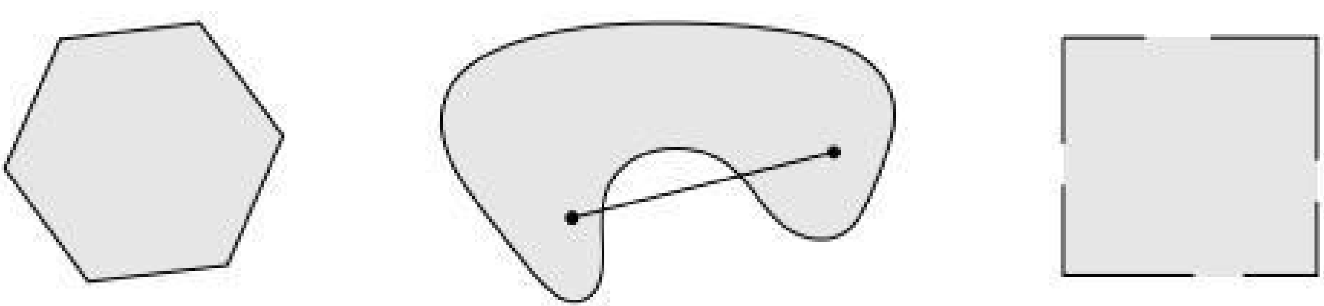
- 凸集相关定义
- 重要的凸集
- 保凸的运算
- 分离超平面定理

§2.3 凸集相关定义

定义：

如果连接集合C中任意两点的线段都在C内，则称C为凸集，即

$$x_1, x_2 \in C \Rightarrow \theta x_1 + (1-\theta)x_2 \in C, 0 \le \theta \le 1$$



§2.3 凸集

- 凸集相关定义
- 重要的凸集
- 保凸的运算
- 分离超平面定理

§2.3 重要的凸集

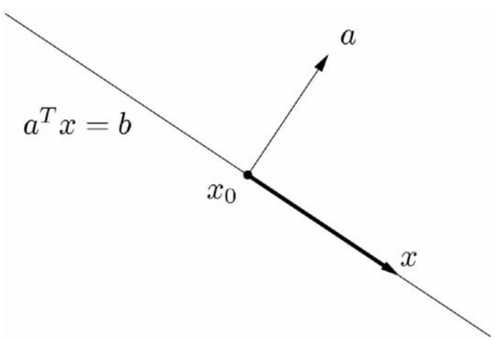
凸集1：超平面和半空间

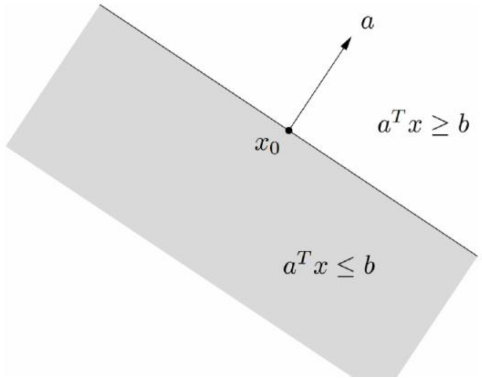
- **超平面**：由非零向量a和标量b定义：

$$H = \{x \in \mathbb{R}^n \mid a^T x = b\}, \quad a \neq 0$$

- **半空间**：由超平面划分的集合：

$$S = \{x \in \mathbb{R}^n \mid a^T x \leq b\} \text{ 或 } S = \{x \in \mathbb{R}^n \mid a^T x \geq b\}$$





§2.3 重要的凸集

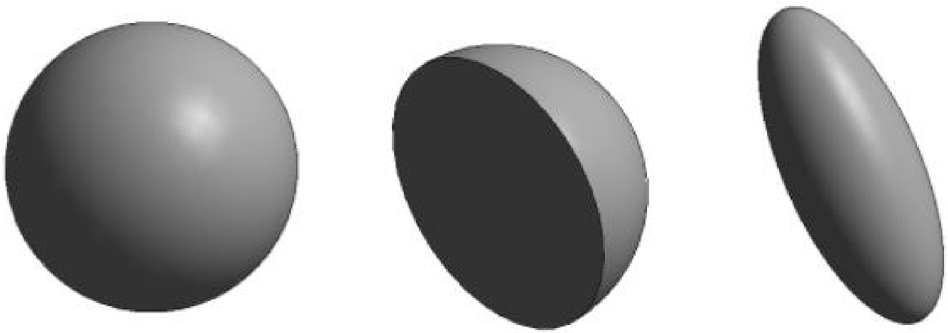
凸集2：球、椭球

- **球**：欧几里得球（以x_c为中心，半径r）：

$$B(x_c, r) = \{x \in \mathbb{R}^n \mid \|x - x_c\|_2 \leq r\}$$

- **椭球**：更一般的形式，涉及正定矩阵P：

$$E(x_c, P) = \{x \in \mathbb{R}^n \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$$

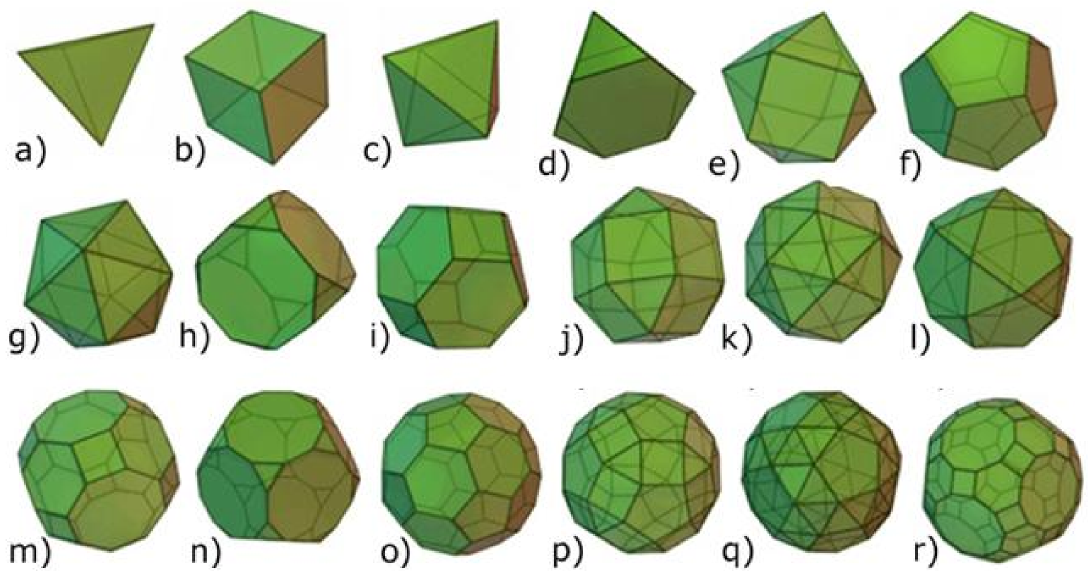


§2.3 重要的凸集

凸集3： 多面体

- 由有限个线性不等式和等式决定的点的集合：

$$P = \{x \in \mathbb{R}^n \mid Ax \leq b,\ Cx = d\}, \quad A \in \mathbb{R}^{m\times n},\ C \in \mathbb{R}^{p\times n}$$



§2.3 凸集

- 凸集相关定义
- 重要的凸集
- **保凸的运算**
- 分离超平面定理

§2.3 保凸的运算

为什么要了解保凸运算：

一般而言，我们证明一个集合（设为C）为凸集有两种方式

第一种是利用定义；

$$x_1, x_2 \in C, 0 \le \theta \le 1 \Rightarrow \theta x_1 + (1-\theta)x_2 \in C$$

第二种方法是说明集合C可由简单的凸集（超平面、半空间、范数球等）通过保凸运算后得到；

定理2.3：任意多个凸集的交为凸集，即若$C_i, i \in \mathcal{I}$是凸集，则

$$\bigcap_{i \in \mathcal{I}} C_i$$

§2.3 保凸的运算

定理： 设 $\boldsymbol{f}$: $\mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}^{\boldsymbol{m}}$是仿射变换

$$f(x) = Ax + b, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

则

(1) 凸集在f下的像是凸集:

$$S \subseteq \mathbb{R}^{\boldsymbol{n}} \text{ 是凸集} \Rightarrow f(S) \overset{\text{def}}{=} \{f(x) | x \in S\} \text{是凸集}$$

(2) 凸集在f下的原像是凸集:

$$C \subseteq \mathbb{R}^{\boldsymbol{m}} \text{ 是凸集} \Rightarrow f^{-1}(C) \overset{\text{def}}{=} \{x \in \mathbb{R}^{\boldsymbol{n}} | f(x) \in C\} \text{是凸集}$$

§2.3 凸集

- 凸集相关定义
- 重要的凸集
- 保凸的运算
- **分离超平面定理**

§2.3 分离超平面定理

定理：

如果C和D是不相交的两个凸集，则存在非零向量a和常数b，使得

$$a^Tx \leq b, \forall x \in C \text{ 且 } a^Tx \geq b, \forall x \in D$$

即超平面 $\{x|a^Tx = b\}$ 分离了C和D

§2.4 凸函数

§2.4 凸函数

- 凸函数的定义
- 凸函数判定定理
- 保凸的运算

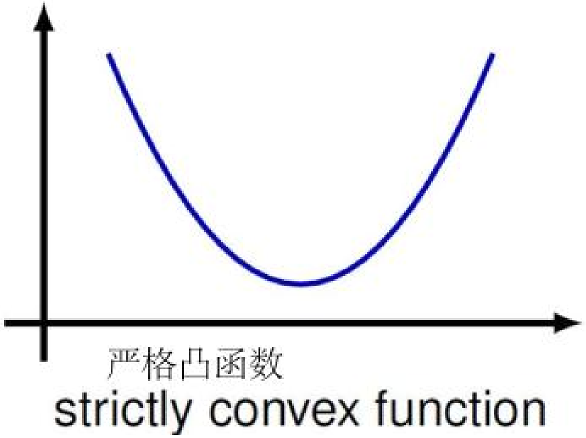
§2.4 凸函数的定义

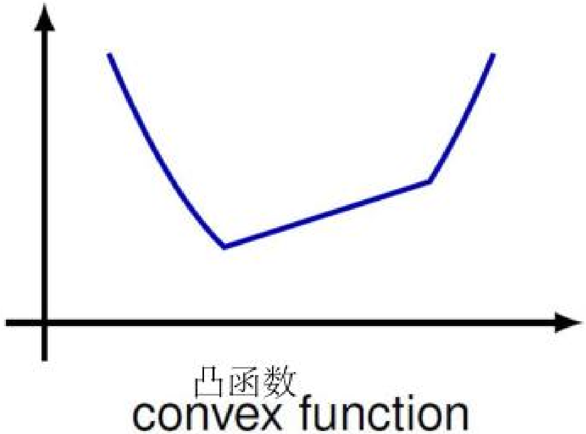
定义：

设函数f为**适当函数**，如果dom f是凸集，且

$$f(\theta x+(1-\theta)y) \leq \theta f(x)+(1-\theta)f(y)$$

对所有 $x, y \in f, 0 \leq \theta \leq 1$ 都成立，则称f是凸函数





§2.4 凸函数

- 凸函数的定义
- 凸函数判定定理
- 保凸的运算

§2.4 凸函数判定定理

定理：

$f(x)$是凸函数当且仅当对任意 $x \in \operatorname{dom} f, v \in \mathbb{R}^n, g : \mathbb{R} \to \mathbb{R}^n$,

$$g(t) = f(x + tv), \quad \operatorname{dom} g = \{t \mid x + tv \in \operatorname{dom} f\}$$

是凸函数。

§2.4 凸函数

- 凸函数的定义
- 凸函数判定定理
- 保凸的运算

§2.4 保凸的运算

定理：

(1) 若 f 是凸函数，则 αf 是凸函数，其中 $\alpha \geq 0$.

(2) 若 f_1, f_2 是凸函数，则 $f_1 + f_2$ 是凸函数.

(3) 若 f 是凸函数，则 $f(Ax + b)$ 是凸函数.

(4) 若$f_1, f_2, \ldots, f_m$均是凸函数，则

$$f(x) = max\{f_1, f_2, \ldots, f_m\} \text{ 是凸函数.}$$