矩阵代数与应用

丁广太

上海大学计算机工程与科学学院

gtding@shu.edu.cn, 2023年3月

开课宗旨

研究对象

信息世界的数据表示形式

物理量之间的关系

数学模型

线性系统与非线性系统？**函数、方程、关系**

算法

数值计算、人工智能→机器学习

实现

程序、电路、以其它方式实现的数据处理规则

典型应用问题

- 解线性方程组
- 信号滤波
- 控制系统设计
- 傅里叶变换
- 最优化问题求解

授课内容及课时计划

- 代数与矩阵基础……………………………………………10学时

  （代数、矩阵、初等变换、性能指标、内积与范数）

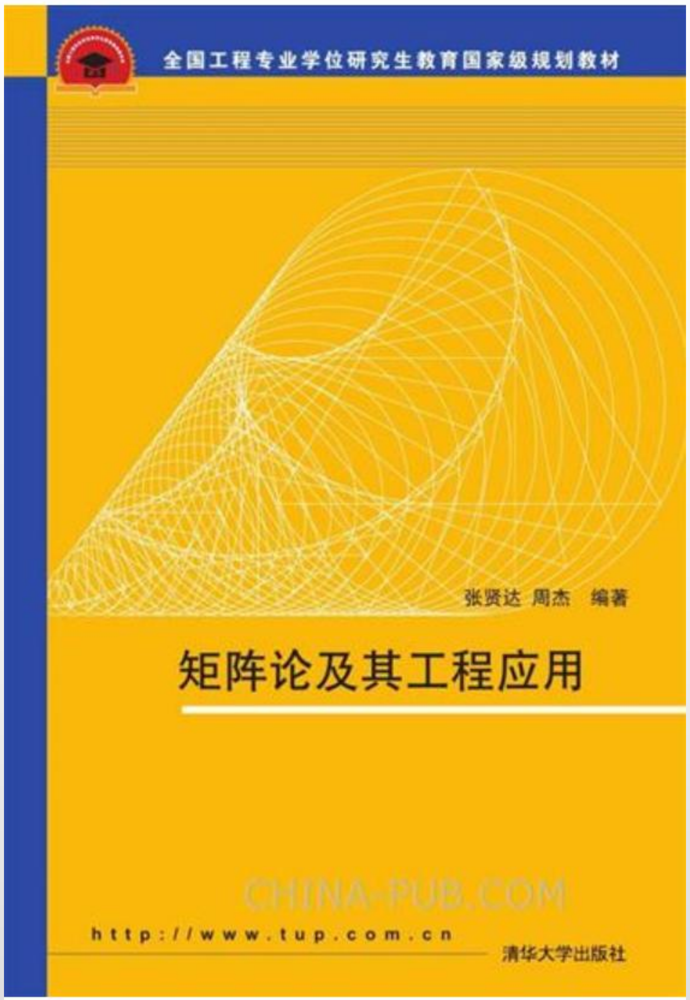
- 特殊矩阵……………………………………………………5学时

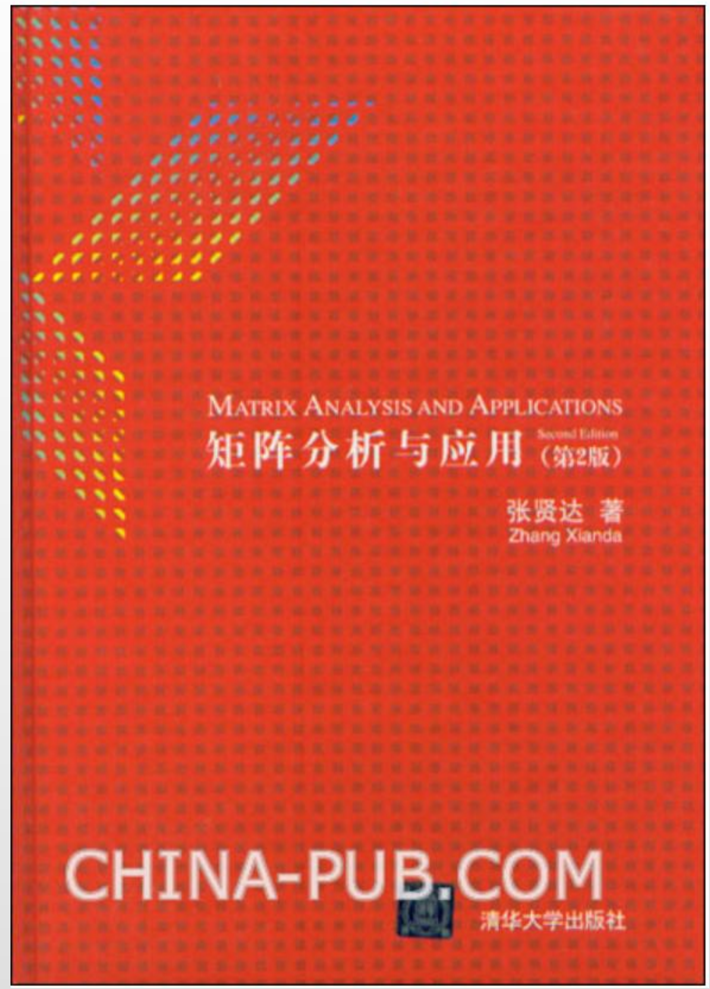
  （置换矩阵、互换矩阵、选择矩阵、正交矩阵、三角阵等）

- 特征分析与奇异值分析…………………………………5学时
- 广义逆与矩阵方程求解…………………………………15学时
- 矩阵微分与梯度分析……………………………………15学时
- 典型应用……………………………………………………10学时

教材与参考书籍

课程目标：矩阵理论、思维习惯、科学素养。反对死记硬背

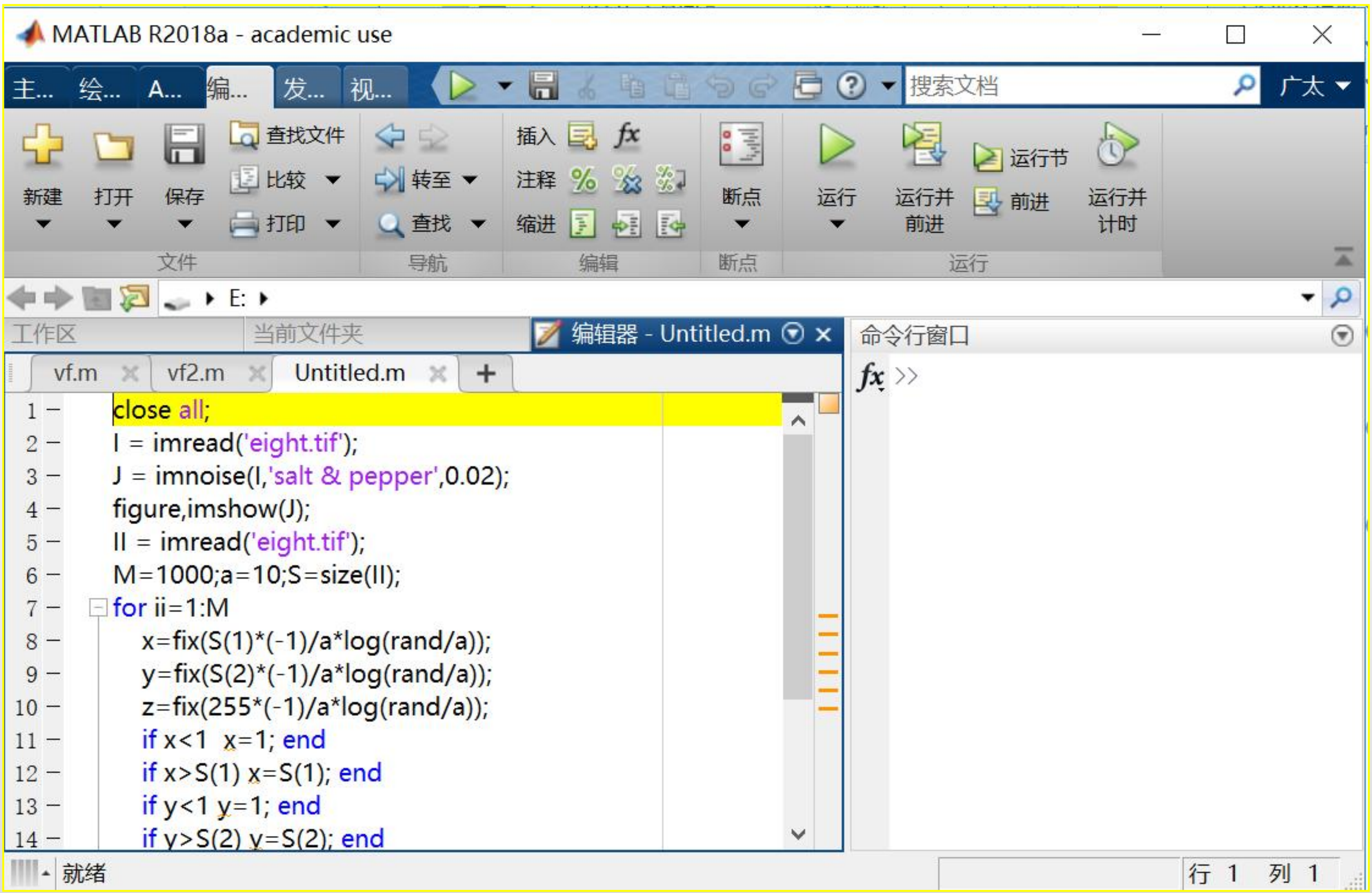




考核方式：待定

课堂形式

星期二(3-4):以讨论为主，讲解教材内容
星期四(3-4):讲解教材内容
星期五(3-4):上机实验，使用Matlab软件对教学内容进行知识验证



感谢清华大学自动化系**韩芳明、张贤达**两位老师

为本课程提供了课件资料！

(本课件相关章节版权属于以上两位老师)

清華大學　　研究生课程教学课件

矩阵分析与应用

韩 芳 明　　张 贤 达

清华大学 自动化系 · 研究生院

办公室：李兆基科技大楼 B220-2

E-mail: fmhan@tsinghua.edu.cn

课堂纪律

不看手机、手机静音

关闭电脑

不得拍照

积极参与讨论

联系方式

- 东区计算机大楼1023
- gtding@shu.edu.cn （工作信箱）
- **gtding@t.shu.edu.cn(用于学生提交作业的信箱)**

第一章 代数与矩阵基础

1.1 代数与矩阵

1.2 矩阵的初等变换

1.3 矩阵的性能指标

1.4 内积与范数

1.5 矩阵和向量的应用案例

数据+运算+指标

目录

CONTENTS

1.1 代数与矩阵

代数是研究数、数量、关系、结构及其之间的联系(如代数方程组等)的性质极其相关问题的通用解法的数学分支.

代数代数—以符号代表数及数量等.

只关心各种关系及其性质，而对于“数本身是什么”这样的问题并不关心.
常见的代数结构类型有：群、环、域、模、线性空间等.

——百度百科

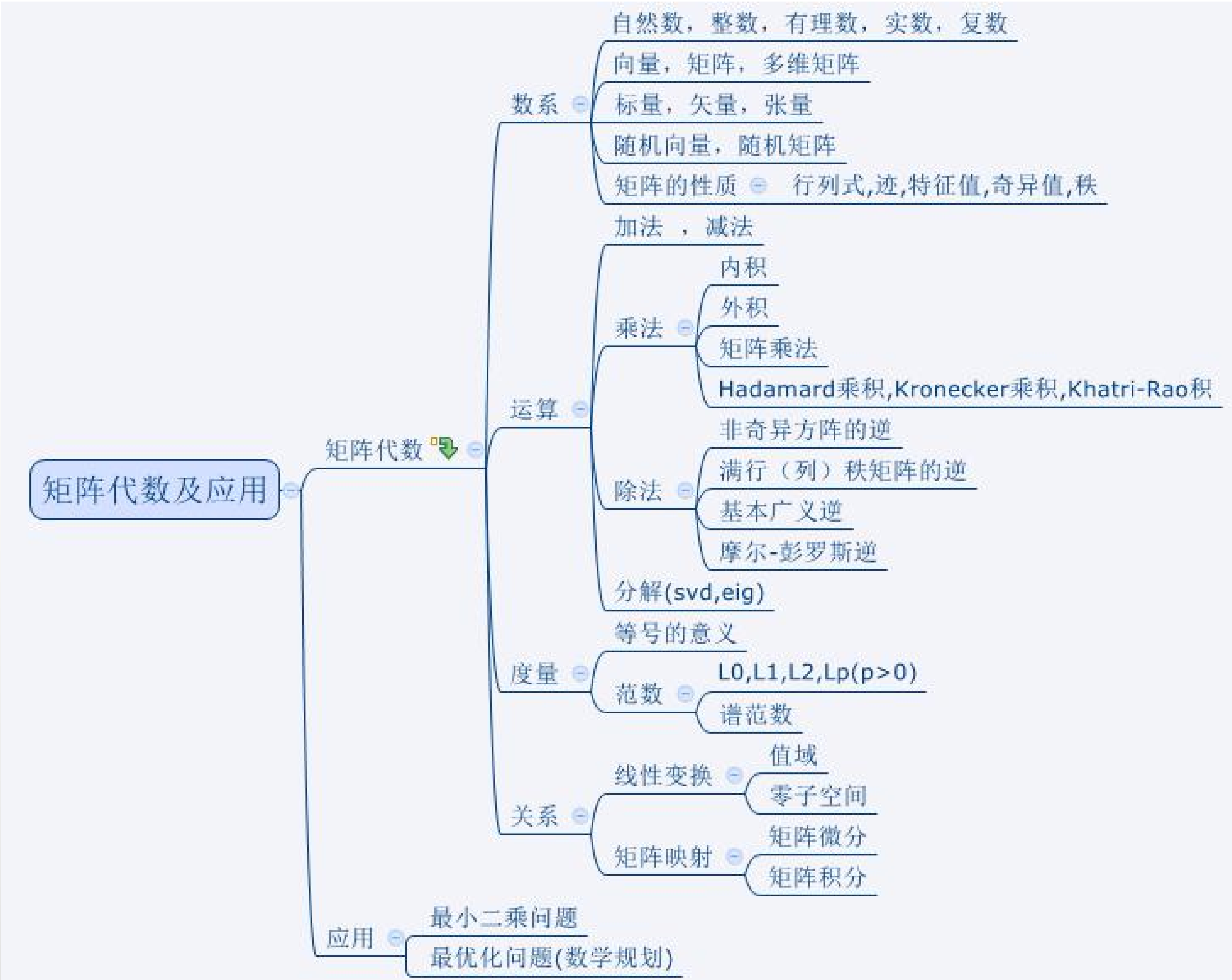
量 physical quantity
- 向量
  - 物理向量
  - 几何向量
  - 代数向量
    - 常数向量
    - 函数向量
    - 随机向量
- 矩阵、张量、其他

数：自然数，整数，有理数，实数，复数

运算：加，减，乘，除

运算规则：交换律，结合律，分配律

等号规则：等号两边施加相同运算，等式仍然成立


矩阵代数及应用
矩阵代数
数系
自然数，整数，有理数，实数，复数
向量，矩阵，多维矩阵
标量，矢量，张量
随机向量，随机矩阵
矩阵的性质
行列式,迹,特征值,奇异值,秩
运算
加法 ，减法
乘法
内积
外积
矩阵乘法
Hadamard乘积,Kronecker乘积,Khatri-Rao积
除法
非奇异方阵的逆
满行（列）秩矩阵的逆
基本广义逆
摩尔-彭罗斯逆
分解(svd,eig)
度量
等号的意义
范数
L0,L1,L2,Lp(p>0)
谱范数
关系
线性变换
值域
零子空间
矩阵映射
矩阵微分
矩阵积分
应用
最小二乘问题
最优化问题(数学规划)

矩阵代数纲要

向量与矩阵的符号表示及基本运算

符号表示

R：实数集合； C：复数集合

$R^{m\times n}\left(C^{m\times n}\right)$：所有m×n实（复）数矩阵的全体

向量：小写字母加黑字母，如**a**，**x**，**y**，**z** 等；

矩阵：大写字母加黑字母，如**A**，**B**，**C** 等(**一般情况下，黑体表示变换，正体表示该变换对应的矩阵,即变换的表示矩阵)**.

$$\mathbf{A}\in R^{m\times n}\Leftrightarrow \mathbf{A}=\left(a_{ij}\right)=\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, a_{ij}\in R$$

若m=n，称矩阵A为正方矩阵（square matrix）；

若m<n，则称A为宽矩阵（broad matrix）或扁矩阵；

若m>n，便称A是高矩阵（high matrix）；

若n=1，则称矩阵A退化为列向量 $a\in C^{m\times 1}$；

若m=1，则矩阵A退化为行向量 $a\in C^{1\times n}$.

数学符号的使用原则：

（1）意义明确、定义清晰

（2）不产生二义性

（3）简单明了、遵从习惯

目的：

便于思想的表达和科学交流

向量/矩阵的基本运算

逆运算?

（1）加减运算： $A \pm B = (a_{ij} \pm b_{ij})$

（2）数乘运算： $kA = (k \cdot a_{ij}), k \in C$

（3）矩阵乘法： $A_{m\times n}, B_{n\times p} \Rightarrow AB = \left(\sum_{k=1}^{n} a_{ik} b_{kp}\right)_{m\times p}$

（4）转置 A^T 与复共轭转置(Hermitian转置)： A^H

$$A^T = (a_{ji}) = \begin{bmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{mn} \end{bmatrix}_{n\times m} \qquad A^H = (a_{ji}^*) = \begin{bmatrix} a_{11}^* & \cdots & a_{m1}^* \\ \vdots & & \vdots \\ a_{1n}^* & \cdots & a_{mn}^* \end{bmatrix}_{n\times m}$$

$$(AB)^T = B^T A^T, \qquad (AB)^H = B^H A^H$$

学习目的：认知世界状态、把握世界运行规律、学以致用…

思维方式：演绎、归纳

理科思维：注重理论和演绎推理（发现、创立…)

工科思维：以实验评判是非（发明、创立…)

文科思维：用心去认知、理解和表达世界

医科思维：注重临床经验(不明…)

艺术科思维：注重想象、以技法表达对世界的认知(不明…)

- 基于理论的逻辑推理
- 基于理论的计算机模拟、仿真
- 物理实验
- 数据实验

何以正确?

何以立功立言?

发明、发现、创立…

问题1：用于进行数据实验的数据从哪里来？

（1）基于规则生成

（2）通过观测仪器对研究对象的状态进行数据采集

（3）计算过程中的中间数据

随机变量

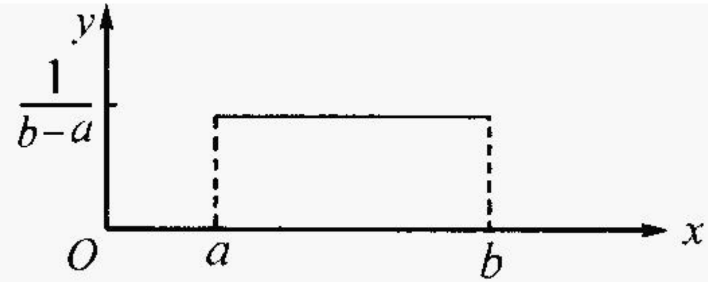
- 随机变量（random variable）表示随机试验各种结果的实值单值函数。随机事件不论与数量是否直接有关，都可以数量化，即都能用数量化的方式表达。
- 随机事件数量化的好处是可以用数学分析的方法来研究随机现象。例如某一时间内公共汽车站等车乘客人数，电话交换台在一定时间内收到的呼叫次数，灯泡的寿命等等，都是随机变量的实例。

(百度百科-随机变量)

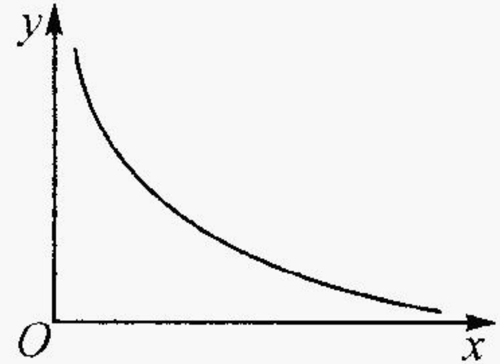
概率空间的公理化描述

- 概率空间由三要素(Ω,$\mathscr{F}$,P)构成。Ω:全体样本;$\mathscr{F}$:全体事件，事件是样本的集合；P：概率测度。
- 概率空间(Ω,$\mathscr{F}$,P)上的随机变量X是定义于Ω上的实值可测函数，即对任意$\omega\in\Omega$，$X(\omega)$为实数，且对任意实数x，使不等式$X(\omega)\leqslant x$成立的一切ω组成的Ω的子集$\{\omega:X(\omega)\leqslant x\}$是事件，也即是$\mathscr{F}$中的元素。
- 事件$\{\omega:X(\omega)\leqslant x\}$常简记作$\{X\leqslant x\}$，并称函数$F(x)=P(X\leqslant x)$，$-\infty<x<\infty$，为X的分布函数。
- 概率密度函数：随机变量X的分布函数的导数。

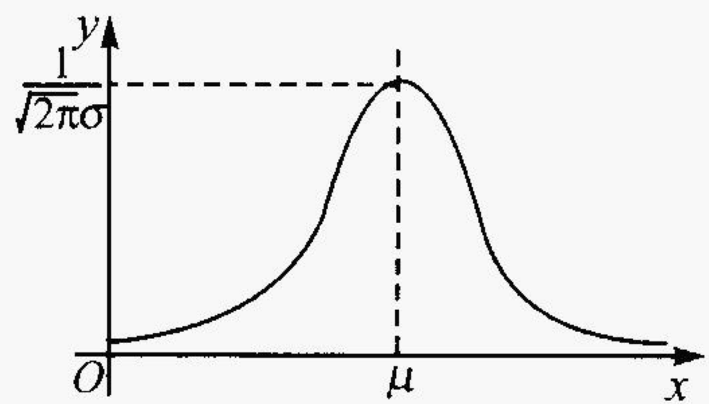
均匀分布：$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x\leqslant b,\\ 0 & 其他;\end{cases}$



指数分布：$f(x)=\begin{cases}a\mathrm{e}^{-ax} & x>0,\\ 0 & 其他;\end{cases}$



正态分布：$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$.



（图片来源：百度百科–随机变量）

随机向量 $\mathbf{x}(\xi)=[x_1(\xi),\cdots,x_m(\xi)]^T$ 的统计描述

◆ 均值向量

$$\mu_{\mathbf{x}} = E\{\mathbf{x}(\xi)\} = \begin{bmatrix} E\{x_1(\xi)\} \\ \vdots \\ E\{x_m(\xi)\} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_m \end{bmatrix}$$

```
A=rand(1000,10);
S=size(A);
tic
B=sum(A,1);
toc %计算时间
B=B/S(1);
disp(S);disp(B);
```

式中，数学期望定义为

$$E\{x_i(\xi)\} \stackrel{def}{=} \int_{-\infty}^{\infty} t f_i(t)dt$$

```
A=rand(1000,10);
S=size(A);
B=zeros(1,S(2));
tic
for i=1:S(1)
    B(1,:)=B(1,:)+A(i,:);
end
toc %计算时间
B=B/S(1);
disp(S);disp(B);
```

```
时间已过 0.005261 秒。
        1000          10
    0.4984    0.4970    0.4964    0.5053    0.5065
    0.5078    0.4816    0.4903    0.5040    0.4863

时间已过 0.000142 秒。
        1000          10
    0.4984    0.4970    0.4964    0.5053    0.5065
    0.5078    0.4816    0.4903    0.5040    0.4863
```

随机向量 $\mathbf{x}(\xi)=[x_1(\xi),\cdots,x_m(\xi)]^T$ 的统计描述

- 自相关矩阵

$$R_{\mathbf{x}} \overset{def}{=} E\{\mathbf{x}(\xi)\mathbf{x}(\xi)^H\} = \begin{bmatrix} r_{11} & r_{12} & \cdots r_{1m} \\ \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & \cdots r_{mm} \end{bmatrix}$$

式中，$r_{ii} = E\{|x_i(\xi)|^2\}$ 表示$x_i(\xi)$ 的自相关函数，而 $r_{ij} = E\{x_i(\xi)x_j^*(\xi)\}$ 表示$x_j(\xi)$ 和 $x_i(\xi)$之间的互相关函数.

```
rng(0);%设置随机数种子
A=rand(5,1000); S=size(A);
B=zeros(S(1),S(1));
for i=1:S(2)
   B=B+A(:,i)*A(:,i)';
end
B=B/S(2);
disp(B); disp(A*A'/S(2)); %???
```

0.3409	0.2596	0.2556	0.2476	0.2529
0.2596	0.3410	0.2600	0.2558	0.2566
0.2556	0.2600	0.3449	0.2525	0.2557
0.2476	0.2558	0.2525	0.3266	0.2492
0.2529	0.2566	0.2557	0.2492	0.3334

0.3409	0.2596	0.2556	0.2476	0.2529
0.2596	0.3410	0.2600	0.2558	0.2566
0.2556	0.2600	0.3449	0.2525	0.2557
0.2476	0.2558	0.2525	0.3266	0.2492
0.2529	0.2566	0.2557	0.2492	0.3334

伪随机数算法

均匀分布：线性同余法

$$\begin{cases} x_n = (ax_{n-1} + b) \bmod m \\ \xi_n = x_n / m \end{cases}$$

a:乘数,b:增量,m:模数,x_0:**种子数**

```
a=12;b=20;m=2^27+1;
x=123456;
for i=1:10
    x=rem(a*x+b,m);
    disp(x/m);
end
```

0.0110, 0.1325, 0.5895, 0.0736,
0.8837, 0.6049, 0.2586, 0.1031,
0.2367, 0.8399

https://www.cnblogs.com/forget406/p/5294143.html

随机数的产生和检验(略)

一般分布

若$P(\xi \le x) = F(x)$,则

$$P(F(\xi) \le x) = P(\xi \le F^{-1}(x)) = x$$

若$\xi \sim F$,则$F(\xi) \sim$均匀分布；

若$\xi \sim$均匀分布,则$F^{-1}(\xi) \sim F$.

例.指数分布

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x}, x \ge 0 \\ 0, \qquad x < 0 \end{cases}$$

$$x = -\log(1 - F(x;\lambda))/\lambda, F(x;\lambda) \ge 0$$

```
for j=1:10
    y=-log(1-rand())/10;
    disp(y);
end
```

0.1066，0.0036，0.1891，0.2718，0.1135
0.1418，0.1359，0.0498，0.1066，0.0188

随机向量 $\mathbf{x}(\xi)=\left[x_1(\xi),\cdots,x_m(\xi)\right]^T$ 的统计描述

◆ 自协方差矩阵

$$C_{\mathbf{x}} \overset{def}{=} E\{[\mathbf{x}(\xi)-\mu_{\mathbf{x}}]\cdot[\mathbf{x}(\xi)-\mu_{\mathbf{x}}]^H\}$$

$$=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ \vdots & \vdots & & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{bmatrix}$$

其中，$c_{ii}=E\{\left|x_i(\xi)-\mu_i\right|^2\}$ 表示 $x_i(\xi)$ 的方差 σ_i^2 ，而

$$c_{ij}=E\{[x_i(\xi)-\mu_i]\cdot[x_j(\xi)-\mu_j]^*\}=E\{x_i(\xi)x_j^*(\xi)\}-\mu_i\mu_j^*=c_{ji}^*$$

表示随机变量 $x_i(\xi)$ 和 $x_j(\xi)$ 之间的协方差.

随机向量 $\mathbf{x}(\xi)=[x_1(\xi),\cdots,x_m(\xi)]^T$ 的统计描述

- 容易看出，自相关矩阵 R_x 和自协方差矩阵 C_x 均为复共轭对称（或Hermitian）矩阵，即

$$R_{\mathbf{x}}^H = R_{\mathbf{x}},\quad C_{\mathbf{x}}^H = C_{\mathbf{x}}$$

- 自相关矩阵和自协方差矩阵之间的关系：

$$C_{\mathbf{x}} = R_{\mathbf{x}} - \mu_{\mathbf{x}}\mu_{\mathbf{x}}^H$$

```
rng(0);%设置随机数种子
A=randn(5,1000);
S=size(A);
B=zeros(S(1),S(1));
C=zeros(S(1),1);
for i=1:S(2)
    B=B+A(:,i)*A(:,i)';
    C=C+A(:,i);
end
B=B/S(2);
C=C/S(2);
disp(B);disp(B-C*C');
```

0.9747	-0.0190	-0.0092	0.0237	-0.0274
-0.0190	**1.0256**	-0.0171	0.0302	-0.0062
-0.0092	-0.0171	**0.9636**	0.0295	-0.0139
0.0237	0.0302	0.0295	**0.9809**	0.0706
-0.0274	-0.0062	-0.0139	0.0706	**1.0013**

```
rng(0);%设置随机数种子
A=rand(5,1000);
S=size(A);
B=zeros(S(1),S(1));
C=zeros(S(1),1);
for i=1:S(2)
    B=B+A(:,i)*A(:,i)';
    C=C+A(:,i);
end
B=B/S(2);
C=C/S(2);
disp(B);disp(B-C*C');
```

0.0854	-0.0003	-0.0020	-0.0021	-0.0002
-0.0003	0.0768	-0.0020	0.0019	-0.0006
-0.0020	-0.0020	0.0852	0.0008	0.0006
-0.0021	0.0019	0.0008	0.0827	0.0020
-0.0002	-0.0006	0.0006	0.0020	0.0829

随机向量 $\mathbf{x}(\xi)=[x_1(\xi),\cdots,x_m(\xi)]^T$ 的统计描述

- 自相关矩阵和自协方差矩阵的推广

  - **互相关矩阵**

$$R_{\mathbf{xy}} \overset{def}{=} E\{\mathbf{x}(\xi)\mathbf{y}^H(\xi)\} = \begin{bmatrix} r_{x_1,y_1} & r_{x_1,y_2} & \cdots & r_{x_1,y_m} \\ r_{x_2,y_1} & r_{x_2,y_2} & \cdots & r_{x_2,y_m} \\ \vdots & \vdots & & \vdots \\ r_{x_m,y_1} & r_{x_m,y_2} & \cdots & r_{x_m,y_m} \end{bmatrix}$$

  - **互协方差矩阵**

$$C_{\mathbf{xy}} \overset{def}{=} E\{[\mathbf{x}(\xi)-\mu_{\mathbf{x}}]\cdot[\mathbf{y}(\xi)-\mu_{\mathbf{y}}]^H\}$$

$$= R_{\mathbf{xy}} - \mu_{\mathbf{x}}\mu_{\mathbf{y}}^H = \begin{bmatrix} c_{x_1,y_1} & c_{x_1,y_2} & \cdots & c_{x_1,y_m} \\ c_{x_2,y_1} & c_{x_2,y_2} & \cdots & c_{x_2,y_m} \\ \vdots & \vdots & & \vdots \\ c_{x_m,y_1} & c_{x_m,y_2} & \cdots & c_{x_m,y_m} \end{bmatrix}$$

随机向量 $\mathbf{x}(\xi)=\left[x_1(\xi),\cdots,x_m(\xi)\right]^T$ 的统计描述

两个随机变量（向量）之间的关系

随机变量 $x(\xi)$ 和 $y(\xi)$ 之间的相关系数定义为

$$\rho_{xy} \overset{def}{=} \frac{c_{xy}}{\sqrt{E\{|x(\xi)-\mu_x|^2\}E\{|y(\xi)-\mu_y|^2\}}}=\frac{c_{xy}}{\sigma_x\sigma_y}$$

- 两个随机变量 $x(\xi)$ 和 $y(\xi)$ 统计不相关
- 两个随机向量 $\mathbf{x}(\xi)$ 和 $\mathbf{y}(\xi)$ 统计不相关

- 两个随机变量 $x(\xi)$ 与 $y(\xi)$ 正交
- 两个随机向量 $\mathbf{x}(\xi)$ 与 $\mathbf{y}(\xi)$ 正交

假设X和Y为两个随机变量。

- 正交

  定义R(X,Y)=E[XY]为相关函数。若R=0,则称X和Y正交。

- 不相关

  若E[XY]=E[X]E[Y]，则称X和Y为不相关。

- 独立

  X和Y的联合分布等于它们各自的独立分布的乘积，即

$$p(XY)=p_X(X)p_Y(Y),$$

  则称他们相互独立。

结论：

（1）独立⇒不相关

（2）对于高斯随机变量，独立和不相关等价：独立⇔不相关

请自己举正例、反例，研究正交、不相关和独立等概念。

```
N=10000;
a=rand(1,N);
x=3*sin(2*pi*a);
y=cos(2*pi*a);
Ex=0;
Ey=0;
Exy=0;
for ii=1:N
    Ex=Ex+x(ii);
    Ey=Ey+y(ii);
    Exy=Exy+x(ii)*y(ii);
end
Ex=Ex/N;
Ey=Ey/N;
Exy=Exy/N;
disp(Ex);
disp(Ey);
disp(Exy);
disp(Ex*Ey-Exy);
```

0.0034

8.0238e-04

0.0094

-0.0094

```
N=10000;
a=rand(1,N);
x=3*sin(2*pi*a)+3;
y=cos(2*pi*a)+7;
Ex=0;
Ey=0;
Exy=0;
for ii=1:N
    Ex=Ex+x(ii);
    Ey=Ey+y(ii);
    Exy=Exy+x(ii)*y(ii);
end
Ex=Ex/N;
Ey=Ey/N;
Exy=Exy/N;
disp(Ex);
disp(Ey);
disp(Exy);
disp(Ex*Ey-Exy);
```

2.9864

6.9960

20.9064

-0.0133

问题2：如何基于矩阵理论对某种工程问题进行数学建模？

矩阵方程：$AX = b$

二次型：$f(X) = (X - X_0)^T A(X - X_0)$

1.2 矩阵的初等变换

矩阵的初等变换有三种:

(1)交换矩阵中某两行(列)的位置;

(2)用一个非零常数k乘以矩阵的某一行(列);

(3)将矩阵的某一行(列)乘以常数k后加到另一行(列)上去.

三类初等矩阵都是可逆矩阵，即非奇异阵.

三类初等矩阵行列式的值是: (1): -1 (2): k (3): 1

设

1、单位矩阵第i,j两行互换得到的方阵为P_{ij}

2、单位矩阵第i行乘以常数k得到初等方阵$D_i(k)$

3、将单位矩阵的第i行的k倍加到第j行得到初等方阵$T_{ij}(k)$

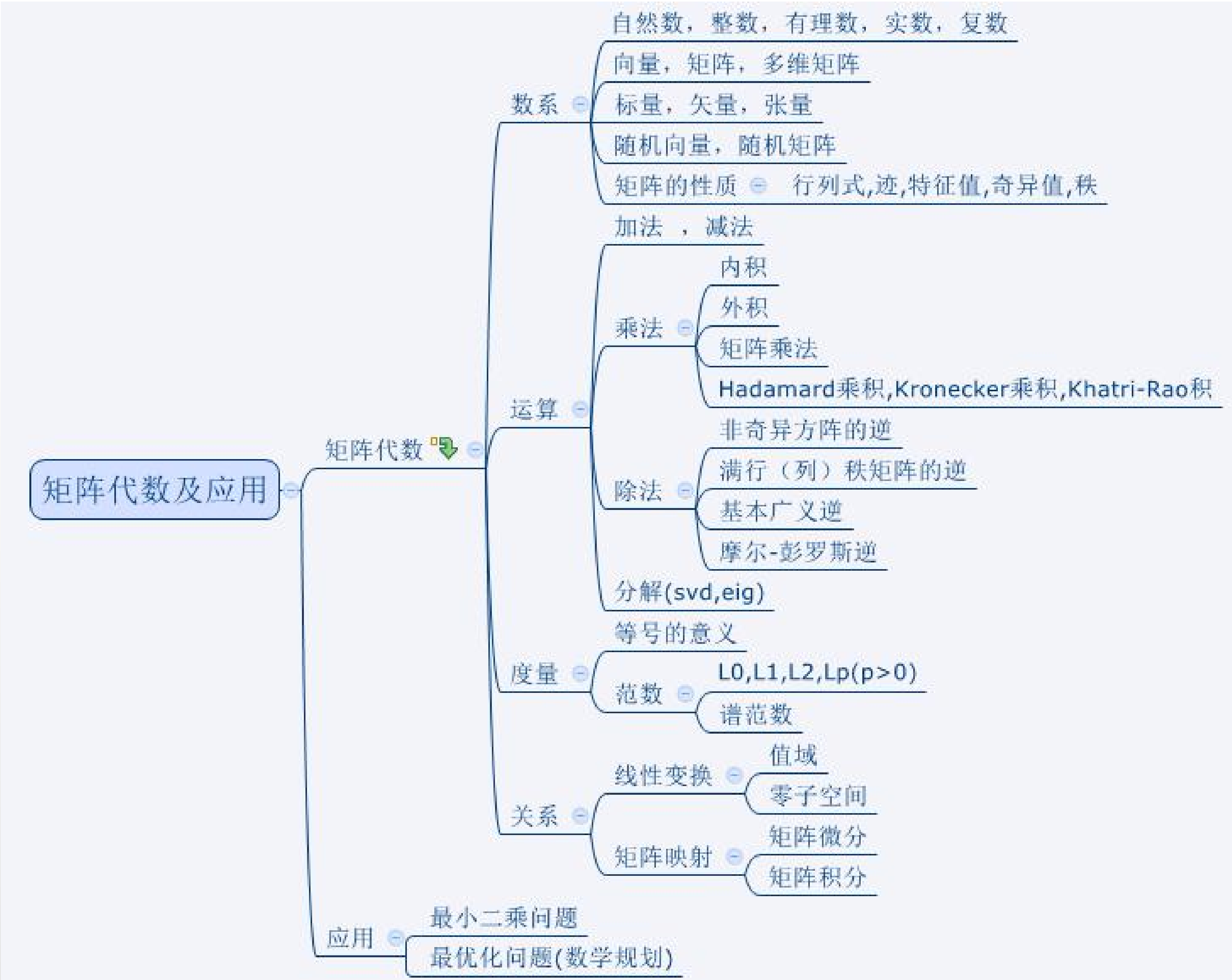
则

将矩阵A的第i,j两行互换所得矩阵为$P_{ij}A$.

操作规则：左行右列

将矩阵A的第i行乘以k得到矩阵$D_i(k)A$.

将矩阵A的第i行的k倍加到第j行得到矩阵$T_{ij}(k)A$.


矩阵代数及应用
矩阵代数
数系
自然数，整数，有理数，实数，复数
向量，矩阵，多维矩阵
标量，矢量，张量
随机向量，随机矩阵
矩阵的性质
行列式,迹,特征值,奇异值,秩
运算
加法 ，减法
乘法
内积
外积
矩阵乘法
Hadamard乘积,Kronecker乘积,Khatri-Rao积
除法
非奇异方阵的逆
满行（列）秩矩阵的逆
基本广义逆
摩尔-彭罗斯逆
分解(svd,eig)
度量
等号的意义
范数
L0,L1,L2,Lp(p>0)
谱范数
关系
线性变换
值域
零子空间
矩阵映射
矩阵微分
矩阵积分
应用
最小二乘问题
最优化问题(数学规划)

矩阵代数纲要

1.3 矩阵及其性能指标

0.线性变换

向量空间

◆ **定义**

令U和V是两个向量空间，L是从U到V的一个函数(映射)，即$L:\mathbf{U}\mapsto\mathbf{V}$.
若对于所有标量a_1,a_2和U中的所有向量$\mathbf{x}_1,\mathbf{x}_2$，有

$$L[a_1\mathbf{x}_1+a_2\mathbf{x}_2]=a_1L[\mathbf{x}_1]+a_2L[\mathbf{x}_2]$$

则称L为线性算子(线性变换、线性映射).

◆ **定理**

设$\mathbf{U}=\{\mathbf{u}_1,\cdots,\mathbf{u}_n\}$和$\mathbf{V}=\{\mathbf{v}_1,\cdots,\mathbf{v}_m\}$分别为$n$维向量空间U和$m$维向量空间V的一组基，则对每一个线性变换$L:\mathbf{U}\mapsto\mathbf{V}$存在一个$m\times n$矩阵A，使得对$\forall r\in U$，有

$$[L(r)]_{\mathbf{V}}=A[r]_{\mathbf{U}}$$

其中$[\cdot]_{\mathrm{U}}$和$[\cdot]_{\mathrm{V}}$分别表示各自相应向量在基U和基V下的**坐标表示**;
而$a_j=[L(u_j)]_{\mathrm{V}}, j=1,2,\cdots,n$，且A称为$L$相对应于U和V的**表示矩阵**.

矢量（向量）：$X \in R^n, Y \in R^m$, n和m都是大于零的整数

$$
\text{线性变换：}\begin{cases} L: R^n \to R^m, \\ L(\alpha_1 X_1 + \alpha_2 X_2) = \alpha_1 L(X_1) + \alpha_2 L(X_2), \\ X_1, X_2 \in R^n, \alpha_1, \alpha_2 \in R(\text{注：复数域类同}) \end{cases}
$$

- 空间上的基向量是选取的，不唯一
- 矢量在不同基下的表示也是不一样的
- 同一矢量在不同基下的代数表示可以互相转换（计算）得到，称为坐标变换
- 对于线性变换，以上说法也正确

设$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n), \widetilde{\mathbf{U}} = (\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \cdots, \tilde{\mathbf{u}}_n)$是选定的两组$R^n$上的基

$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m), \widetilde{\mathbf{V}} = (\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \cdots, \tilde{\mathbf{v}}_m)$是选定的两组$R^m$上的基

则矢量 **X** 和 **Y** 在 R^n 和 R^m 上有如下表示（用黑体表示矢量或基向量，正体表示向量或矩阵）

$$
\mathbf{X} = x_1 \mathbf{u}_1 + x_2 \mathbf{u}_2 + \cdots + x_n \mathbf{u}_n = \sum_{i=1}^{n} x_i \mathbf{u}_i = (\mathbf{u}_1, \mathbf{u}_1, \cdots, \mathbf{u}_1) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \mathrm{U} X
$$

$$
\mathbf{X} = \tilde{x}_1 \tilde{\mathbf{u}}_1 + \tilde{x}_2 \tilde{\mathbf{u}}_2 + \cdots + \tilde{x}_n \tilde{\mathbf{u}}_n = \sum_{i=1}^{n} \tilde{x}_i \tilde{\mathbf{u}}_i = (\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_1, \cdots, \tilde{\mathbf{u}}_1) \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{pmatrix} = \widetilde{\mathrm{U}} \widetilde{X}
$$

$$\mathbf{Y} = y_1\mathbf{v}_1 + y_2\mathbf{v}_2 + \cdots + y_m\mathbf{v}_m = \sum_{i=1}^{m} x_i\mathbf{v}_i = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathbf{V}Y$$

$$\mathbf{Y} = \tilde{y}_1\tilde{\mathbf{v}}_1 + \tilde{y}_2\tilde{\mathbf{v}}_2 + \cdots + \tilde{y}_m\tilde{\mathbf{v}}_m = \sum_{i=1}^{m} \tilde{y}_i\tilde{\mathbf{v}}_i = (\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \cdots, \tilde{\mathbf{v}}_m)\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_m \end{pmatrix} = \tilde{\mathbf{V}}\tilde{Y}$$

$$L(\mathbf{u}_i) = a_{i1}\mathbf{v}_1 + a_{i2}\mathbf{v}_2 + \cdots + a_{im}\mathbf{v}_m = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m)(a_{i1}, a_{i2}, \cdots, a_{im})^T, i = 1,2,\cdots,n$$

简记为

$$L(\mathbf{U}) = \mathbf{V}A$$

也就是

$$L((\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n)) = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m)\begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix}$$

这个 A 称为线性变换在基 **U** 和 **V** 下的表示矩阵。如果自变量空间和应变量空间各选两组基，则一共有四种表示矩阵（A, A_1, A_2, A_3）。

$$L(\mathbf{U}) = \mathbf{V}A$$

$$L(\mathbf{U}) = \tilde{\mathbf{V}}A_1$$

$$L(\tilde{\mathbf{U}}) = \mathbf{V}A_2$$

$$L(\tilde{\mathbf{U}}) = \tilde{\mathbf{V}}A_3$$

它们都表示同一线性变换 L

0.线性变换

◆ 推论

设n维向量空间U和m维向量空间V中各自有如下基变换

$$\text{在U中：}\ (\tilde{\mathbf{u}}_1,\cdots,\tilde{\mathbf{u}}_n)=(\mathbf{u}_1,\cdots,\mathbf{u}_n)P$$

$$\text{在V中：}\ (\tilde{\mathbf{v}}_1,\cdots,\tilde{\mathbf{v}}_m)=(\mathbf{v}_1,\cdots,\mathbf{v}_m)Q$$

再设线性变换 $L:\mathbf{U}\mapsto\mathbf{V}$ 在基 $\mathbf{U}=[\mathbf{u}_1,\cdots,\mathbf{u}_n]$和$\mathbf{V}=[\mathbf{v}_1,\cdots,\mathbf{v}_m]$下的表示矩阵为A；而在基 $\tilde{\mathbf{U}}=[\tilde{\mathbf{u}}_1,\cdots,\tilde{\mathbf{u}}_n]$ 和 $\tilde{\mathbf{V}}=[\tilde{\mathbf{v}}_1,\cdots,\tilde{\mathbf{v}}_m]$下的表示矩阵为$\tilde{A}$，则

$$\tilde{A}=Q^{-1}AP$$

◆ 定义

若A和B均为$m\times n$矩阵，且存在非奇异矩阵P和Q，使得 $B=Q^{-1}AP$，则称B<u>等价于</u>A，并表示为 $B\cong A$.

等价的矩阵表示相同的线性变换！

◆ 推论(的解释和证明)

设n维向量空间U和m维向量空间V中各自有如下基变换

$$\text{在U中}: (\tilde{\mathbf{u}}_1, \cdots, \tilde{\mathbf{u}}_n) = (\mathbf{u}_1, \cdots, \mathbf{u}_n)P$$

$$\text{在V中}: (\tilde{\mathbf{v}}_1, \cdots, \tilde{\mathbf{v}}_m) = (\mathbf{v}_1, \cdots, \mathbf{v}_m)Q$$

再设线性变换$L: \mathbf{U} \mapsto \mathbf{V}$ 在基$\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_n]$和 $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_m]$ 下的表示矩阵为A;
而在基 $\tilde{\mathbf{U}} = [\tilde{\mathbf{u}}_1, \cdots, \tilde{\mathbf{u}}_n]$和 $\tilde{\mathbf{V}} = [\tilde{\mathbf{v}}_1, \cdots, \tilde{\mathbf{v}}_m]$ 下的表示矩阵为 $\tilde{A}$，则

$$\tilde{A} = Q^{-1}AP$$

设$\mathbf{U}, \mathbf{V}, \tilde{\mathbf{U}}, \tilde{\mathbf{V}}$中列向量都是基向量.

$\forall X, L(\mathbf{U}X) = \mathbf{V}AX, L(\tilde{\mathbf{U}}X) = \tilde{\mathbf{V}}\tilde{A}X$ (这是线性变换的“**表示矩阵**”定义的正规写法).

特别地,若基向量$\mathbf{U}, \mathbf{V}$矩阵中列向量取欧氏空间的单位坐标向量，则

$\mathbf{U} = \mathbf{V} = \mathbf{I}, L(X) = AX$ (“**表示矩阵**”定义的简写), $L(\tilde{\mathbf{U}}X) = \tilde{\mathbf{V}}\tilde{A}X$.

已知$\tilde{\mathbf{U}} = \mathbf{U}P, \tilde{\mathbf{V}} = \mathbf{V}Q$,因而

$$L(\mathbf{U}PX) = \mathbf{V}Q\tilde{A}X \Rightarrow L(\mathbf{U}PX) = \mathbf{V}APX = \mathbf{V}Q\tilde{A}X,$$

因为$\mathbf{V}$中向量是基向量，所以$APX = Q\tilde{A}X$.又因为X是任意向量，所以$AP = Q\tilde{A}$.

0.线性变换

◆ **定理**

设 $\mathbf{U}=[\mathbf{u}_1,\cdots,\mathbf{u}_n]$和 $\mathbf{V}=[\mathbf{v}_1,\cdots,\mathbf{v}_n]$的各列分别构成$n$维向量空间W的两组基，而S为从U到V的转移（过渡）矩阵，即**V**=**U**S. 若W上的线性算子(或线性变换)L相应于U和V的表示矩阵分别为A和B，则

$$B=S^{-1}AS$$

1.矩阵的相似

◆ **定义**

若A和B均为$n\times n$矩阵，且存在非奇异矩阵S，使得$B=S^{-1}AS$，则称B相似于（be similar to）A，并表示为 $B\sim A$.

若B~A，则 $A=\left(S^{-1}\right)^{-1}BS^{-1}\sim B$.因此，常简称$A$与$B$相似.

线性变换在不同基下的表示矩阵是相似的！

2.行列式

正方矩阵 $A_{n\times n}$ 的行列式表示为

$$\det(A)=|A|=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

矩阵A去掉第i行和第j列之后得到的剩余矩阵 A_{ij} 的“加权”行列式

$A_{ij}=(-1)^{i+j}\det(A_{ij})$ 称为元素 a_{ij} 的代数余子式(algebraic cofactor). 特别地，当j=i时，$A_{ij}=A_{ii}$ 称为A的代数主子式.

2.行列式

◆ 行列式的计算公式

$$\det(A) = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}$$

$$= \sum_{j=1}^{n} a_{ij}(-1)^{i+j}\det(A_{ij})$$

或者

$$\det(A) = a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj}$$

$$= \sum_{i=1}^{n} a_{ij}(-1)^{i+j}\det(A_{ij})$$

◆ 行列式与线性方程组的关系

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_2 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases} \quad (|A| \neq 0)$$

则

$$x_j = \frac{A_{1j}b_1 + A_{2j}b_2 + \cdots + A_{nj}b_n}{|A|}, j = 1,2,\cdots,n$$

Gramer法则

2.行列式

◆ 行列式的主要性质

（1）如果矩阵的两行（或列）互换位置，则行列式数值保持不变，但符号改变.

（2）将矩阵的任一行（列）乘以任一常数后加到任一其他行（列），则行列式数值保持不变.

（3）对任意的常数c，有 $\det(cA)=c^n\det(A)$.

（4）上/下三角矩阵的行列式等于其对角线元素的乘积.

（5）单位矩阵的行列式等于1，即det(**I**)=1.

（6）转置矩阵与复共轭（Hermitian）转置矩阵的行列式

$$\det(A)=\det(A^T), \det(A^H)=[\det(A^T)]^*$$

（7）行列式等于特征值之积，即 $|A|=\prod_{i=1}^{n}\lambda_i$.

* 表示“共轭”

2.行列式

◆ 行列式的主要性质

（7）行列式等于特征值之积，即 $|A|=\prod_{i=1}^{n}\lambda_i$.

（8）逆矩阵的行列式 $\left|A^{-1}\right|=1/|A|$.

（9）两个矩阵乘积的行列式等于它们的行列式的乘积，即

det(*AB*)=det(*A*)det(*B*).

（10）对于矩阵 $A_{m\times m}, B_{m\times n}, C_{n\times m}, D_{n\times n}$，分块矩阵的行列式满足

A非奇异： $\det\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A)\det\left(D - CA^{-1}B\right)$

D非奇异： $\det\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(D)\det\left(A - BD^{-1}C\right)$

（11）Hadamard不等式：若A正定或半正定，则 $\det(A)\le\prod_{i}A_{ii}$

小结：作为一个性能指标，行列式在一定程度上刻画了矩阵的奇异性.

3.秩

定理1：在p维行（或列）向量的集合之中，最多存在p个线性无关的行（或列）向量.

定理2：矩阵$A_{m\times n}$的线性无关行数与线性无关列数相同.

定义：矩阵$A_{m\times n}$的秩定义为该矩阵中线性无关的行和列的数目，并表示为rank(A).

- 关于矩阵A的秩，下列叙述等价：

  （1）$rank(A)=k$;

  （2）存在A的k列（行）且不多于k列（行）组成一线性无关组;

  （3）存在A的一个k×k非奇异子矩阵，而且A的所有（$(k+1)\times(k+1)$）子矩阵奇异.

3.秩

◆ 秩的主要性质

（1）秩是一个正整数.

（2）秩等于或小于矩阵的行数或列数.

（3）若 $rank(A_{n\times n}) = n$，则称A满秩（full rank），此时A非奇异，且其n个列向量 $a_1,\cdots,a_n$ 线性无关.

（4）若 $rank(A_{m\times n}) < \min\{m,n\}$，则称A是秩亏缺的.一个秩亏缺的正方矩阵为奇异矩阵.

（5）若 $rank(A_{m\times n}) = m(< n)$ 则称矩阵A具有满行秩.

（6）若 $rank(A_{m\times n}) = n(< m)$，则称矩阵A具有满列秩.

（7）任何矩阵A左乘满列秩矩阵或者右乘满行秩矩阵后，矩阵A的秩保持不变.

3.秩

◆ 秩的等式与不等式关系

1.$rank\left(A^{H}\right)=rank\left(A^{T}\right)=rank\left(A^{*}\right)=rank\left(A\right)$ *表示"共轭"

2.$rank\ \left(cA\right)=rank\ \left(A\right)(c\neq 0)$

3.$rank\left(AA^{T}\right)=rank\left(A^{T}A\right)=rank\left(A\right)$

$rank\left(AA^{H}\right)=rank\left(A^{H}A\right)=rank\left(A\right)$

4.$rank\left(AB\right)\leq \min\{rank\left(A\right),rank\left(B\right)\}$

5.$rank\left(A+B\right)\leq rank([A,B])\leq rank\left(A\right)+rank\left(B\right)$

小结：作为一个性能指标，矩阵的秩刻画矩阵行与行之间或者列与列之间的线性无关性，从而反映矩阵的满秩性和秩亏缺性.

4.二次型$x^H Ax$

◆ **定义**

一个复共轭对称矩阵A称为

正定矩阵（A>0）：若二次型$x^H Ax > 0, \forall x \neq 0;$

半正定（非负定）矩阵（A≥0）：若$x^H Ax \geq 0, \forall x \neq 0;$

负定矩阵（A<0）：若二次型 $x^H Ax < 0, \forall x \neq 0;$

半负定（非正定）矩阵（A≤0）：若 $x^H Ax \leq 0, \forall x \neq 0;$

不定矩阵：若二次型$x^H Ax$ 既可能取正值，也可能取负值.

小结：作为一个性能指标，二次型刻画Hermitian矩阵的正定性.

```
for ii=1:5
    for jj=1:5
        AA(ii,jj)=rand();
    end
end
BB=AA*AA';
det(AA)
eig(BB)
trace(BB)

ans =
    0.0335


ans =
    0.0161
    0.0678
    0.1938
    0.7444
    7.1631


ans =
    8.1851
```

$$f(x,y) = ax^2 + bxy + cy^2 + dx + ey + g$$

$$= (x,y)\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + (x,y)\begin{pmatrix} d \\ e \end{pmatrix} + g$$

$$= (x,y,1)\begin{pmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & g \end{pmatrix}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$= (x,y,1)\begin{pmatrix} a & b/2 & 0 \\ b/2 & c & 0 \\ d & e & g \end{pmatrix}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$f(x,y,z) = ax^2 + bxy + cy^2 + dxz + eyz + gz^2$$

$$= (x,y,z)\begin{pmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & g \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

目标函数，价值函数：求(X,Y,Z), 使得f(X,Y,Z)最小.

$$X^H AX = (X^H AX + X^H AX)/2$$

$$= (X^H AX + (X^H AX)^H)/2$$

$$= (X^H AX + X^H A^H X)/2$$

$$= X^H \frac{A + A^H}{2} X$$

$$= X^H BX$$

$$B = \frac{A + A^H}{2}$$

数学规划(Mathematical Programming)：线性规划、非线性规划、多目标规划、动态规划、参数规划、组合优化和整数规划、随机规划、模糊规划、非光滑优化、多层规划、全局优化、变分不等式和互补问题等.

一个数学规划问题一般具有以下形式

$$\begin{array}{ll} \text{最小化} & f(x) \\ \text{满足} & c_i(x) \le b_i,\ i = 1, \ldots, k \\ & x \in \mathbb{R}^n \end{array}$$

$$\begin{cases} X = [x_1, x_2, \ldots, x_n]^T \\ s.t. g_i(X) \le 0(i, 1, 2, \ldots, n) \\ h_j(X) = 0(j = 1, 2, \ldots, n) \\ max f(X) \quad or \quad min f(X) \end{cases}$$

满足以下格式的规划问题都被称作*二次规划*(quadratic programming)：

$$\begin{array}{ll} \text{最小化} & \frac{1}{2} x^\top Q x + c^\top x \\ \text{满足} & Ax \le b \\ & x \in \mathbb{R}^n \end{array}$$

5.特征值

◆特征值的第一定义公式：对于$n \times n$矩阵A，若存在标量λ，使得

$$A\mathbf{u} = \lambda\mathbf{u}$$

具有n×1非零解（向量）$\mathbf{u}$，则称λ为矩阵A的特征值.

◆特征值的第二定义公式：关于λ的多项式

$$\det(A - \lambda I) = 0$$

的根即为A的特征值.

矩阵A的特征值常用符号eig(A)表示.

特征值的第二定义公式反映了以下事实：

(1) 若矩阵A存在零特征值，则该矩阵奇异.

(2) 只有零矩阵的全部特征值为0，任何奇异的非零矩阵A一定存在在非零的特征值.（如何证明？）

(3) 若矩阵A所有特征值都不为0，则它一定非奇异.

5.特征值

◆ 特征值的一些基本性质

（1）矩阵A奇异，当且仅当存在零特征值.

（2）若$rank(A)$=r，则A最多有r个非零特征值.

（3）$eig\left(A^T\right)=eig(A)$

（4）$eig(AB)=eig(BA)$

（5）令I为单位矩阵，则

$$eig(I+cA)=1+c\,eig(A)$$
$$eig(A-cI)=eig(A)-c$$

（6）逆矩阵的特征值$eig\left(A^{-1}\right)=1/eig(A)$.

5.特征值

◆ **引理**：正定矩阵的所有特征值都是正实数.

从而可以根据矩阵的特征值来判定其正定性.

- 正定矩阵：所有特征值取正实数的矩阵；
- 半正定矩阵：各特征值取非负实数的矩阵；
- 负定矩阵：全部特征值为负实数的矩阵；
- 半负定矩阵：各特征值取非正实数的矩阵；
- 不定矩阵：有些特征值取正实数，另一些取负实数的矩阵.

小结：作为一个性能指标，矩阵的特征值既刻画原矩阵的奇异性，又反映原矩阵所有对角元素的结构，还关联矩阵的正定性.

6.迹

$n\times n$矩阵A的迹定义为其对角线元素之和

$$tr(A)=a_{11}+a_{22}+\cdots+a_{nn}=\sum_{i=1}^{n}a_{ii}$$

◆ 迹的重要性质

（1）矩阵A的转置、复共轭和复共轭转置的迹分别为

$$tr\left(A^T\right)=tr(A)$$

$$tr\left(A^*\right)=[tr(A)]^*$$

$$tr\left(A^H\right)=[tr(A)]^*$$

(2) $x^H Ax=tr\left(Axx^H\right)$和 $y^H x=tr\left(xy^H\right)$.

（3）迹是**相似不变量**，即

$$tr\left(BAB^{-1}\right)=tr\left(B^{-1}AB\right)=tr(A)$$

用性质5证明性质2、3、4

6.迹

◆ 迹的重要性质

（4）迹等于特征值之和，即有

$$tr(A)=\lambda_1+\lambda_2+\cdots+\lambda_n$$

（5）tr(AB)=tr(BA).特别地，当 $B=A^H$ 时有

$$tr(A^H A)=tr(AA^H)=\sum_{i=1}^{n}\sum_{j=1}^{n}\left|a_{ij}\right|^2$$

讨论 1：性质5对A和B的维数有要求吗?

讨论 2：下面的迹哪些相等?

$$tr(ABC),\quad tr(ACB),\quad tr(BAC)$$

$$tr(BCA),\quad tr(CAB),\quad tr(CBA)$$

讨论 3：写出矩阵乘积$ABCD$的等迹关系式.

小结：作为一个性能指标，矩阵的迹反映矩阵所有特征值之和.

用矩阵乘法的定义来证明

$$tr(AB)=tr(\sum_k a_{ik}b_{kj})$$

$$=\sum_l\sum_k a_{lk}b_{kl}$$

$$=\sum_l\sum_k b_{kl}a_{lk}$$

$$=\sum_k\sum_l b_{kl}a_{lk}$$

$$=tr(\sum_l b_{il}a_{lj})$$

$$=tr(BA)$$

1.4 内积与范数

A.向量的内积与范数

定义：令V是复向量空间.函数 $<\mathbf{x},\mathbf{y}>: \mathbf{V}\times\mathbf{V}\mapsto C$ 称为向量 $\mathbf{x}$与 $\mathbf{y}$的内积，若对所有 $\mathbf{x},\mathbf{y},\mathbf{z}\in\mathbf{V}$，满足以下公理:

(1) $<\mathbf{x},\mathbf{x}>\geq 0$, 当且仅当 $\mathbf{x}=0$ 时等号成立;

(2) $<\mathbf{x}+\mathbf{z},\mathbf{y}>=<\mathbf{x},\mathbf{y}>+<\mathbf{z},\mathbf{y}>$;

(3) $<c\mathbf{x},\mathbf{y}>=c^{*}<\mathbf{x},\mathbf{y}>,\forall c\in C$;

(4) $<\mathbf{x},\mathbf{y}>=<\mathbf{y},\mathbf{x}>^{*}$.

定义：令V是复向量空间.函数 $\|\mathbf{x}\|: \mathbf{V}\mapsto R$ 称为向量 $\mathbf{x}$ 的范数，若对所有 $\mathbf{x},\mathbf{y}\in\mathbf{V}$，满足以下公理:

(1) $\|\mathbf{x}\|\geq 0$, 当且仅当 $\mathbf{x}=0$ 时等号成立;

(2) $\|c\mathbf{x}\|=|c|\cdot\|\mathbf{x}\|,\forall c\in C$;

(3) $\|\mathbf{x}+\mathbf{y}\|\leq\|\mathbf{x}\|+\|\mathbf{y}\|$.

支撑，支撑集合，支撑集，支集
z=f(x,y)

◆ 几种常用的向量范数

l_0 （准）范数： $\|\mathbf{x}\|_0 = Supp(\mathbf{x})$

l_1 范数： $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$

l_2 （Euclidean范数或Frobenius范数）：

$$\|\mathbf{x}\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2}$$

l_∞ 范数： $\|\mathbf{x}\|_\infty = \max\left(|x_1|, |x_2|, \cdots, |x_n|\right)$

l_p 范数（Holder范数）：

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, (p \geq 1)$$

加权范数(椭圆范数)： $\|\mathbf{x}\|_A = \sqrt{\mathbf{x}^H A \mathbf{x}}$ （A为Hermitian正定)

◆ 常用向量内积

复常数向量：$<\mathbf{x},\mathbf{y}>=\mathbf{x}^H\mathbf{y}=\sum_{i=1}^{n}x_i^* y_i$

复函数向量：$<\mathbf{x}(t),\mathbf{y}(t)>=\int_a^b \mathbf{x}^H(t)\mathbf{y}(t)dt$

复随机向量：$<\mathbf{x}(\xi),\mathbf{y}(\xi)>=E\{\mathbf{x}^H(\xi)\mathbf{y}(\xi)\}$

两个向量的Euclidean距离：

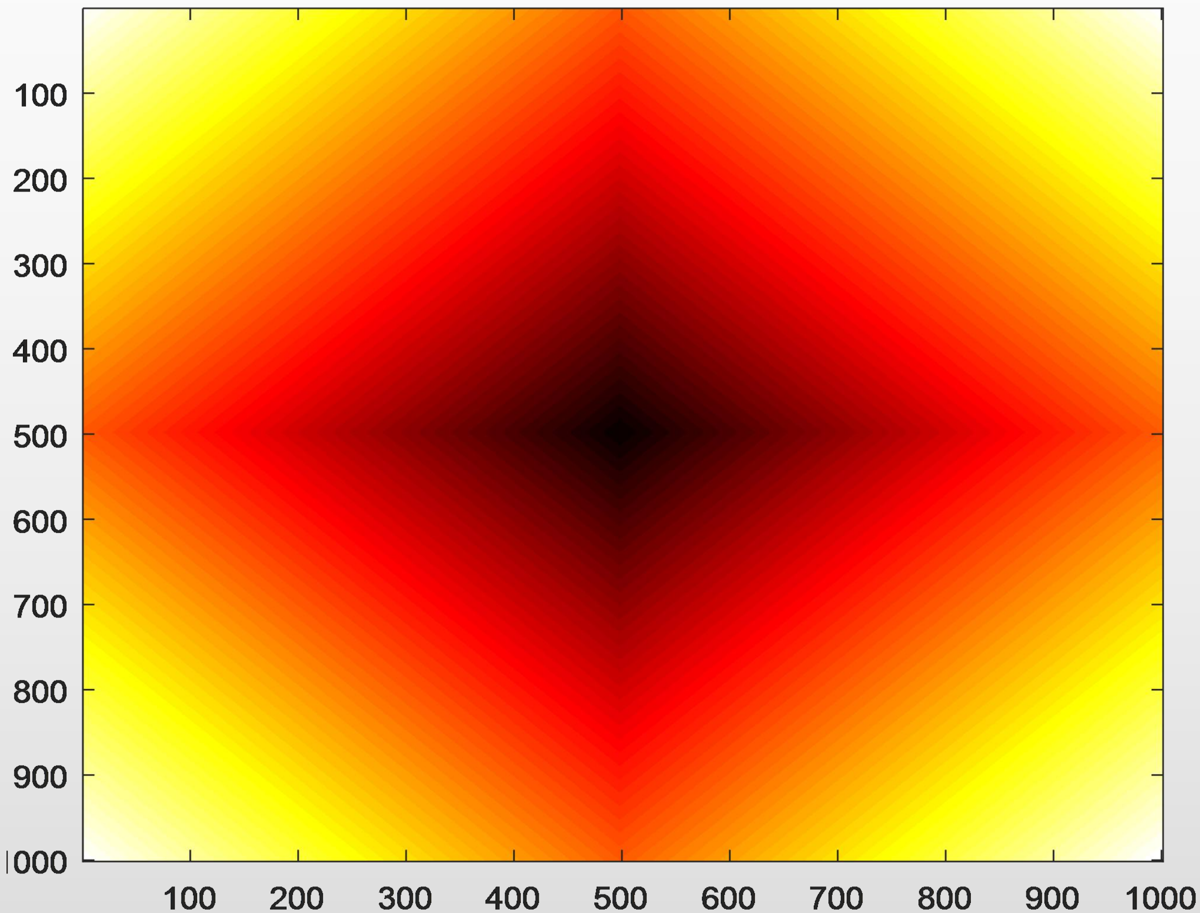
$$\|\mathbf{x}-\mathbf{y}\|_2=\sqrt{(x_1-y_1)^2+\cdots+(x_n-y_n)^2}$$

两个向量之间的夹角：

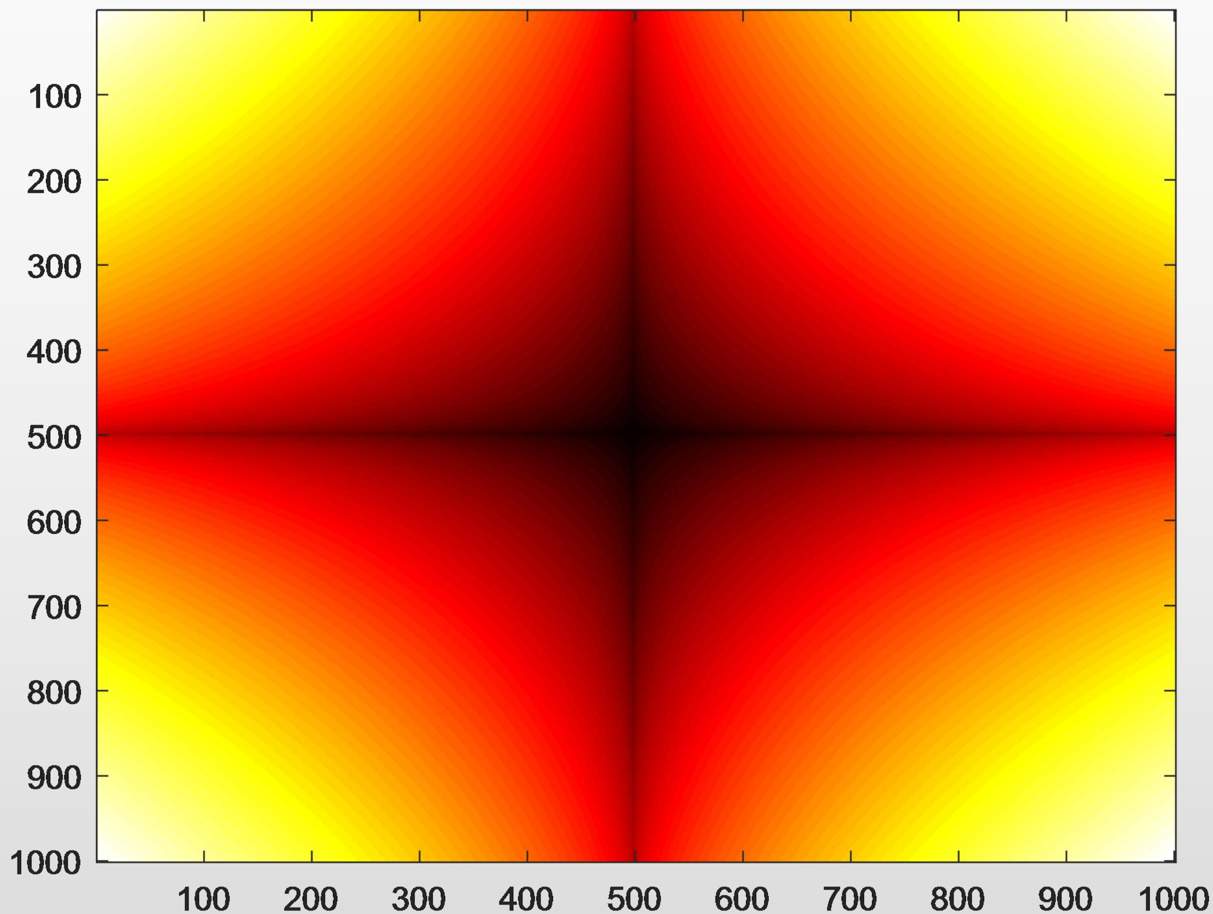
$$\cos\theta \stackrel{def}{=} \frac{<\mathbf{x},\mathbf{y}>}{\sqrt{<\mathbf{x},\mathbf{x}>}\sqrt{<\mathbf{y},\mathbf{y}>}}$$

正交：两个向量之间的夹角等于π/2（内积为0）.

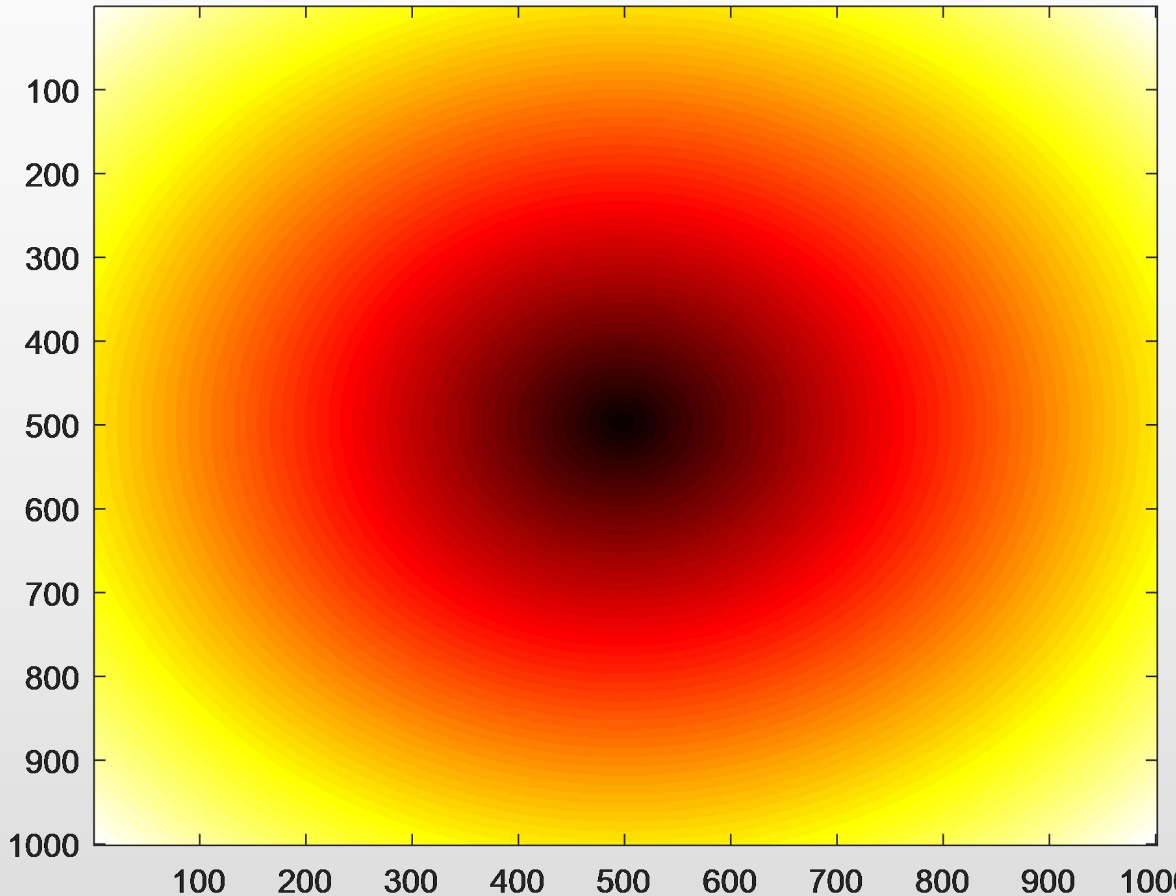
```
p=1;
for ii=1:1001
    for jj=1:1001
        AAx(ii,jj)=ii-500;
        AAy(ii,jj)=jj-500;
        AA(ii,jj)=( ( abs(AAx(ii,jj)) )^p+(abs(AAy(ii,jj))^p) )^(1/p);
    end
end
colormap(hot);
imagesc(AA);
```



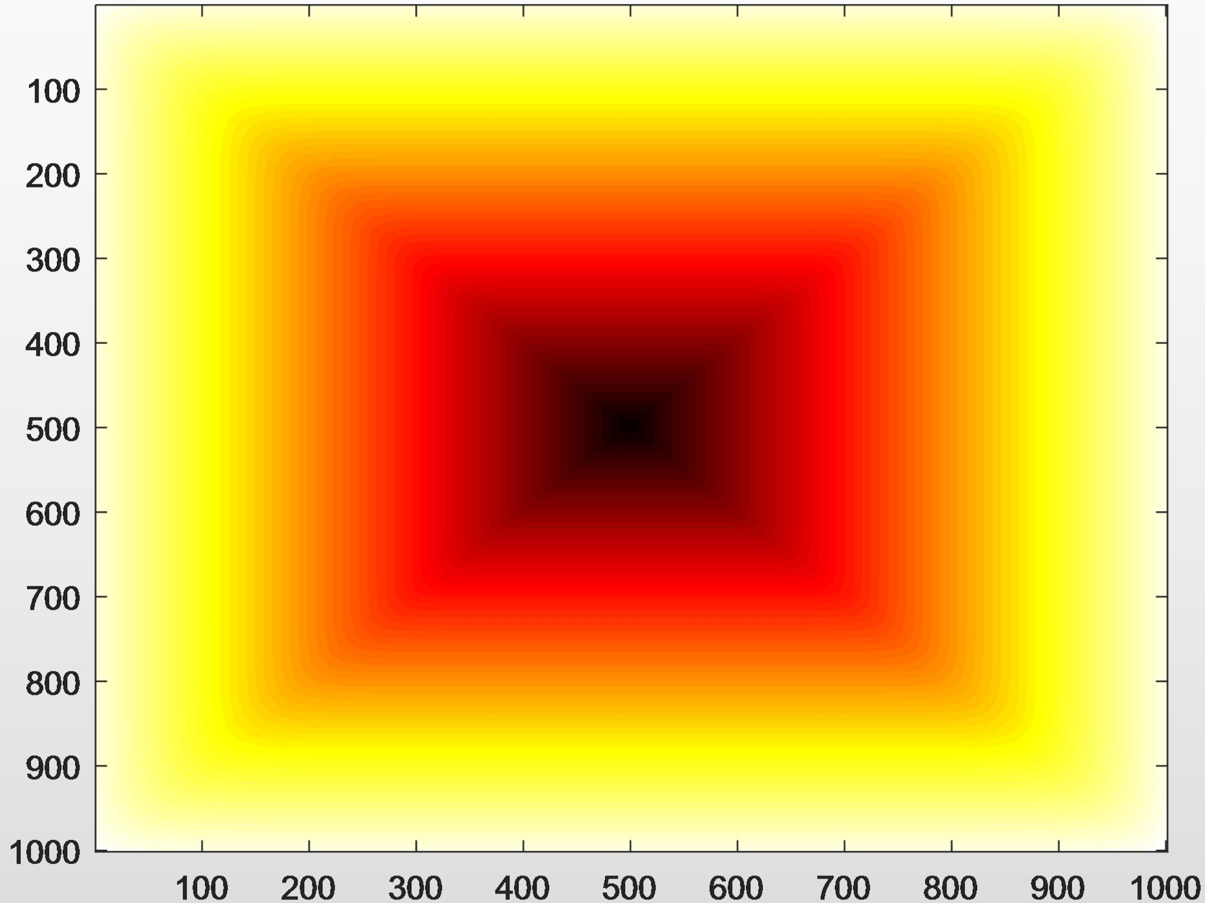
```
p=0.5;
for ii=1:1001
    for jj=1:1001
        AAx(ii,jj)=ii-500;
        AAy(ii,jj)=jj-500;
        AA(ii,jj)=( ( abs(AAx(ii,jj)) )^p+(abs(AAy(ii,jj))^p) )^(1/p);
    end
end
colormap(hot);
imagesc(AA);
```



```
p=2;
for ii=1:1001
    for jj=1:1001
        AAx(ii,jj)=ii-500;
        AAy(ii,jj)=jj-500;
        AA(ii,jj)=( ( abs(AAx(ii,jj)) )^p+(abs(AAy(ii,jj))^p) )^(1/p);
    end
end
colormap(hot);
imagesc(AA);
```



```
p=20;
for ii=1:1001
    for jj=1:1001
        AAx(ii,jj)=ii-500;
        AAy(ii,jj)=jj-500;
        AA(ii,jj)=( ( abs(AAx(ii,jj)) )^p+(abs(AAy(ii,jj))^p) )^(1/p);
    end
end
colormap(hot);
imagesc(AA);
```



向量的相似度

聚类：基于相似性原则，切分数据集(类别标准未知)
分类：判断数据类别，划分数据到已知目标类别中(类别标准已知)

◆ 问题的提出

高维观测数据向量→低维特征向量.
假定有多个特征向量，如何对它们进行聚类和分类?

◆ 向量相似性测度

（1）Euclidean距离：

$$D(\mathbf{s}_i,\mathbf{x})=\|\mathbf{x}-\mathbf{s}_i\|_2=\sqrt{(\mathbf{x}-\mathbf{s}_i)^T(\mathbf{x}-\mathbf{s}_i)}$$

基于欧氏距离的近邻分类法：若 $k=\arg\min_{\mathbf{x}} D(\mathbf{s}_i,\mathbf{x}), i=1,\cdots,M$ 则称$\mathbf{x}$ 是到 $\mathbf{s}_k$ 的近邻（最近的邻居）. 并将未知类型的模式向量 $\mathbf{x}$ 归为其近邻所属的模式类型.

（2）两个向量的夹角的余弦函数：

$$S(\mathbf{s}_i,\mathbf{x})=\cos(\theta_i)=\frac{\mathbf{x}^T\mathbf{s}_i}{\|\mathbf{x}\|_2\cdot\|\mathbf{s}_i\|_2}$$

若 $\cos(\theta_k)>\cos(\theta_i), \forall i\neq k$ 成立，则判未知模式向量 $\mathbf{x}$ 与样本模式向量 $\mathbf{s}_k$ 最为相似.

向量的相似度

◆ 向量相似性测度

（3）Tanimoto测度：

$$S(\mathbf{s}_i,\mathbf{x})=\frac{\mathbf{x}^T\mathbf{s}_i}{\mathbf{x}^T\mathbf{x}+\mathbf{s}_i^T\mathbf{s}_i+\mathbf{x}^T\mathbf{s}_i}$$

Tanimoto测度广泛用于信息恢复、疾病分类、动物和植物分类等.

（4）Mahalanobis距离：

M个样本模式向量的均值向量与协方差矩阵为

$$\mathbf{m}=\frac{1}{M}\sum_{k=1}^{M}\mathbf{s}_i,C=\frac{1}{M}\sum_{i=1}^{M}(\mathbf{s}_i-\mathbf{m})(\mathbf{s}_i-\mathbf{m})^T$$

而

$$D(\mathbf{m},\mathbf{x})=\sqrt{(\mathbf{x}-\mathbf{m})^T C^{-1}(\mathbf{x}-\mathbf{m})}$$

$$D(\mathbf{m},\mathbf{s}_i)=\sqrt{(\mathbf{s}_i-\mathbf{m})^T C^{-1}(\mathbf{s}_i-\mathbf{m})}$$

分别表示未知模式向量 $\mathbf{x}$ 与第i个样本模式向量 $\mathbf{s}_i$ 到均值向量m的Mahalanobis距离.

向量的相似度

◆ 基于马氏距离的近邻分类法

$$D(\mathbf{s}_k,\mathbf{x})=\min_i\left|D(\mathbf{m},\mathbf{s}_i)-D(\mathbf{m},\mathbf{x})\right|, i=1,\cdots,M$$

将 $\mathbf{x}$ 归为近邻 $\mathbf{s}_k$ 所属的模式类型.

◆ 向量的线性相关性

定义：一组m维向量 $\{\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_n\}$ 称为线性无关，若方程

$$c_1\mathbf{u}_1+c_2\mathbf{u}_2+\cdots+c_n\mathbf{u}_n=\mathbf{0}$$

只有零解 $c_1=\cdots=c_n=0$.反之，若能找到一组不全为零的系数 $c_1,\cdots,c_n$ 使得上述方程成立，则称m维向量组 $\{\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_n\}$ 线性相关.

备注：（1）一组向量线性相关意味着一定存在其中的某些向量可以由其他向量线性表出.

例如，若 c_1,c_2,c_n 不为零时，上式仍然能够成立，则 $\mathbf{u}_1=\frac{c_2}{c_1}\mathbf{u}_2+\frac{c_n}{c_1}\mathbf{u}_n$

即 $\mathbf{u}_1$ 可以表示成 $\mathbf{u}_2$ 和 $\mathbf{u}_n$ 的线性组合.

（2）然而，一组向量线性相关并不意味着其中的所有向量均可由其他向量线性表出.

(1)内积

dot

表示两个向量的内积

(2)范数

norm(x, 1)， norm(x, 2)， norm(x, inf)，
norm(x, -inf)， norm(x, p)

(3) 距离（相似性）

pdist2

help pdist2

pdist2 - Pairwise distance between two sets of observations. This MATLAB function returns the distance between each pair of observations in X and Y using the metric specified by Distance.

```
D = pdist2(X,Y,Distance)
D = pdist2(X,Y,Distance,DistParameter)
D = pdist2(___,Name,Value)
[D,I] = pdist2(___,Name,Value)
```

```
d1 = pdist2(x,y,'Euclidean');
d2 = pdist2(x,y,'seuclidean');
d3 = pdist2(x,y,'mahalanobis');
d4 = pdist2(x,y,'cityblock');
d5 = pdist2(x,y,'minkowski',p);
d6 = pdist2(x,y,'chebychev');
d7 = pdist2(x,y,'cosine');
d8 = pdist2(x,y,'correlation');
d9 = pdist2(x,y,'hamming');
d10 = pdist2(x,y,'jaccard');
d11 = pdist2(x,y,'spearman');
```

- Minkowski distance

$$d_{st} = \sqrt[p]{\sum_{j=1}^{n} |x_{sj} - y_{tj}|^p}.$$

For the special case of $p = 1$, the Minkowski distance gives the city block distance. For the special case of $p = 2$, the Minkowski distance gives the Euclidean distance. For the special case of $p = \infty$, the Minkowski distance gives the Chebychev distance.

- Chebychev distance

$$d_{st} = \max_j\{|x_{sj} - y_{tj}|\}.$$

The Chebychev distance is a special case of the Minkowski distance, where $p = \infty$.

- Cosine distance

$$d_{st} = \left(1 - \frac{x_s y'_t}{\sqrt{(x_s x'_s)(y_t y'_t)}}\right).$$

- Correlation distance

$$d_{st} = 1 - \frac{(x_s - \bar{x}_s)(y_t - \bar{y}_t)'}{\sqrt{(x_s - \bar{x}_s)(x_s - \bar{x}_s)'}\sqrt{(y_t - \bar{y}_t)(y_t - \bar{y}_t)'}},$$

where

$$\bar{x}_s = \frac{1}{n}\sum_j x_{sj}$$

and

$$\bar{y}_t = \frac{1}{n}\sum_j y_{tj}.$$

思考:

（1）在明可夫斯基距离下，一条直线的平行线?

（2）在余弦距离下，一条直线的平行线?

（3）作图画出在余弦距离意义下的圆。

B.矩阵的内积与范数

◆ **定义**

令 $\mathbb{C}^{m\times n}$ 是复矩阵空间.函数 $<X,Y>:\mathbb{C}^{m\times n}\times\mathbb{C}^{m\times n}\mapsto\mathbb{C}$ 称为矩阵X与Y的内积，若对所有 $X,Y,Z\in\mathbb{C}^{m\times n}$, 满足以下公理:

(1) $<X,X>\geq 0$， i.f.f. X=0时等号成立;

(2) $<X+Z,Y>=<X,Y>+<Z,Y>$;

(3) $<cX,Y>=c^*<X,Y>,\forall c\in C$;

(4) $<X,Y>=<Y,X>^*$.

◆ **定义**

令 $\mathbb{C}^{m\times n}$ 是复矩阵空间.函数 $\|X\|:\mathbb{C}^{m\times n}\mapsto R$ 称为矩阵X的范数，若对所有 $X,Y\in\mathbb{C}^{m\times n}$，满足

(1) $\|X\|\geq 0$， i.f.f. X=0时等号成立;

(2) $\|cX\|=|c|\|X\|,\forall c\in C$;

(3) $\|X+Y\|\leq\|X\|+\|Y\|$;

(4) $\|XZ\|\leq\|X\|\cdot\|Z\|,\forall Z\in V^{n\times k}$.

◆ 几种常用的矩阵范数$\left(A \in C^{m \times n}\right)$

(1) $\| \mathrm{A} \|_{m_1} = \sum_{i}^{n} \sum_{j}^{n} \left| a_{i,j} \right|, (m = n)$

(2) $\| \mathrm{A} \|_{m_\infty} = n \cdot \max_{i,j} \left| a_{i,j} \right|, (m = n)$

(3) Frobenius范数: $\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} \left| a_{ij} \right|^2}$

(4) Mahalanobis范数: $\|A\|_\Omega = \sqrt{tr(A^H \Omega A)}, (\Omega > 0)$

(5) 从属范数: $\|A\|_P = \max_{\|\mathbf{x}\|_P = 1} \|A\mathbf{x}\|_P$

- ◆ 列和范数: $\|A\|_1 = \max_{j} \sum_{i=1}^{m} \left| a_{i,j} \right|$
- ◆ 谱范数: $\|A\|_2 = \sqrt{\lambda_{\max}\left(A^H A\right)}$
- ◆ 行和范数: $\|A\|_\infty = \max_{i} \sum_{j=1}^{n} \left| a_{i,j} \right|$

◆ 矩阵内积与范数之间的关系

（1）Cauchy-Schwartz不等式

$$\left\| < A, B > \right\|^2 \le \left\| A \right\|^2 \left\| B \right\|^2$$

等号成立，当且仅当A=cB（c为某复常数）.

（2）Pathagoras定理

$$< A, B >= 0 \Rightarrow \left\| A + B \right\|^2 = \left\| A \right\|^2 + \left\| B \right\|^2$$

（3）极化恒等式

$$\mathrm{Re}\left(< A, B > \right) = \frac{1}{4}\left(\left\| A + B \right\|^2 - \left\| A - B \right\|^2 \right)$$

$$\mathrm{Re}\left(< A, B > \right) = \frac{1}{2}\left(\left\| A + B \right\|^2 - \left\| A \right\|^2 - \left\| B \right\|^2 \right)$$

小结：作为一个性能指标，矩阵的范数刻画矩阵元素的结构.

1.5 矩阵和向量的应用案例

人脸识别的稀疏表示 sparse vector/matrix



(1)人脸图像向量化

(2)构造训练数据矩阵

$$D=[D_1,D_2,\cdots,D_c]=[d_{1,1},d_{1,2},\cdots,d_{1,N_1},d_{2,1},d_{2,2},\cdots,d_{2,N_2},\cdots,d_{c,1},d_{c,2},\cdots,d_{c,N_c}]$$

(3)人脸识别（模型)

已知数据向量y和数据矩阵D,求线性方程组y=Dα的稀疏解向量α_0，即

$$\min \|\alpha_0\|_0 \ \text{ subject to } y=D\alpha_0$$

思考:

（1）在明可夫斯基距离下，一条直线的平行线?

（2）在余弦距离下，一条直线的平行线?

（3）作图画出在余弦距离意义下的圆。

作业

(1)请用矩阵符号推演如下定理证明过程

定理：设 $\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n]$ 和 $\boldsymbol{V} = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n]$ 的各列分别构成 n 维向量空间 $\mathbb{V}$ 的两组基, 而 $\boldsymbol{S}$ 为从 $\boldsymbol{U}$ 到 $\boldsymbol{V}$ 的转移 (过渡) 矩阵, 即 $\boldsymbol{V} = \boldsymbol{U}\boldsymbol{S}$。若 $\mathbb{V}$ 上的线性算子 $\mathcal{L}$ 相应于 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的表示矩阵分别为 $\boldsymbol{A}$ 和 $\boldsymbol{B}$, 则

$$\boldsymbol{B} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$$

(2)请用矩阵符号推演如下公式

令 $\boldsymbol{I}$ 为单位矩阵，则

$$\mathrm{eig}(\boldsymbol{I} + c\boldsymbol{A}) = 1 + c\,\mathrm{eig}(\boldsymbol{A})$$
$$\mathrm{eig}(\boldsymbol{A} - c\boldsymbol{I}) = \mathrm{eig}(\boldsymbol{A}) - c$$

(3）1.1，1.2, 1.3，1.7, 1.8，1.10, 1.11

提交截止日期：2023年4月9日（过期不候)