矩阵分析及其应用

孔贺

南科大，自动化与智能制造学院

kongh@sustech.edu.cn

课程结构

- 一. 线性代数基本知识回顾（3 个学时）
- 二. 线性空间与线性映射（12 个学时）
- 三. λ-矩阵与 Jordan 标准型（8 个学时）
- 四. 内积空间与矩阵分解（12 个学时）
- 五. 范数及其他（8 个学时）

1. 第一章：线性代数知识回顾
2. 第二章：线性空间与线性映射
3. 第三章：λ-矩阵与 Jordan 标准型
4. 第四章：内积空间
   - 4.1 内积与 Gram 矩阵
   - 4.2 标准正交基与酉矩阵
   - 4.3 正交投影与最佳逼近
   - 4.4 受限最小二乘
   - 4.5 矩阵的广义逆
   - 4.6 正规矩阵
   - 4.7 Hermite 矩阵与 Hermite 二次型
   - 4.8 矩阵的奇异值分解

4.1 内积与 Gram 矩阵

定义：假设 $\mathbf{V}$ 是实数域 $\mathbf{R}$ 上的线性空间。若存在映射

$$\tau : \mathbf{V} \times \mathbf{V} \to \mathbf{R}$$

且其满足以下条件：

- 1. 对称性，即对任意 $\alpha,\beta \in \mathbf{V}$，有

$$\langle \alpha,\beta \rangle = \langle \beta, \alpha \rangle$$

- 2. 对第二个变元线性，即对任意 $\alpha,\beta, v \in \mathbf{V}$，$g, h \in \mathbf{R}$，有

$$\langle \alpha, \beta g + vh \rangle = \langle \alpha, \beta \rangle g + \langle \alpha, v \rangle h$$

- 3. 正定性，即对任意的 $\alpha \in \mathbf{V}$，有

$$\langle \alpha,\alpha \rangle \geq 0;$$

当且仅当 $\alpha = 0$ 的时候有 $\langle \alpha,\alpha \rangle = 0$

则我们称其为 $\mathbf{V}$ 上的一个**内积**，并记为 $\tau(\alpha,\beta) = \langle \alpha,\beta \rangle$。特别的，称定义有如上内积的（有限维）线性空间为**（欧几里得）内积空间**。

内积性质的几点说明

- 对任意 $\alpha \in \mathbf{V}$，有 $\langle 0,\alpha \rangle = 0$。
- 双线性：注意以下事实，若我们固定 α，让 β 变化，则由内积可决定以下的映射

$$\langle \alpha, \cdot \rangle : \mathbf{V} \to \mathbf{R}$$

也可记为

$$\beta \mapsto \langle \alpha, \beta \rangle$$

那么内积定义中的第 2 个性质是在说上述映射是从 $\mathbf{V}$ 到 $\mathbf{R}$ 的线性映射，因此称其对第二个变元线性。综合第 1，2 个性质，我们可以得出内积对第一个变元也是线性的。
- 根据双线性的性质，计算

$$\langle u_1 3 + u_2 2, v_1 4 + v_2 5 \rangle$$

欧式空间举例

- 实数域 $\mathbf{R}$ 上的标准欧式空间 $\mathbf{R}^n$（$\mathbf{R}^n$ 上的标准内积）

$$\begin{aligned}\langle x, y\rangle &= x^{\mathrm{T}} y \\ &= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}\end{aligned}$$

- 非标准内积

$$\langle x, y\rangle = x_1 y_1 + \frac{1}{2} x_1 y_2 + \frac{1}{2} x_2 y_1 + x_2 y_2$$

- 几何空间作为欧式空间
- 函数空间的内积

定义：假设 $\mathbf{V}$ 是复数域 $\mathbf{C}$ 上的线性空间。若存在映射

$$\tau : \mathbf{V} \times \mathbf{V} \rightarrow \mathbf{C}$$

且其满足以下条件：

- 1. 共轭对称性，即对任意 $\alpha,\beta \in \mathbf{V}$，有

$$\langle\alpha,\beta\rangle = \overline{\langle\beta,\alpha\rangle}$$

- 2. 对第二个变元线性，即对任意 $\alpha,\beta,v \in \mathbf{V}$，$g,h \in \mathbf{C}$，有

$$\langle\alpha,\beta g + vh\rangle = \langle\alpha,\beta\rangle g + \langle\alpha,v\rangle h$$

- 3. 对任意的 $\alpha \in \mathbf{V}$，有

$$\langle\alpha,\alpha\rangle \geq 0;$$

  当且仅当 $\alpha = 0$ 的时候有 $\langle\alpha,\alpha\rangle = 0$

则我们称其为 $\mathbf{V}$ 上的一个**复内积**。特别的，称定义有如上复内积的有限维线性空间为**复欧几里得空间**，简称**酉空间**。欧式空间与酉空间通称为内积空间（以下大多数情况下，我们会讨论酉空间）。

复内积性质与举例

- 对任意 $\alpha \in \mathbf{V}$，有 $\langle 0, \alpha \rangle = 0$。
- 复内积对第一个变元是共轭线性的。
- 复数域 $\mathbf{C}$ 上的标准酉空间 $\mathbf{C}^n$（$\mathbf{C}^n$ 上的标准内积）

$$
\begin{aligned}
\langle x, y \rangle &= \bar{x}^{\mathrm{T}} y \\
&= \begin{bmatrix} \overline{x_1} & \overline{x_2} & \cdots & \overline{x_n} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}
\end{aligned}
$$

线性组合的内积的矩阵表示

$$
\begin{aligned}
&\left\langle \sum_{i=1}^{s} \alpha_i g_i, \sum_{j=1}^{t} \beta_j h_j \right\rangle \\
&= \begin{bmatrix} \overline{g_1} & \overline{g_2} & \cdots & \overline{g_s} \end{bmatrix} [\langle \alpha_i, \beta_j \rangle]_{s\times t} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_t \end{bmatrix}
\end{aligned}
$$

因此，只要知道两个向量组之间，每对向量的内积，那么由这两个向量组各自任意做线性组合得到的一对向量的内积，可以由上述各对向量的内积（向量组的 Gram 矩阵）和线性组合的系数得出。

向量组的 Gram 矩阵

- 假设 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 是 n 维复内积空间的一个向量组，我们称矩阵

$$\mathcal{G}\left(\alpha_1,\alpha_2,\cdots,\alpha_s\right) := \left[\langle\alpha_i,\alpha_j\rangle\right]_{s\times s} \in \mathbf{C}^{s\times s}$$

  为上述向量组的 Gram 矩阵，记为 $\mathcal{G}\left(\{\alpha_i\}\right)$。
- 假设 $\alpha_1,\alpha_2,\cdots,\alpha_s$；$\beta_1,\beta_2,\cdots,\beta_t$ 是 n 维复内积空间的两个向量组，我们称矩阵

$$\mathcal{G}\left(\alpha_1,\alpha_2,\cdots,\alpha_s;\beta_1,\beta_2,\cdots,\beta_t\right) := \left[\langle\alpha_i,\beta_j\rangle\right]_{s\times t} \in \mathbf{C}^{s\times t}$$

  为上述两个向量组的交互 Gram 矩阵，记为 $\mathcal{G}\left(\{\alpha_i\},\{\beta_j\}\right)$。
  - 注：矩阵乘积作为两向量组的交互 Gram 矩阵。

向量的长度和夹角

- 定义：假设 $\mathbf{V}$ 为酉空间，给定向量 $\alpha, \beta \in \mathbf{V}$，称 $\|\alpha\| := \sqrt{\langle \alpha,\alpha \rangle}$ 为向量的长度（模）；称 $d(\alpha, \beta) := \|\alpha - \beta\|$ 为向量 α 与 β 之间的距离。
- 定理：假设 $\mathbf{V}$ 为酉空间，给定向量 $\alpha \in \mathbf{V}$，则其长度 $\|\alpha\|$ 具有以下性质
  - (1) 非负性：$\|\alpha\| \geq 0$，当且仅当 $\alpha = 0$ 时 $\|\alpha\| = 0$
  - (2) 齐次性：$\|\alpha k\| = |k| \, \|\alpha\|$，其中 k 为任意数
  - (3) 满足三角不等式：$\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$
  - (4) 满足 Cauchy-Schwartz 不等式：$|\langle \alpha,\beta \rangle| \leq \|\alpha\| \, \|\beta\|$
  - (5) 满足平行四边形公式：$\|\alpha + \beta\|^2 + \|\alpha - \beta\|^2 = 2(\|\alpha\|^2 + \|\beta\|^2)$

- 基于 Cauchy-Schwartz 不等式，对实内积空间中的向量有

$$-1 \leq \frac{\langle\alpha,\beta\rangle}{\|\alpha\| \, \|\beta\|} \leq 1$$

- 定义：假设 α, β 为欧式空间中的两个非零向量，我们称

$$\varphi := \arccos \frac{\langle\alpha,\beta\rangle}{\|\alpha\| \, \|\beta\|}$$

  为两个向量的夹角。
- 定义：假设 α, β 为酉空间中的两个非零向量，若 $\langle\alpha,\beta\rangle = 0$，则称 α, β 相互正交，记为 $\alpha \perp \beta$。
  - 在欧式空间中，两个向量正交意味着其夹角为 $\frac{\pi}{2}$。

4.2 标准正交基与酉矩阵

定义：假设 V 为 C 上的内积空间，给定 V 中的非零向量 $\alpha_1,\alpha_2,\cdots,\alpha_s$，考虑其组成的向量组。

- 若此向量组中的任意两个向量都正交，即 $\alpha_i \perp \alpha_j$，其中 $i,j=1,2,\cdots,s$ 且 $i \neq j$，则称其为正交向量组，简称**正交组**。
- 若正交向量组又满足标准性，即 $\|\alpha_i\|=1$，$i=1,2,\cdots,s$，则称其为标准正交向量组，简称**标准正交组**。
- 若标准正交向量组同时还是 V 的一组基，则称其为**标准正交基**。

定理：给定内积空间中的一个向量组。

- 若其为正交向量组，则其线性无关（逆否命题 ?）。
- 其为正交组的充要条件是其 Gram 矩阵为非奇异对角矩阵。
- 其为标准正交组的充要条件是其 Gram 矩阵为单位阵。

内积空间中标准正交组的好处

定理：给定复内积空间 $\mathbf{V}$ 中的一个正交组 $\alpha_1,\alpha_2,\cdots,\alpha_s$，以及 $\alpha \in \mathbf{V}$。若有

$$\begin{aligned}\alpha &= \alpha_1 k_1 + \alpha_2 k_2 + \cdots + \alpha_s k_s \\ &= \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_s \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_s \end{bmatrix}\end{aligned}$$

则

$$k_i = \frac{\langle \alpha_i, \alpha \rangle}{\langle \alpha_i, \alpha_i \rangle},\ i = 1, 2, \cdots, s$$

特别的，若 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 为标准正交组，则有

$$k_i = \langle \alpha_i, \alpha \rangle,\ i = 1, 2, \cdots, s$$

举例及问题：对内积空间，标准正交基的存在性？

Gram-Schmidt 正交化方法

给定 n 维内积空间 $\mathbf{V}$ 中的 s 个线性无关的向量 $\alpha_1,\alpha_2,\cdots,\alpha_s$，那么可以通过 Gram-Schmidt 正交化方法将其改造成标准向量组：

- （1）正交化：

$$
\begin{aligned}
\beta_1 &= \alpha_1 \\
\beta_2 &= \alpha_2 - \beta_1\square \\
\beta_3 &= \alpha_3 - (\beta_1\square + \beta_2\square) \\
&\vdots \\
\beta_s &= \alpha_s - (\beta_1\square + \beta_2\square + \cdots + \beta_{s-1}\square)
\end{aligned}
$$

- （2）单位化：

$$\widetilde{\beta}_1 = \frac{\beta_1}{\|\beta_1\|},\ \widetilde{\beta}_2 = \frac{\beta_2}{\|\beta_2\|},\cdots,\widetilde{\beta}_s = \frac{\beta_s}{\|\beta_s\|}$$

注：举例。对任意 n 维内积空间，Gram-Schmidt 正交化方法保证了标准正交基的存在性，从而使得任一内积空间同构于相同维数的标准内积空间。

Gram-Schmidt 正交化的矩阵表达—矩阵的 QR 分解

定理：给定 $A \in \mathbf{C}^{n\times s}$ 且 A 列满秩，则存在**唯一**的矩阵对 $Q \in \mathbf{C}^{n\times s}$ 和 $R \in \mathbf{C}^{s\times s}$ 满足如下条件

- （1）$A = QR$
- （2）$Q^{\mathrm{H}}Q = I_s$，即 Q 的列向量是标准酉空间 $\mathbf{C}^n$ 中的标准正交组中的 s 个向量
- （3）R 是对角线为正实数的上三角矩阵

注明：

- （1）需要证明什么，如何证？
- （2）特别的，如果 $A \in \mathbf{C}^{n\times n}$ 且 A 满秩，则 $\overline{Q}^{\mathrm{T}}Q = I_n$，即 Q 为通常所说的**酉矩阵**

酉矩阵

定义：给定 $U \in \mathbf{C}^{n\times n}$，若其满足

$$U^{\mathrm{H}} U = I_n$$

则称 U 为酉矩阵。相应的，实矩阵被称为正交矩阵，若

$$U^{\mathrm{T}} U = I_n$$

注：酉矩阵的两种解读

- (1) 由标准酉空间中的任一标准正交基所拼成（回忆标准内积）。
- (2) U 为酉矩阵，若 $U^{-1} = U^{\mathrm{H}}$。

定义：给定 A，$B \in \mathbf{C}^{n\times n}$，若有酉矩阵 $U \in \mathbf{C}^{n\times n}$ 使得

$$U^{\mathrm{H}} A U = B$$

则称 A 与 B 酉相似。

酉矩阵保持度量

定理：给定 $U \in \mathbf{C}^{n\times n}$，则下列说法等价

- （1）U 为酉矩阵，即其列向量组是标准酉空间 $\mathbf{C}^n$ 中的一个标准正交基
- （2）U^{H} 为酉矩阵，即其行向量组的共轭转置是标准酉空间 $\mathbf{C}^n$ 中的一个标准正交基
- （3）$U^{\mathrm{H}}U = I_n$
- （4）$UU^{\mathrm{H}} = I_n$
- （5）对任意 $x \in \mathbf{C}^n$，有 $\|Ux\| = \|x\|$
- （6）对任意 $x, y \in \mathbf{C}^n$，有 $\langle Ux, Uy\rangle = \langle x, y\rangle$

酉矩阵的例子

例 1： 回忆 2D 平面中旋转矩阵的表示（第一章）

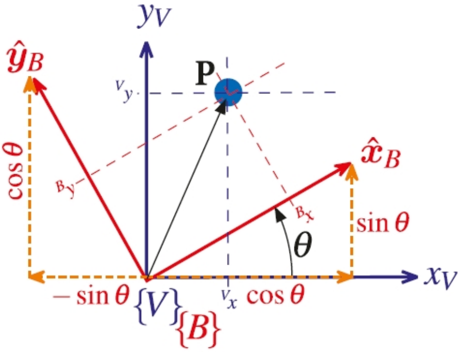

图: Fig. 2.7, Peter Corke, Robotics, Vision and Control (Second Edition)

例 2： 3D 空间中绕 x 轴的旋转矩阵

4.3 正交投影与最佳逼近

问题：假设 $\mathbf{V}$ 是 $\mathbf{C}$ 上的内积空间，给定 $\beta \in \mathbf{V}$，$\mathbf{W}$ 是 $\mathbf{V}$ 的一个有限维子空间

- 求 $\alpha^* \in \mathbf{W}$ 使得

$$d(\alpha^*, \beta) = \min\{d(\alpha, \beta) : \alpha \in \mathbf{W}\}$$

  其中，$d(\alpha, \beta) := \|\alpha - \beta\|$ 表示向量 α，β 之间的距离。
- 换言之，我们要找到 α^* 使得

$$d(\alpha^*, \beta) \leq d(\alpha, \beta), \text{对任意 } \alpha \in \mathbf{W}$$

- 用优化问题的形式可表示为

$$\alpha^* = \arg\min_{\alpha \in \mathbf{W}} \|\alpha - \beta\|$$

对这样一个问题，我们关心的是什么？

$\mathbf{W}$ 为一维子空间的情形（用欧式空间为例来说明）

对此情形，有 $\mathbf{W} = span\{\alpha\}, \alpha \neq 0$，且 $\mathbf{W}$ 中的所有元素可表示为

$$\mathbf{W} = span\{\alpha k\}，其中 k \in \mathbf{R}$$

- 解法一：化为函数极值问题

$$\begin{aligned} f(k) &= \|\alpha k - \beta\|^2 = \langle \alpha k - \beta, \alpha k - \beta \rangle \\ &= \langle \alpha, \alpha \rangle k^2 - 2 \langle \alpha, \beta \rangle k + \langle \beta, \beta \rangle \end{aligned}$$

- 解法二：勾股定理

  假设 $\beta - \alpha k^* \perp \mathbf{W}$，则有

$$\beta - \alpha k = \beta - \alpha k^* + \alpha(k^* - k)$$

  其中 $\alpha(k^* - k) \in \mathbf{W}$，且 $\beta - \alpha k^* \perp \alpha(k^* - k)$。由勾股定理可知

$$\begin{aligned} \|\beta - \alpha k\|^2 &= \|\beta - \alpha k^*\|^2 + \|\alpha(k^* - k)\|^2 \\ &\geq \|\beta - \alpha k^*\|^2 \end{aligned}$$

  即，$\beta - \alpha k^* \perp \mathbf{W}$ 中的 k^* 为所求。

- 例子

$\mathbf{W}$ 为多维子空间的情形

定理：假设 $\mathbf{V}$ 是 $\mathbf{C}$ 上的内积空间，给定 $\beta \in \mathbf{V}$，$\mathbf{W}$ 是 $\mathbf{V}$ 的一个 s 维子空间，那么

- （1）存在唯一的 $\alpha^* \in \mathbf{W}$ 使得

$$d(\alpha^*, \beta) = \min\left\{d(\alpha, \beta) : \alpha \in \mathbf{W}\right\}$$

- （2）假设 $w_1, w_2, \cdots, w_s$ 为 $\mathbf{W}$ 的一组基，记

$$\mathbf{W} = \begin{bmatrix} w_1 & w_2 & \cdots & w_m \end{bmatrix}, \alpha^* = \mathbf{W}k^* = \mathbf{W}\begin{bmatrix} k_1^* \\ k_2^* \\ \vdots \\ k_m^* \end{bmatrix}$$

则

$$k^* = \mathcal{G}^{-1}\left(\{w_i\}\right)\mathcal{G}\left(\{w_i\}, \beta\right), \alpha^* = \mathbf{W}k^*$$

- 若 $w_1, w_2, \cdots, w_s$ 均为数域上的向量，则有

$$k^* = \left(\mathbf{W}^{\mathrm{H}}\mathbf{W}\right)^{-1}\mathbf{W}^{\mathrm{H}}\beta, \ \alpha^* = \mathbf{W}k^* = \mathbf{W}\left(\mathbf{W}^{\mathrm{H}}\mathbf{W}\right)^{-1}\mathbf{W}^{\mathrm{H}}\beta$$

- 例子：给定复数域上的矩阵 $A \in \mathbf{C}^{n\times s}$，假设其列满秩，那么子空间 $im(A)$ 对应的正交投影矩阵为

$$P_A = A\left(A^{\mathrm{H}}A\right)^{-1}A^{\mathrm{H}}$$

即，给定向量 $\beta \in \mathbf{C}^n$，其在子空间 $im(A)$ 所对应的投影为

$$\alpha^* = P_A\beta = A\left(A^{\mathrm{H}}A\right)^{-1}A^{\mathrm{H}}\beta$$

对于给定的 A，线性映射 $\beta \mapsto \alpha^*$ 是从酉空间 $\mathbf{C}^n$ 到子空间 $\mathbf{W} = im(A)$ 的投影算子，由 P_A 决定。
- 投影矩阵的性质
  - (1) Hermite 性
  - (2) 幂等的
  - (3) 与子空间秩的一致性

最小二乘数据拟合

对“因变量” y 和“自变量” x_1，x_2，$\cdots$，x_p 做 n 次观测，得到如下数据

编号	y	x_0	x_1	x_2	$\cdots$	x_p
1	y_1	1	x_{11}	x_{12}	$\cdots$	x_{1p}
2	y_2	1	x_{21}	x_{22}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
n	y_n	1	x_{n1}	x_{n2}	$\cdots$	x_{np}

实际中，为了方便，经常补充常值变量 $x_0 = 1$。我们要寻找“经验公式”

$$y = x_0 m_0 + x_1 m_1 + x_2 m_2 + \cdots + x_p m_p = \begin{bmatrix} 1 & x_1 & x_2 & \cdots & x_p \end{bmatrix} \begin{bmatrix} m_0 \\ m_1 \\ m_2 \\ \vdots \\ m_p \end{bmatrix}$$

即求参数 $m_0, m_1, m_2, \cdots, m_p$ 实现“最优拟合”。

最小二乘准则的几何解释

最小二乘的几何本质是“因变量” y 的观测值组成的向量向“自变量” x_0，x_1，x_2，$\cdots$，x_p 的 n 次观测值所组成的矩阵的像空间做正交映射：

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} \in \mathbf{V} = \mathbf{R}^n, \ m = \begin{bmatrix} m_0 \\ m_1 \\ \vdots \\ m_p \end{bmatrix}$$

$$a_0 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \ a_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix}, \ a_2 = \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix}, \cdots, \ a_p = \begin{bmatrix} x_{1p} \\ x_{2p} \\ \vdots \\ x_{np} \end{bmatrix},$$

$$X = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_p \end{bmatrix} \in \mathbf{R}^{n\times(p+1)}, \ \mathbf{W} = im(X)$$

假设 X 列满秩，因此有

$$m^* = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y$$

最小二乘最优解的计算

继续考虑上面的最优拟合的例子，最优解

$$m^* = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y \tag{1}$$

- 实际中，矩阵求逆主要面临以下弊端：
  - (1) 复杂度较高
  - (2) 易引进计算误差
- 工程应用需要有复杂度更低、更数据稳定的方法
- 注意：这里我们假设 X 为列满秩，那么其必有唯一的 QR 分解：

$$X = QR$$

  其中 Q 为标准正交阵，R 是对角线正实数的上三角矩阵。带入式 (1) 得

$$m^* = R^{-1}Q^{\mathrm{T}}Y$$

- 通过 QR 分解求解最小二乘问题的方法
  - (1) QR 分解
  - (2) 计算 $Q^{\mathrm{T}}Y$
  - (3) 代入法求解 $Rm^* = Q^{\mathrm{T}}Y$

应用举例 1：k–均值聚类

- 假设有 N 个 n 维向量，x_1，$\cdots$，x_n，聚类的目标是
  - 将这些向量分为 k 组或者 k 类
  - 且每组或者每类中的向量都离彼此比较接近

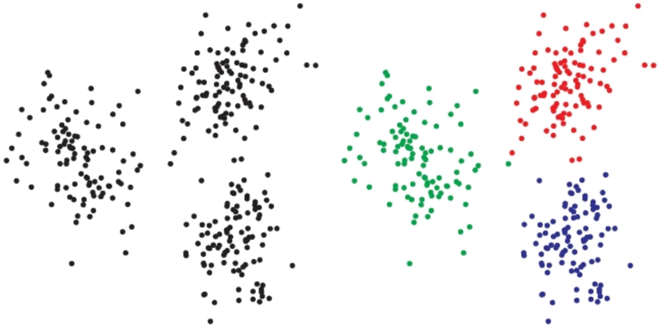

图: Fig. 4.1, Boyd and Vandenberghe, 2018

- 举例：家用能源消耗数据分析与处理 $\mathbf{R}^{24}$

聚类的核心要素

- 确认分组（cluster assignments）：c_1，$\cdots$，c_N
- 组代表（group representative），其不一定是现有的数据点
- 聚类的目标函数/代价函数

$$J^{clust} = \frac{\|x_1 - z_{c_1}\|^2 + \cdots + \|x_N - z_{c_N}\|^2}{N}$$

如何确认分组和组代表，以及二者的先后顺序？

- 当组代表固定下来的时候

$$J^{clust} = \frac{\min_{j=1,\cdots,k} \|x_1 - z_j\|^2 + \cdots + \min_{j=1,\cdots,k} \|x_N - z_j\|^2}{N}$$

- 分组固定的时候，合理选择组代表

$$J^{clust} = J_1 + \cdots + J_k$$

$$J_j = \frac{1}{N} \sum_{i \in G_j} \|x_i - z_j\|^2$$

$$z_j = \frac{1}{|G_j|} \sum_{i \in G_j} x_i$$

k–均值聚类

Algorithm 4.1 k-MEANS ALGORITHM

given a list of N vectors $x_1, \ldots, x_N$, and an initial list of k group representative vectors $z_1, \ldots, z_k$

repeat until convergence

1. *Partition the vectors into k groups.* For each vector $i = 1, \ldots, N$, assign x_i to the group associated with the nearest representative.
2. *Update representatives.* For each group $j = 1, \ldots, k$, set z_j to be the mean of the vectors in group j.

图: Algorithm. 4.1, Boyd and Vandenberghe, 2018

k–均值聚类

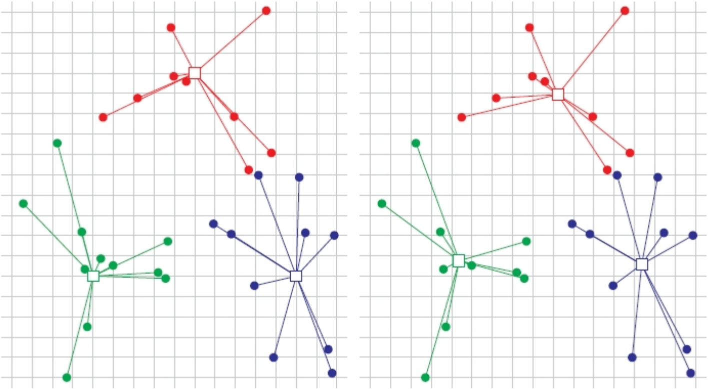

图: Fig. 4.2, Boyd and Vandenberghe, 2018

应用举例 2：卡尔曼滤波（Kalman filtering）

考虑如下系统

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + w_k \\ y_k = Cx_k + Du_k + v_k \end{cases}$$

其中

- $x_k \in \mathbf{R}^n$，$u_k \in \mathbf{R}^m$，$y_k \in \mathbf{R}^p$ 分别代表系统的状态、已知控制输入、系统输出。
- $w_k \in \mathbf{R}^n$，$v_k \in \mathbf{R}^p$ 分别代表系统的过程噪声、量测噪声。
- 假设

$$\mathcal{E}\begin{bmatrix} w_k \\ v_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathcal{E}\left[\begin{bmatrix} w_k \\ v_l \end{bmatrix}\begin{bmatrix} w_i^{\mathrm{T}} & v_j^{\mathrm{T}} \end{bmatrix}\right] = \begin{bmatrix} Q\delta_{ki} & S\delta_{kj} \\ S^{\mathrm{T}}\delta_{li} & R\delta_{lj} \end{bmatrix}$$

滤波、预测与平滑

- 所谓的“滤波与预测（filtering and prediction）”

$$\widehat{x}_{k|k-1} := \mathcal{E}\left[x_k \mid y_0, y_1, y_2, \cdots, y_{k-1}\right]$$
$$\widehat{x}_{k|k} := \mathcal{E}\left[x_k \mid y_0, y_1, y_2, \cdots, y_k\right]$$

- 对应的估计误差的协方差

$$\Sigma_{k|k-1} := \mathcal{E}\left[(x_k - \widehat{x}_{k|k-1})(x_k - \widehat{x}_{k|k-1})^{\mathrm{T}}\right]$$
$$\Sigma_{k|k} := \mathcal{E}\left[(x_k - \widehat{x}_{k|k})(x_k - \widehat{x}_{k|k})^{\mathrm{T}}\right]$$

- 另有所谓的“平滑”（smoothing）
  - （1） Fix-point smoothing

$$\widehat{x}_{0|k} := \mathcal{E}\left[x_0 \mid y_0, y_1, y_2, \cdots, y_k\right]$$

  - （2） Fix-lag smoothing

$$\widehat{x}_{k-T|k} := \mathcal{E}\left[x_{k-T} \mid y_0, y_1, y_2, \cdots, y_k\right]$$

  - （3） Fix-length smoothing

$$\left[\widehat{x}_{k-T|k}, \widehat{x}_{k-T+1|k}, \cdots, \widehat{x}_{k|k}\right] := \mathcal{E}\left[x_{k-T}, x_{k-T+1}, \cdots, x_k \mid y_0, y_1, y_2, \cdots, y_k\right]$$

卡尔曼滤波器的迭代过程

- 1. Initialization
- 2. Filtering
  - Measurement update

$$\widehat{x}_{k|k} = \widehat{x}_{k|k-1} + L_k(y_k - C\widehat{x}_{k|k-1} - Du_k) \tag{2}$$

  - Filter gain and filtered error covariance

$$L_k = \Sigma_{k|k-1} C^{\mathrm{T}} (C\Sigma_{k|k-1} C^{\mathrm{T}} + R)^{-1}$$
$$\Sigma_{k|k} = \Sigma_{k|k-1} - L_k C\Sigma_{k|k-1}$$

- 3. Prediction:
  - Time update

$$\widehat{x}_{k+1|k} = A\widehat{x}_{k|k} + Bu_k = A\widehat{x}_{k|k-1} + K_k(y_k - \widehat{y}_{k|k-1}) \tag{3}$$

  - Prediction gain and prediction error covariance

$$K_k = (A\Sigma_{k|k-1} C^{\mathrm{T}} + S)(C\Sigma_{k|k-1} C^{\mathrm{T}} + R)^{-1}$$

$$\Sigma_{k+1|k} = A\Sigma_{k|k-1} A^{\mathrm{T}} - K_k(A\Sigma_{k|k-1} C^{\mathrm{T}} + S)^{\mathrm{T}} + Q$$

- 4. New measurement arrives and go to step 2

卡尔曼滤波器的最优性—三种不同的解读

- (a) 基于正交投影
  - R. E. Kalman, A new approach to linear filtering and prediction problems, 1960 (one of the 25 seminal papers in control, 1932-1981)
- (b) 基于概率论：最大化后验概率状态估计
- (c) 基于最小二乘
  - (1) 噪声为高斯的情况下，卡尔曼滤波器是无偏的、估计误差协方差最小的线性滤波器
  - (2) 噪声为非高斯且均值为零的时候，卡尔曼滤波器依然是无偏的、估计误差协方差最小的线性滤波器

理论、直觉和应用的“三位一体”

INTRODUCTION

The Nature of Probability Theory

1. THE BACKGROUND

Probability is a mathematical discipline with aims akin to those, for example, of geometry or analytical mechanics. In each field we must carefully distinguish three aspects of the theory: (*a*) the formal logical content, (*b*) the intuitive background, (*c*) the applications. The character, and the charm, of the whole structure cannot be appreciated without considering all three aspects in their proper relation.

图: William Feller, An Introduction to Probability Theory and Its Applications, 1950

威廉 · 费勒

图: William Feller, 1906-1970，克罗地亚裔美国数学家，20 世纪最伟大的概率学家之一。师从著名数学家希尔伯特和柯朗，年仅 20 岁就获得哥廷根大学的博士学位，对近代概率论的发展做出了卓越贡献。特别是他的两本专著（《概率论及其应用》，共 2 卷），影响了世界各国几代概率论及相关领域的人士。

4.4 受限最小二乘

- 线性受限最小二乘问题为:

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|^2 \\ \text{subject to} & Cx=d \end{array}$$

- 待求解变量为 n 维向量 x
- 已知: A 为 $m\times n$ 矩阵, b 为 m 维向量, C 为 $p\times n$ 矩阵, d 为 p 维向量
- 目标函数为 $\|Ax-b\|^2$
- 等式约束为 $Cx=d$
- 如果 x 满足 $Cx=d$, 则 x 是可行解
- 若 $\hat{x}$ 满足 $C\hat{x}=d$, 并且对于任意满足 $Cx=d$ 的 n 维向量 x, 有 $\|A\hat{x}-b\|^2\leq\|Ax-b\|^2$, 则 $\hat{x}$ 为问题的解。

通过微积分求最优条件

求解约束优化问题:

$$
\begin{array}{ll}
\text{minimize} & f(x) = \|Ax - b\|^2 \\
\text{subject to} & c_i^{\mathrm{T}} x = d_i, \quad i = 1, \ldots, p
\end{array}
$$

1. 引入拉格朗日乘子 $z_1, \ldots, z_p$,得到拉格朗日函数:

$$
L(x, z) = f(x) + z_1(c_1^{\mathrm{T}} x - d_1) + \ldots + z_p(c_p^{\mathrm{T}} x - d_p)
$$

2. 对 x 和 z 分别求偏导数,并令其等于零:

$$
\frac{\partial L}{\partial x_i}(\hat{x}, z) = 0, \quad i = 1, \ldots, n; \qquad \frac{\partial L}{\partial z_i}(\hat{x}, z) = 0, \quad i = 1, \ldots, p
$$

通过微积分求最优条件

- 对于损失函数 L，我们已知 $\frac{\partial L}{\partial z_i}(\hat{x}, z) = c_i^{\mathrm{T}}\hat{x} - d_i = 0$
- 前 n 个方程更为重要：

$$\frac{\partial L}{\partial x_i}(\hat{x}, z) = 2\sum_{j=1}^{n}\left(A^{\mathrm{T}}A\right)_{ij}\hat{x}_j - 2\left(A^{\mathrm{T}}b\right)_i + \sum_{j=1}^{p} z_j c_i = 0$$

- 以矩阵-向量形式表示：$2\left(A^{\mathrm{T}}A\right)\hat{x} - 2A^{\mathrm{T}}b + C^{\mathrm{T}}z = 0$
- 将其与 $C\hat{x} = d$ 相结合，得到 KKT 条件：

$$\begin{bmatrix} 2A^{\mathrm{T}}A & C^{\mathrm{T}} \\ C & 0 \end{bmatrix}\begin{bmatrix} \hat{x} \\ z \end{bmatrix} = \begin{bmatrix} 2A^{\mathrm{T}}b \\ d \end{bmatrix}$$

  这是一个包含 $n+p$ 个变量 $\hat{x}, z$ 的方程组
- KKT 方程是无约束最小二乘问题正规方程的扩展

约束最小二乘问题的解法

- 假设 KKT 矩阵是可逆的，我们有：

$$\begin{bmatrix} \hat{x} \\ z \end{bmatrix} = \begin{bmatrix} 2A^{\mathrm{T}}A & C^{\mathrm{T}} \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2A^{\mathrm{T}}b \\ d \end{bmatrix}$$

- 当且仅当 C 的行线性无关且 $\begin{bmatrix} A \\ C \end{bmatrix}$ 的列线性无关时，KKT 矩阵是可逆的
- 这意味着 $m+p \geq n$，且 $p \leq n$

通过 QR 分解求解约束最小二乘问题

Algorithm Constrained least squares via QR factorization

given an $m \times n$ matrix A and a $p \times n$ matrix C that satisfying the invertibility condition of the KKT matrix, an m-vector b, and a p-vector d.

1. QR factorizations. Compute the QR factorizations

$$\begin{bmatrix} A \\ C \end{bmatrix} = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} R, \quad Q_2^{\mathrm{T}} = \tilde{Q}\tilde{R}$$

2. Compute $\tilde{R}^{-T}d$ by forward substitution.
3. Form right-hand side and solve

$$\tilde{R}w = 2\tilde{Q}^{\mathrm{T}} Q_1^{\mathrm{T}} b - 2\tilde{R}^{-T}d$$

   via back substitution.
4. Compute $\hat{x}$. Form right-hand side and solve

$$R\hat{x} = Q_1^{\mathrm{T}} b - (1/2) Q_2^{\mathrm{T}} w$$

   by back substitution.

非线性最小二乘

非线性方程组

- n 个未知数 $x_1, \ldots, x_n$ 的 m 个非线性方程组：

$$f_i(x_1, \ldots, x_n) = 0, \quad i = 1, \ldots, m$$

- $f_i(x) = 0$ 是第 i 个方程；$f_i(x)$ 是第 i 个残差
- $x = (x_1, \ldots, x_n)$ 是 n 维向量
- 可以写成向量方程 $f(x) = 0$，其中 $f : \mathbf{R}^n \to \mathbf{R}^m$，

$$f(x) = (f_1(x), \ldots, f_m(x))$$

- 当 f 是仿射时，简化为 m 个线性方程组
- 若 $m > n$，方程组超定；若 $m < n$，方程组欠定；若 $m = n$，方程组"方阵"（square）

非线性最小二乘

- 寻找 $\hat{x}$ 满足：

$$\text{minimize} \quad \|f(x)\|^2 = f_1(x)^2 + \cdots + f_m(x)^2$$

- 包含解决方程 $f(x) = 0$ 的问题作为特例
- 类似（线性）最小二乘，本身非常有用

最优性条件

- 最优性的必要条件:$\nabla \| f(\hat{x}) \|^2 = 0$
- 任何最优点都满足这个条件
- 有些点可能满足这个条件但不是最优点
- 可以表示为 $2Df(\hat{x})^{\mathrm{T}} f(\hat{x}) = 0$
- $Df(\hat{x})$ 是 $m \times n$ 的导数或雅可比矩阵,

$$Df(\hat{x})_{ij} = \frac{\partial f_i}{\partial x_j}(\hat{x}), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n$$

- 当 f 是仿射时,最优性条件简化为正规方程

解决非线性最小二乘问题的困难性

- 解决非线性方程或非线性最小二乘问题（一般而言）要比解线性方程困难得多
- 甚至确定是否存在解都很困难
- 因此，实际中经常使用启发式算法（heuristics）：
  - 不能保证始终有效
  - 但在实践中通常表现良好

  （例如 k-means）

基本思路

- 在任意点 z 处，我们可以形成仿射近似

$$\hat{f}(x;z) \approx f(z) + Df(z)(x-z)$$

- 只要 x 接近 z，$\hat{f}(x;z) \approx f(x)$
- 我们可以通过线性最小二乘最小化 $\|\hat{f}(x;z)\|^2$
- 我们将迭代，其中 z 是当前迭代的点

基本高斯-牛顿算法

- 迭代 $x^{(1)}, x^{(2)}, \ldots$
- 在第 k 次迭代时，形成 f 在 $x^{(k)}$ 处的仿射近似：

$$\hat{f}(x; x^{(k)}) = f(x^{(k)}) + Df(x^{(k)})(x - x^{(k)})$$

- 将 $x^{(k+1)}$ 设置为 $\|\hat{f}(x; x^{(k)})\|^2$ 的最小化器，

$$x^{(k+1)} = x^{(k)} - \left(Df(x^{(k)})^{\mathrm{T}} Df(x^{(k)})\right)^{-1} Df(x^{(k)})^{\mathrm{T}} f(x^{(k)})$$

牛顿法（$m = n$）

- 迭代 $x^{(1)}, x^{(2)}, \ldots$
- 在第 k 次迭代时，形成 f 在 $x^{(k)}$ 处的仿射近似：

$$\hat{f}(x; x^{(k)}) = f(x^{(k)}) + Df(x^{(k)})(x - x^{(k)})$$

- 将 $x^{(k+1)}$ 设置为 $\hat{f}(x; x^{(k)}) = 0$ 的解

$$x^{(k+1)} = x^{(k)} - \left(Df(x^{(k)})\right)^{-1} f(x^{(k)})$$

Levenberg-Marquardt 算法

- 迭代 $x^{(1)}, x^{(2)}, \ldots$
- 在第 k 次迭代时，形成 f 在 $x^{(k)}$ 处的仿射近似：

$$\hat{f}(x; x^{(k)}) = f(x^{(k)}) + Df(x^{(k)})(x - x^{(k)})$$

- 选择 $x^{(k+1)}$ 为

$$\|\hat{f}(x; x^{(k)})\|^2 + \lambda^{(k)}\|x - x^{(k)}\|^2$$

  的最小化器，其中

$$\lambda^{(k)} > 0$$

- 我们希望 $\|\hat{f}(x; x^{(k)})\|^2$ 很小，但我们不希望 $x^{(k+1)}$ 离 $x^{(k)}$ 太远，因为这时 $\hat{f}(x; x^{(k)}) \approx f(x)$ 不再成立。

调整 $\lambda^{(k)}$

思路：

- 如果 $\lambda^{(k)}$ 太大，$x^{(k+1)}$ 太接近 $x^{(k)}$，进展缓慢
- 如果太小，$x^{(k+1)}$ 可能离 $x^{(k)}$ 很远，仿射近似很差

更新机制：

- 如果 $\|f(x^{(k+1)})\|^2 < \|f(x^{(k)})\|^2$，接受新的 x 并减小 λ：

$$\lambda^{(k+1)} = 0.8\lambda^{(k)}$$

- 否则，增加 λ 并不更新 x：

$$\lambda^{(k+1)} = 2\lambda^{(k)}, \quad x^{(k+1)} = x^{(k)}$$

矩阵的广义逆

- 问题：给定 $A \in \mathbf{C}^{n\times n}$，若 A 可逆，则线性方程组 $Ax = b$ 存在唯一解 $x = A^{-1}b$。那么
  - 1. 若 A 不是方阵，即 $A \in \mathbf{C}^{m\times n}$，且线性方程组 $Ax = b$ 的解若存在，可否用 $x = Hb$ 的形式表示出来？
  - 2. 若 A 不是方阵且线性方程组 $Ax = b$ 的解不存在，如何求其最好的近似解？
- 定义：考虑线性方程组

$$Ax = b$$

其中，$A \in \mathbf{C}^{m\times n}$，$b \in \mathbf{C}^{m}$。对于任意的 $b \in R(A)$，存在 $A^{-} \in \mathbf{C}^{n\times m}$ 使得 $x = A^{-}b$ 为线性方程组的解，则称 A^{-} 为矩阵 A 的广义逆。

广义逆的存在性与表示

矩阵的广义逆

- 定理：给定 $A \in \mathbf{C}^{m\times n}$，广义逆存在的充要条件是存在 $A^{-} \in \mathbf{C}^{n\times m}$ 满足
$$AA^{-}A = A$$
即上述矩阵方程有解。
- 定理：给定 $A \in \mathbf{C}^{m\times n}$，假设其秩为 r，$P \in \mathbf{C}^{m\times m}$ 和 $Q \in \mathbf{C}^{n\times n}$ 为可逆矩阵使得
$$PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$
则
$$A^{-} = Q\begin{bmatrix} I_r & X \\ Y & Z \end{bmatrix}P$$
为矩阵 A 的广义逆，其中 X, Y, Z 为任意维数相容的矩阵。

矩阵的伪逆

- 定义：考虑 $A \in \mathbf{C}^{m\times n}$，若存在 $A^+ \in \mathbf{C}^{n\times m}$ 满足以下等式

$$
\begin{aligned}
AA^+A &= A \\
A^+AA^+ &= A^+ \\
(AA^+)^{\mathrm{H}} &= AA^+ \\
(A^+A)^{\mathrm{H}} &= A^+A
\end{aligned}
$$

  则称 A^+ 是矩阵 A 的伪逆，也称 Moore-Penrose pseudoinverse。

- 定理：考虑 $A \in \mathbf{C}^{m\times n}$，则其伪逆存在且唯一，其可以表示为

$$A^+ = C^{\mathrm{H}}(CC^{\mathrm{H}})^{-1}(B^{\mathrm{H}}B)^{-1}B^{\mathrm{H}}$$

  其中，$A = BC$ 为 A 的一个满秩分解。

线性方程组的最小二乘解

- 定义：给定 $A \in \mathbf{C}^{m\times n}$，假设线性方程组 $Ax = b$ 无解。若 x^* 满足

$$x^* = \min_{x\in\mathbf{C}^n} \|Ax - b\|_2$$

则称 x^* 是线性方程组的最小二乘解；长度最小的最小二乘解称为极小最小二乘解。

- 定理：定义：给定 $A \in \mathbf{C}^{m\times n}$，假设线性方程组 $Ax = b$ 无解。则
  - 1. 其最小二乘解的通解为

$$x^* = A^+ b + (I - A^+ A) y$$

  其中 $y \in \mathbf{C}^n$ 为任意向量
  - 2. 其唯一的极小最小二乘解为

$$\tilde{x}^* = A^+ b$$

4.6 正规矩阵

- 回忆：若当标准型
- 问题：如果只允许用标准正交基，可将一个矩阵相似变换成什么？
- 定理：任一复方阵均酉相似于上三角矩阵。更具体地讲，给定 $A \in \mathbf{C}^{n \times n}$，一定有酉矩阵 $U \in \mathbf{C}^{n \times n}$ 使得

$$U^{-1}AU$$

为上三角矩阵（以上即 Schur 定理）。

- 问题：给定任一复方阵，什么时候存在酉矩阵将其化为对角阵（几何视角如何解读）？
- 定义：若 $A \in \mathbf{C}^{n\times n}$ 满足

$$A^{\mathrm{H}}A = AA^{\mathrm{H}} \tag{4}$$

即一个给定复方阵，若其共轭转置与其可交换，则称其为正规矩阵。

- 引理：若 A 同时是正规矩阵和上三角矩阵，则其是对角阵。
- 定理：给定 $A \in \mathbf{C}^{n\times n}$，存在酉矩阵 $U \in \mathbf{C}^{n\times n}$ 使得 $U^{-1}AU$ 为对角阵的充要条件是其为正规矩阵（几何视角如何解读）。

三种特殊正规矩阵的特征值

- 正规矩阵的例子
  - Hermite 矩阵
  - 反 Hermite 矩阵
  - 酉矩阵
- 给定正规矩阵 $A \in \mathbf{C}^{n\times n}$，则
  - A 是 Hermite 矩阵的充要条件是其特征值是实数
  - A 是反 Hermite 矩阵的充要条件是其特征值的实部为零
  - A 是酉矩阵的充要条件是其特征值的模长等于 1

正规矩阵的酉相似对角化的具体过程

首先，给定 n 阶正规矩阵 A，由前述定理知

- （a） A 的特征值的代数重数等于其几何重数
- （b） A 的对应于不同特征值的特征子空间相互正交

于是，将 A 酉相似对角化的具体过程是

- （1） 计算 $|\lambda I_n - A|$ 的所有不同的根，记为 $\lambda_1, \lambda_2, \cdots, \lambda_p$
- （2） 对每个 λ_i，计算 $(\lambda_i I_n - A)x = 0$ 的基础解系，记为

$$v_1^i, v_2^i, \cdots, v_{r_i}^i$$

- （3） 通过 Gram—Schmidt 标准正交化，求得

$$u_1^i, u_2^i, \cdots, u_{r_i}^i$$

- （4） 拼出 Λ

$$\Lambda = diag\{\lambda_i I_{r_i}\}$$

其中 $i = 1, 2, \cdots, p$；拼出 U 如下

$$U = \left[\begin{array}{ccccccccc} u_1^1 & \cdots & u_{r_1}^1 & u_1^2 & \cdots & u_{r_2}^2 & \cdots & u_1^p & \cdots & u_{r_p}^p \end{array}\right]$$

4.7 Hermite 矩阵与 Hermite 二次型

- 定理：给定 $A \in \mathbf{C}^{n\times n}$，则下列条件等价
  - (1) A 为 Hermite 矩阵，即
$$A = A^{\mathrm{H}}$$
  - (2) 对任意的 $x, y \in \mathbf{C}^n$，都有
$$\langle Ax, y\rangle = \langle x, Ay\rangle$$
- 定义：给定 n 阶 Hermite 矩阵 A。映射
$$x \mapsto \langle Ax, x\rangle = x^{\mathrm{H}}Ax$$
  确定了从 $\mathbf{C}^n$ 到 $\mathbf{R}$ 的函数，称为 Hermite 二次型，写成分量形式有
$$x^{\mathrm{H}}Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\bar{x}_i x_j$$
- 为什么 Hermite 二次型一定为实数？
  - (a) A 为正规矩阵
  - (b) A 的特征值均为实数

- 定义：给定 n 阶 Hermite 矩阵 A。若对任意的 $x \in \mathbf{C}^n$ 且 $x \neq 0$，都有

$$x^{\mathrm{H}} Ax > 0 \ (\geq 0)$$

则称 $x^{\mathrm{H}} Ax$ 为 Hermite 正定（半正定）二次型，称 A 为 Hermite 正定（半正定）矩阵。

- 定理：给定 n 阶 Hermite 矩阵 A。其为 Hermite 正定（半正定）矩阵的充要条件是其特征值均为正实数（非负实数）。
- 定理：给定 n 阶 Hermite 正定（半正定）矩阵 A。则存在唯一的 Hermite 正定（半正定）矩阵 H 使得

$$A = B^2$$

Hermite 矩阵特征值的极值刻画

- 定理：给定 n 阶 Hermite 矩阵 A，其特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 均为实数，不妨假设
$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$$
则有
$$\lambda_{\min}(A) = \lambda_1 = \min_{x\in\mathbf{C}^n,\ x\neq 0} \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x} = \min_{x\in\mathbf{C}^n,\ \|x\|=1} x^{\mathrm{H}}Ax$$
$$\lambda_{\max}(A) = \lambda_n = \max_{x\in\mathbf{C}^n,\ x\neq 0} \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x} = \max_{x\in\mathbf{C}^n,\ \|x\|=1} x^{\mathrm{H}}Ax$$
- 注：Hermite 矩阵的 Rayleigh 商
$$R(x) := \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x},\ \ x \in \mathbf{C}^n,\ \ x \neq 0$$

4.8 矩阵的奇异值分解

- 回忆：两个矩阵 $A, B \in \mathbf{F}^{m\times n}$ 被称为是等价的，记为 $A \sim B$，如果存在 n 阶非奇异矩阵 $P \in \mathbf{F}^{n\times n}$ 和 m 阶非奇异矩阵 $Q \in \mathbf{F}^{m\times m}$ 使得

$$AP = QB$$

- 初等行列变换的标准型：对任意矩阵 $A \in \mathbf{F}^{m\times n}$，可以经过初等行列变换将 A 化为标准型，即存在 n 阶非奇异矩阵 $P \in \mathbf{F}^{n\times n}$ 和 m 阶非奇异矩阵 $Q \in \mathbf{F}^{m\times m}$ 使得

$$Q^{-1}AP = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$
$$\leftrightarrow AP = Q\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

- 问题：若 P 和 Q 都是酉矩阵，那么 A 可以被变换成什么样子？

- 引理：给定矩阵 $X \in \mathbf{C}^{m\times n}$ 且 $rank(X) = r$，则有
  - (a) $X^{\mathrm{H}}X$ 和 XX^{H} 分别是 m 阶和 n 阶的 Hermite 半正定矩阵
  - (b) $rank(X^{\mathrm{H}}X) = rank(XX^{\mathrm{H}}) = r$
- 定理：给定矩阵 $X \in \mathbf{C}^{m\times n}$ 且 $rank(X) = r$，则存在酉矩阵 $U \in \mathbf{C}^{m\times m}$ 和 $V \in \mathbf{C}^{n\times n}$ 使得

$$U^{\mathrm{H}}XV = \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$$

其中，

$$\Sigma_r = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$$

且 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$。特别的，我们称

$$\sigma_i = \sqrt{\lambda_i(X^{\mathrm{H}}X)},\ i = 1, 2, \cdots, r$$

为矩阵 X 的奇异值。