第一章 内积空间

1.1 内积与 Gram 矩阵

定义 1.1.1 (欧几里德空间)
设V是实数域$\mathbb{R}$上的线性空间. 映射
$\tau: V\times V\to\mathbb{R}$
称为V上的一个内积, 并记$\tau(v_1,v_2)=\langle v_1,v_2\rangle$, 如果它满足:
1) 对称性: 对任意$v_1,v_2\in V$, $\langle v_2,v_1\rangle=\langle v_1,v_2\rangle$;
2) 对第二个变元的线性性: 对任意$v_1,v_2,v_3\in V$, $k,l\in\mathbb{R}$, $\langle v_1,v_2k+v_3l\rangle=\langle v_1,v_2\rangle k+\langle v_1,v_3\rangle l$;
3) 正定性: 对任意$v\in V, v\neq 0$, $\langle v,v\rangle>0$.
若在V上指定了一个内积, 则称V关于该内积为一个实内积空间. 有限维的实内积空间也称为欧几里德空间, 简称为欧氏空间.

注 1.1.1 (双线性)
首先注意下面事实: 固定v_1, 让v变化, 则由内积可决定一个映射如下
$\langle v_1,\cdot\rangle: V\to\mathbb{R}$
$\quad v\mapsto\langle v_1,v\rangle.$
定义中的性质 2)就是说该映射$\langle v_1,\cdot\rangle$是从V到$\mathbb{R}$的线性映射, 因此称之为对第二个变元是线性的. 由性质 1), 2)可推出内积对第一个变元也是线性的.

注 1.1.2 ($\mathbb{R}$上的标准欧几里德空间$\mathbb{R}^n$)
首先, $\mathbb{R}^n$是$\mathbb{R}$上的标准线性空间. 定义

$$\begin{aligned}\langle x,y\rangle &= x^{\mathrm{T}}y\\ &=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix}\\ &= x_1y_1+x_2y_2+\cdots+x_ny_n.\end{aligned}$$

容易验证, 这样定义的$\langle\cdot,\cdot\rangle$是$\mathbb{R}^n$上的内积, 称为$\mathbb{R}^n$上的标准内积, 相应地, $\mathbb{R}^n$称为标准欧几里德空间.

注 1.1.3 (几何空间作为欧几里德空间)
回忆第一章的例??, 考虑几何空间. 集合

$$\begin{aligned}V&=\{\text{有向线段}\}\\ &=\left\{\overrightarrow{AB}: A,B\text{取遍几何空间}\right\},\end{aligned}$$

关于平行四边形法则或三角形法则定义加法, 同向或反向伸缩定义数乘法, 构成$\mathbb{R}$上的线性空间. 对任意$\overrightarrow{OA},\overrightarrow{OB}\in V$定义
$\langle\overrightarrow{OA},\overrightarrow{OB}\rangle=|\overrightarrow{OA}|\cdot|\overrightarrow{OB}|\cdot\cos\angle AOB$,
称为向量$\overrightarrow{OA},\overrightarrow{OB}$的点乘, 或数量积. 这里, 符号$|\cdot|$表示有向线段的长度. 可以验证, $\langle\cdot,\cdot\rangle$是$V$上的内积.
只验证性质 2). 验证过程并非显然(其实, 在中学数学课本或大学高等数学中向量代数与空间解析几何章节中就有). 可基于实数$|\overrightarrow{OB}|\cdot\cos\angle AOB$的如下的几何意义: 以$O$为原点, 以$\overrightarrow{OA}$方向为正方向, 使得$\overrightarrow{OA}$所在的直线成为一数轴, 则$|\overrightarrow{OB}|\cdot\cos\angle AOB$就是有向线段$\overrightarrow{OB}$在$\overrightarrow{OA}$方向上的投影矢量(有向线段)在该数轴上的坐标.

注 1.1.4 (线性空间$\mathbb{R}^2$上的非标准内积的例)
定义

$\langle x,y\rangle = x_1y_1+\frac{1}{2}x_1y_2+\frac{1}{2}x_2y_1+x_2y_2$.

验证它是内积:

1) 对称性:

2) x 固定, 作为 y 的函数是线性的:

$$\begin{aligned}\langle x,y\rangle &= x_1y_1+\frac{1}{2}x_1y_2+\frac{1}{2}x_2y_1+x_2y_2\\ &=\left(x_1+\frac{1}{2}x_2\right)y_1+\left(\frac{1}{2}x_1+x_2\right)y_2\end{aligned}$$

3) 正定性:

$$\begin{aligned}\langle x,x\rangle &= x_1x_1+x_1x_2+x_2x_2\\ &=(x_1+\tfrac{1}{2}x_2)^2+\tfrac{3}{4}(x_2)^2.\end{aligned}$$

若 $x\neq 0$, 则 $x_1+\frac{1}{2}x_2$ 和 x_2 两数不全为零, 从而 $\langle x,x\rangle>0$.

注 1.1.5 (平方可积函数空间 $\mathcal{L}^2([a,b],\mathbb{R}^n)$)

$\mathcal{L}^2([a,b],\mathbb{R}^n)$ 是函数空间 $\mathcal{F}([a,b],\mathbb{R}^n)$ 的子集合, 定义为

$$\mathcal{L}^2([a,b],\mathbb{R}^n)=\{f\mid f\in\mathcal{F}([a,b],\mathbb{R}^n),\text{且}\int_a^b (f(t))^{\mathrm{T}}f(t)\mathrm{d}t<+\infty\},$$

即以定义在区间 $[a,b]$ 上取之于 $\mathbb{R}^n$ 的平方可积函数为元素的集合. 记

$$f(t)=[f_1(t),\cdots,f_n(t)]^{\mathrm{T}},$$

则

$$\int_a^b (f(t))^{\mathrm{T}}f(t)\mathrm{d}t=\int_a^b[(f_1(t))^2+\cdots+(f_n(t))^2]\mathrm{d}t.$$

首先证明, 它是 $\mathcal{F}([a,b],\mathbb{R}^n)$ 一个线性子空间(不等式). 它不是有限维的.

其次, 对任意 $f,g\in\mathcal{L}^2([a,b],\mathbb{R}^n)$, 定义

$$\langle f,g\rangle=\int_a^b (f(t))^{\mathrm{T}}g(t)\mathrm{d}t=\int_a^b[f_1(t)g_1(t)+\cdots+f_n(t)g_n(t)]\mathrm{d}t.$$

可以证明, 这是线性空间 $\mathcal{L}^2([a,b],\mathbb{R}^n)$ 上的一个内积.

注 1.1.6 (概率论中的内积空间 $\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$)

$\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$ 是概率空间 $(\Omega,\mathcal{B},P(\cdot))$ 上的有二阶矩的随机变量的集合. 即 $X\in\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$ 的含义是: 首先, $X:\Omega\to\mathbb{R}$ 是随机变量; 其次, 它有有限的二阶矩, 即

$$E(X^2)=\int_\Omega (X(\omega))^2P(d\omega)<+\infty.$$

由概率论知:

(1) 若 $X\in\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$, 则 X 的数学期望 $E(X)$ 和方差 $D(X)$ 都存在;

(2) 若 $X,Y\in\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$, 则 XY 的数学期望 $E(XY)$ 存在.

首先, 可以证明 $\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$ 是线性空间; 其次, 对任意 $X,Y\in\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$, 定义

$$\langle X,Y\rangle=E(XY),$$

可以证明, 这是线性空间 $\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$ 上的一个内积.

回忆 X 与 Y 的协方差

$$\begin{aligned}\mathrm{cov}(X,Y)&=E\{(X-E(X))(Y-E(Y))\}\\ &=E(XY)-E(X)E(Y)\end{aligned}$$

可知, 它与内积的关系. 若只考虑零均值随机变量, 则二者相同.

定义 1.1.2 (酉空间)
设V是复数域$\mathbb{C}$上的线性空间. 映射
$\tau: V\times V\to\mathbb{C}$
称为V上的一个内积, 并记$\tau(v_1,v_2)=\langle v_1,v_2\rangle$, 如果它满足:
1) 共轭对称性: 对任意$v_1,v_2\in V$, $\overline{\langle v_2,v_1\rangle}=\langle v_1,v_2\rangle$;
2) 对第二个变元的线性性: 对任意$v_1,v_2,v_3\in V$, $k,l\in\mathbb{R}$, $\langle v_1, v_2k+v_3l\rangle=\langle v_1,v_2\rangle k+\langle v_1,v_3\rangle l$;
3) 正定性: 对任意$v\in V, v\neq 0$, $\langle v,v\rangle$为实数, 且$\langle v,v\rangle>0$.
定义了内积的复线性空间称为复内积空间. 有限维的复内积空间也称为酉空间.

命题 1.1.1 (非负性)
对任意$v\in V$, $\langle v,v\rangle$为实数, 且$\langle v,v\rangle\geq 0$. 等号成立, 即$\langle v,v\rangle=0$, 当且仅当$v=0$.
证明: 由性质 3), 只需证明$\langle 0,0\rangle=0$. 而由性质 2)有
$\langle 0,0\rangle=\langle 0,0+0\rangle=\langle 0,0\rangle+\langle 0,0\rangle$,
因此, 作为一个复数, 必有$\langle 0,0\rangle=0$. 证毕

命题 1.1.2 (复内积对第一个变元是共轭线性的)
确切地, 对任意$v_1,v_2,v_3\in V$, $k,l\in\mathbb{R}$, $\langle v_1k+v_2l, v_3\rangle=\bar{k}\langle v_1,v_3\rangle+\bar{l}\langle v_2,v_3\rangle$.
证明:
$$\begin{aligned}\langle v_1k+v_2l, v_3\rangle&=\overline{\langle v_3, v_1k+v_2l\rangle}\quad(\text{由性质1)})\\&=\overline{\langle v_3, v_1\rangle k+\langle v_3,v_2\rangle l}\quad(\text{由性质2)})\\&=\bar{k}\cdot\overline{\langle v_3, v_1\rangle}+\bar{l}\cdot\overline{\langle v_3,v_2\rangle}\\&=\bar{k}\cdot\langle v_1,v_3\rangle+\bar{l}\cdot\langle v_2,v_3\rangle\quad(\text{由性质1)}).\end{aligned}$$
证毕

命题 1.1.3 (零向量与其他向量的内积)
对任意$v\in V$, $\langle v, 0\rangle=\langle 0, v\rangle=0$.

注 1.1.7 ($\mathbb{C}$上的标准酉空间$\mathbb{C}^n$)
首先, $\mathbb{C}^n$是$\mathbb{C}$上的标准线性空间. 定义
$$\begin{aligned}\langle x,y\rangle&=\bar{x}^{\mathrm{T}}y\\&=\begin{bmatrix}\bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n\end{bmatrix}\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}\\&=\bar{x}_1y_1+\bar{x}_2y_2+\cdots+\bar{x}_ny_n.\end{aligned}$$
容易验证, 这样定义的$\langle\cdot,\cdot\rangle$是$\mathbb{C}^n$上的内积, 称为$\mathbb{C}^n$上的标准内积, 相应地, $\mathbb{C}^n$称为标准酉空间.

定义 1.1.3 (复矩阵的共轭转置)
设$A\in\mathbb{C}^{m\times n}$. 定义记号
$A^{\mathrm{H}}:=\left(\bar{A}\right)^{\mathrm{T}}=\left[b_{ij}\right]_{n\times m}$,
这里, $b_{ij}=\bar{a}_{ji}$表示复数a_{ij}的共轭.

于是, $\mathbb{C}^n$上的标准内积可写成矩阵乘法
$\langle x,y\rangle=x^{\mathrm{H}}y$.

注 1.1.8 (重要注记: 抽象向量内积的抽象矩阵运算表达)
将两个向量的内积运算写成两个抽象1×1矩阵之间的形式乘法运算如下

$\langle\alpha,\beta\rangle=\alpha^{\mathrm{H}}\beta$.

这种记号是基于将抽象向量类比成列数组的想法. (若是类比成行数组, 则有一套对偶的记号.)

以下我们只对复内积空间来叙述.

命题 1.1.4 (线性组合的内积的矩阵表示)

$$\left\langle\sum_{i=1}^{s}\alpha_i k_i,\ \ \sum_{j=1}^{t}\beta_j l_j\right\rangle=\begin{bmatrix}\bar{k}_1 & \cdots & \bar{k}_s\end{bmatrix}\left[\langle\alpha_i,\beta_i\rangle\right]_{s\times t}\begin{bmatrix}l_1\\ \vdots\\ l_t\end{bmatrix}$$

$$=\overline{\begin{bmatrix}k_1\\ \vdots\\ k_s\end{bmatrix}}^{\mathrm{T}}\left[\langle\alpha_i,\beta_i\rangle\right]_{s\times t}\begin{bmatrix}l_1\\ \vdots\\ l_t\end{bmatrix}.$$

证明: 反复利用定义 1.1.2 中的性质 2)和命题 1.1.2 可得. 证毕

因此, 只要知道两向量组之间, 诸对向量的内积, 那么, 由这两向量组各自任意作线性组合做出的一对向量, 其内积可由上述诸对向量的内积及线性组合的系数来计算. 两向量组之间所有配对的内积拼成矩阵, 命名为:

定义 1.1.4 (两向量组的交互 Gram 矩阵)

设 $\alpha_1,\alpha_2,\ldots,\alpha_s$ 和 $\beta_1,\beta_2,\ldots,\beta_t$ 是复内积空间的两个向量组. 矩阵

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

称为 $\alpha_1,\alpha_2,\ldots,\alpha_s$ 和 $\beta_1,\beta_2,\ldots,\beta_t$ 的交互 Gram 矩阵. 简记为 $G(\{\alpha_i\};\{\beta_j\})$.

注 1.1.9 (标准欧几里德空间 $\mathbb{R}^n$ 及标准酉空间 $\mathbb{C}^n$ 上的两向量组的交互 Gram 矩阵; 矩阵乘积作为两向量组的交互 Gram 矩阵)

作为标准酉空间 $\mathbb{C}^n$ 上向量组, 矩阵 $A\in\mathbb{C}^{n\times s}$ 的列向量组(简称为列向量组 A)和 $B\in\mathbb{C}^{n\times t}$ 的列向量组(简称为列向量组 B)之间的交互 Gram 矩阵为

$G=A^{\mathrm{H}}B$.

反之, 给定任何矩阵乘积关系

$C=AB$,

注意到

$C=AB=\left(A^{\mathrm{H}}\right)^{\mathrm{H}}B$

矩阵 C 均可解释为列向量组 A^{H} 和列向量组 B 之间的交互 Gram 矩阵.

注 1.1.10 (矩阵乘法的三种观点)

观察矩阵乘法关系

$C=AB$

的三种着眼点(观点者, 观察事物的着眼点是也):

- 观点 1: 将 C 按列分块. C 的每一列都是列向量组 A 的线性组合; 或者反过来说将 C 的每一列沿列向量组 A 分解. 这也就是, 视矩阵 A 为列空间上的线性映射, B 和 C 的相应列是原像与映像的关系.
- 观点 2: 将 C 按行分块. C 的每一行都是行向量组 B 的线性组合; 或者反过来说将 C 的每一行沿行向量组 B 分解. 这也就是, 视矩阵 B 为行空间上的线性映射, A 和 C 的相应行是原像与映像的关系.
- 观点 3: 将 C 按逐行逐列分块. C 的第 i 行第 j 列元素 c_{ij} 是 A^{H} 的第 i 列与 B 的第 j 列的内积(列的观点), 或者说是 A 的第 i 行与 B^{H} 的第 j 行的内积(行的观点). 从而 C 就是交互 Gram 矩阵(可有行和列两种说法).

定义 1.1.5 (向量组的 Gram 矩阵, 内积空间的度量矩阵)

设 $\alpha_1,\alpha_2,\ldots,\alpha_s$ 是复内积空间的一个向量组. 矩阵

$G(\alpha_1,\alpha_2,\ldots,\alpha_s):=\left[\langle\alpha_i,\alpha_j\rangle\right]_{s\times s}\in\mathbb{C}^{s\times s}$
称为向量组$\alpha_1,\alpha_2,\ldots,\alpha_s$的 Gram 矩阵. 简记为$G(\{\alpha_i\})$. 基向量组的 Gram 矩阵称为该基的度量矩阵.

注 1.1.11 (标准欧几里德空间$\mathbb{R}^n$及标准酉空间$\mathbb{C}^n$上的向量组的 Gram 矩阵)

注 1.1.12 (重要注记: 抽象向量内积符号的抽象矩阵运算表达)
$G(\{\alpha_i\};\{\beta_j\})=\begin{bmatrix}\alpha_1 & \cdots & \alpha_s\end{bmatrix}^{\mathrm{H}}\begin{bmatrix}\beta_1 & \cdots & \beta_t\end{bmatrix}$.
行体制时对偶.

注 1.1.13 ($\mathcal{L}^2([a,b],\mathbb{R}^n)$和$\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$中的 Gram 矩阵的具体形式)

定理 1.1.1 (Gram 矩阵将抽象线性方程组转化为具体的线性方程组)
设$\alpha_1,\alpha_2,\ldots,\alpha_t$是$\mathbb{F}$上线性空间$V$中的一个向量组, $\beta\in V$. 则对任意$x\in\mathbb{F}^t$,
$\begin{bmatrix}\alpha_1 & \cdots & \alpha_t\end{bmatrix}x=\beta$
等价于
$G(\{\alpha_i\})x=G(\{\alpha_i\},\beta)$.
注意到$G(\{\alpha_i\})\in\mathbb{F}^{t\times t}$, $G(\{\alpha_i\},\beta)\in\mathbb{F}^t$是两个“具体矩阵”.

注 1.1.14 (内积提供了一种将抽象空间中的抽象事物映射成具体空间中具体事物的手段)

推论 1.1.1 (抽象齐次线性方程组解空间维数)
设$\alpha_1,\alpha_2,\ldots,\alpha_t$是$\mathbb{F}$上线性空间$V$中的一个向量组. 则抽象齐次线性方程组
$\begin{bmatrix}\alpha_1 & \cdots & \alpha_t\end{bmatrix}x=0$
的解集合
$S=\left\{x\in\mathbb{F}^t:\begin{bmatrix}\alpha_1 & \cdots & \alpha_t\end{bmatrix}x=0\right\}$
是标准线性空间$\mathbb{F}^t$的子空间, 且
$\dim S=t-\operatorname{rank}(\alpha_1,\ldots,\alpha_t)$.
证明: 记$\operatorname{rank}(\alpha_1,\ldots,\alpha_t)=r$, 且不妨设$\alpha_1,\ldots,\alpha_r$是$\alpha_1,\alpha_2,\ldots,\alpha_t$的一个极大线性无关组. 则有
$\begin{bmatrix}\alpha_1 & \cdots & \alpha_t\end{bmatrix}=\begin{bmatrix}\alpha_1 & \cdots & \alpha_r\end{bmatrix}\begin{bmatrix}I_r & A_2\end{bmatrix}$.
于是, 由$\alpha_1,\ldots,\alpha_r$线性无关,
$\begin{bmatrix}\alpha_1 & \cdots & \alpha_t\end{bmatrix}x=0 \Leftrightarrow \begin{bmatrix}\alpha_1 & \cdots & \alpha_r\end{bmatrix}\begin{bmatrix}I_r & A_2\end{bmatrix}x=0 \Leftrightarrow \begin{bmatrix}I_r & A_2\end{bmatrix}x=0$.
因此
$S=\left\{x\in\mathbb{F}^t:\begin{bmatrix}I_r & A_2\end{bmatrix}x=0\right\}$.
但显然有矩阵$\begin{bmatrix}I_r & A_2\end{bmatrix}$的秩$\operatorname{rank}\begin{bmatrix}I_r & A_2\end{bmatrix}=r$, 因此$S$是线性空间, 且
$\dim S=t-\operatorname{rank}\begin{bmatrix}I_r & A_2\end{bmatrix}=t-r$. 证毕

定理 1.1.2 (Gram 矩阵的性质)
设$\beta_1,\beta_2,\ldots,\beta_s$是复内积空间的一个向量组, $G=G(\beta_1,\ldots,\beta_s)$.
1) Hermite 对称性, 即: $\overline{G}^{\mathrm{T}}=G$.
2) Hermite 非负定性, 即: 对任意$z\in\mathbb{C}^s$, 有$\bar{z}^TGz\geq 0$.
3) Hermite 正定性等价于向量组的线性无关性, 即: $\beta_1,\beta_2,\ldots,\beta_s$线性无关, 当且仅当对任意$z\in\mathbb{C}^s, z\neq 0$, 有$\bar{z}^TGz>0$.
4) $\operatorname{rank}(\beta_1,\beta_2,\ldots,\beta_s)=\operatorname{rank}(G)$
证明: 1) 由
$(\overline{G}^{\mathrm{T}})_{ij}=(\overline{G})_{ji}=\overline{\langle\beta_j,\beta_i\rangle}=\langle\beta_i,\beta_j\rangle=(G)_{ij}$

即得.
2) 由命题 1.1.4 可得
$\bar{z}^T Gz = \langle [\beta_1 \ \cdots \ \beta_s] z, [\beta_1 \ \cdots \ \beta_s] z \rangle$,
再由命题 1.1.1, 即得 $\bar{z}^T Gz \geq 0$.
3) $\beta_1, \beta_2, \ldots, \beta_s$ 线性无关, 当且仅当对任意 $z \in \mathbb{C}^s, z \neq 0$,
$[\beta_1 \ \cdots \ \beta_s] z \neq 0$,
而有命题 1.1.1, $[\beta_1 \ \cdots \ \beta_s] z \neq 0$ 当且仅当
$\langle [\beta_1 \ \cdots \ \beta_s] z, [\beta_1 \ \cdots \ \beta_s] z \rangle = \bar{z}^T Gz > 0$. 证毕
4) 首先证明 $[\beta_1 \ \cdots \ \beta_s] z = 0$, 当且仅当 $Gz = 0$.
若 $[\beta_1 \ \cdots \ \beta_s] z = 0$, 则
$\langle \beta_i, [\beta_1 \ \cdots \ \beta_s] z \rangle = \sum_{j=1}^{s} \langle \beta_i, \beta_j \rangle z_j = [g_{i1} \ \cdots \ g_{is}] z = 0$, $i = 1, \ldots, s$.
从而 $Gz = 0$. 反之, 若 $Gz = 0$, 则有 $\bar{z}^T Gz = 0$. 而
$\bar{z}^T Gz = \langle [\beta_1 \ \cdots \ \beta_s] z, [\beta_1 \ \cdots \ \beta_s] z \rangle$,
从而由命题 1.1.1, 必有 $[\beta_1 \ \cdots \ \beta_s] z = 0$.
于是由普通线性方程组 $Gz = 0$ 与抽象线性方程组 $[\beta_1 \ \cdots \ \beta_s] z = 0$ 同解, 于是由推论 1.1.1 知
$s - \operatorname{rank} G = s - \operatorname{rank}(\beta_1, \beta_2, \ldots, \beta_s)$,
从而 $\operatorname{rank}(\beta_1, \beta_2, \ldots, \beta_s) = \operatorname{rank}(G)$. 证毕

定理 1.1.3 (抽象内积通过坐标系和度量矩阵表现为具体内积)
设 $\alpha_1, \alpha_2, \ldots, \alpha_n$ 是复内积空间 V 的一个基, $G = G(\alpha_1, \ldots, \alpha_n) \in \mathbb{C}^{n \times n}$ 为其度量矩阵. 若向量 $\alpha, \beta \in V$ 在该基下的坐标分别为 $x, y \in \mathbb{C}^n$, 则
$\langle \alpha, \beta \rangle = \bar{x}^{\mathrm{T}} Gy$.
特别地, 若 G 为单位矩阵 I_n, 则 V 上的内积通过坐标化为 $\mathbb{C}^n$ 上的标准内积.

注 1.1.15 (任意抽象内积都可以通过适当选择坐标系表现为 $\mathbb{C}^n$ 上的标准内积)
给定内积空间, 不同的基底一般会有不同的度量矩阵. 可以证明, 总可以选择一组基底, 使得其度量矩阵是单位矩阵. 这样的基底就是标准正交基. 标准正交基的存在性可由 Gram-Schmidt 正交化方法, 或者 Hermit 矩阵的性质来证明.

1.2 向量的长度和夹角

定义 1.2.1 (向量的长度)
设 α 是复内积空间的一个向量, 记
$\|\alpha\| := \sqrt{\langle \alpha, \alpha \rangle}$,
称为向量 α 的长度. 向量 α 和 β 之间的距离定义为
$d(\alpha, \beta) := \| \alpha - \beta \|$.

命题 1.2.1 (向量长度的性质)
1) 正性: 对任意向量 α, $\|\alpha\| \geq 0$. 等号成立, 当且仅当 $\alpha = 0$.
2) 正齐性: 对任意向量 α 及任意复数 k, $\|\alpha k\| = \|\alpha\| \cdot |k|$.
3) 三角不等式: 对任意向量 α, β, $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$.
4) Cauchy-Schwartz 不等式: 对任意向量 α, β, $|\langle \alpha, \beta \rangle| \leq \|\alpha\| \cdot \|\beta\|$.
5) 平行四边形公式: 对任意向量 α, β,

$\|\alpha+\beta\|^2+\|\alpha-\beta\|^2=2\|\alpha\|^2+2\|\beta\|^2$.

证明: 4) 若 $\langle\alpha,\beta\rangle=0$, 则结论显然成立. 故假定复数 $\langle\alpha,\beta\rangle\neq 0$. 此时必有 α,β 都不为零. 先考虑 $\langle\alpha,\beta\rangle$ 是非零实数的情形:

$|\langle\alpha,\beta\rangle|\le\|\alpha\|\cdot\|\beta\| \Leftrightarrow (2\langle\alpha,\beta\rangle)^2\le 4\|\alpha\|^2\cdot\|\beta\|^2$, 即 $(2\langle\alpha,\beta\rangle)^2\le 4\langle\alpha,\alpha\rangle\cdot\langle\beta,\beta\rangle$

$\Leftrightarrow$ 对任意实数 t, 实系数的二次函数 $f(t)=\langle\alpha,\alpha\rangle t^2+2\langle\alpha,\beta\rangle t+\langle\beta,\beta\rangle\ge 0$.

由于当 t, $\langle\alpha,\beta\rangle$ 均为实数时,

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

所以恒有 $f(t)\ge 0$.

再考虑 $\langle\alpha,\beta\rangle$ 是一般的非零复数, 设

$\langle\alpha,\beta\rangle=|\langle\alpha,\beta\rangle|\cdot e^{i\theta}$.

于是 $\langle\alpha e^{i\theta},\beta\rangle=e^{-i\theta}\langle\alpha,\beta\rangle=e^{-i\theta}|\langle\alpha,\beta\rangle|e^{i\theta}=|\langle\alpha,\beta\rangle|$ 为实数. 由已证结论有

$|\langle\alpha e^{i\theta},\beta\rangle|\le\|\alpha e^{i\theta}\|\cdot\|\beta\|$.

再由 $\|\alpha e^{i\theta}\|=\|\alpha\|$ 即得证.

3) 可利用 4)来证明.

注 1.2.1 (实内积空间时, Cauchy-Schwartz 不等式等号成立的条件)

实内积空间时, Cauchy-Schwartz 不等式中, 等号成立的充要条件是 α,β 线性相关.

证明: 充分性. 已知 α,β 线性相关, 即存在 k,l 不全为零, 使

$\alpha k+\beta l=0$.

不妨设 $k\neq 0$, 从而

$\alpha=-\beta\frac{l}{k}$, $\langle\alpha,\beta\rangle=\langle-\beta\frac{l}{k},\beta\rangle=-\frac{l}{k}\langle\beta,\beta\rangle$, $\|\alpha\|=\|-\beta\frac{k}{l}\|=|\frac{k}{l}|\cdot\|\beta\|$, $\|\alpha\|\cdot\|\beta\|=|\frac{k}{l}|\cdot\|\beta\|^2=|\frac{k}{l}|\langle\beta,\beta\rangle$.

必要性. 已知 $|\langle\alpha,\beta\rangle|=\|\alpha\|\cdot\|\beta\|$. 若 α,β 有一个为零向量, 则显然 α,β 线性相关. 下设 α,β 都不是零向量. 则此时, 实系数的二次函数 $f(t)=\langle\alpha,\alpha\rangle t^2+2\langle\alpha,\beta\rangle t+\langle\beta,\beta\rangle$ 的判别式 $(2\langle\alpha,\beta\rangle)^2-4\langle\alpha,\alpha\rangle\langle\beta,\beta\rangle=0$, 于是必有实数 t_0 使得 $f(t_0)=0$. 又 $f(t)=\langle\alpha t+\beta,\alpha t+\beta\rangle$, 因此 $\langle\alpha t_0+\beta,\alpha t_0+\beta\rangle=0$. 由命题 1.1.1, 有

$\alpha t_0+\beta=0$,

从而 α,β 线性相关. 证毕

由 Cauchy-Schwartz 不等式, 我们有

$$-1\le\frac{\langle\alpha,\beta\rangle}{\|\alpha\|\cdot\|\beta\|}\le 1.$$

因此, 我们可以有下面的定义

定义 1.2.2 (实内积空间两向量之间的夹角)

设 α,β 是实内积空间的两个非零向量. 定义 α,β 之间的夹角为

$$\angle_\alpha^\beta:=\arccos\frac{\langle\alpha,\beta\rangle}{\|\alpha\|\cdot\|\beta\|}.$$

定义 1.2.3 (复内积空间两向量正交)

设 α,β 是复(实)内积空间的两个非零向量. 若 $\langle\alpha,\beta\rangle=0$, 则称 α,β 相互正交, 记为 $\alpha\perp\beta$. 在实内积空间的情形, α,β 相互正交, 也就是 $\angle_\alpha^\beta=\frac{\pi}{2}$.

显然, 零向量与任何向量正交. 反之, 与任何向量都正交的向量必是零向量.

命题 1.2.2 (勾股定理)
设 α,β 是实内积空间的两个非零向量. 下列各条件等价:
(1) $\alpha \perp \beta$;
(2) $\angle_{\alpha}^{\beta} = \dfrac{\pi}{2}$;
(3) $\langle \alpha,\beta \rangle = 0$;
(4) $\|\alpha\|^2 + \|\beta\|^2 = \|\alpha+\beta\|^2$;
(5) $\|\alpha\|^2 + \|\beta\|^2 = \|\alpha-\beta\|^2$.
在复内积空间情形, 有
(1) $\Leftrightarrow$ (3) $\Rightarrow$ (4) $\Leftrightarrow$ (5).

注 1.2.2 (复内积空间勾股定理)
在复内积空间的情形, 易见

$$\begin{aligned}\|\alpha+\beta\|^2 &= \|\alpha\|^2 + \|\beta\|^2 + \langle \alpha,\beta \rangle + \langle \beta,\alpha \rangle \\ &= \|\alpha\|^2 + \|\beta\|^2 + 2\,\mathrm{Re}\langle \alpha,\beta \rangle.\end{aligned}$$

于是, $\|\alpha\|^2 + \|\beta\|^2 = \|\alpha+\beta\|^2$ 等价于 $\mathrm{Re}\langle \alpha,\beta \rangle = 0$. 只需举例说明有 $\mathrm{Re}\langle \alpha,\beta \rangle = 0$, 但 $\langle \alpha,\beta \rangle \neq 0$ 的情形. 考虑标准复内积空间 $\mathbb{C}^2$. 取

$$\alpha = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \beta = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} i.$$

则

$\langle \alpha,\beta \rangle = \overline{\alpha}^{\mathrm{T}}\beta = i \neq 0$, 但 $\mathrm{Re}\langle \alpha,\beta \rangle = 0$.

1.3 标准正交基与酉矩阵

定义 1.3.1 (正交组，标准正交组，标准正交基)
给定复内积空间 V .
(1) V 中若干个非零向量构成的向量组称为正交向量组, 简称正交组, 若组中任意两个向量正交.
(2) 正交向量组称为标准正交组, 若其中每个向量的长度为 1.
(3) 标准正交组若还是 V 的一个基, 则称为标准正交基.

定理 1.3.1 (正交组的 Gram 矩阵)
(1) 一个向量组为正交向量组的充要条件是其 Gram 矩阵为非奇异对角矩阵.
(2) 一个向量组为标准正交组的充要条件是其 Gram 矩阵为单位矩阵.

推论 1.3.1 (正交组是线性无关组)

此推论亦可直接证明如下:

内积空间中标准正交组的好处体现在

定理 1.3.2 (沿正交组展开时坐标的解耦性)
设 $\alpha_1,\alpha_2,\ldots,\alpha_s$ 是复内积空间 V 的一个正交组, $\beta\in V$. 若有

$$\beta=\alpha_1k_1+\alpha_2k_2+\cdots+\alpha_sk_s$$
$$=\begin{bmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_s\end{bmatrix}\begin{bmatrix}k_1\\k_2\\\vdots\\k_s\end{bmatrix}$$

则

$$k_i=\frac{\langle\alpha_i,\beta\rangle}{\langle\alpha_i,\alpha_i\rangle}, i=1,2,\ldots,s.$$

特别地, 若 $\alpha_1,\alpha_2,\ldots,\alpha_s$ 进而是标准正交组, 则

$$k_i=\langle\alpha_i,\beta\rangle, i=1,2,\ldots,s.$$

例 1.3.1 (对比 $\mathbb{R}^n$ 中正交组和非正交组的坐标计算，也可从 Cramer 法则加以说明)

注 1.3.1 (正交坐标计算的解耦性的矩阵形式)
直接将 V 中的抽象方程

$$\beta=\alpha_1k_1+\alpha_2k_2+\cdots+\alpha_sk_s$$
$$=\begin{bmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_s\end{bmatrix}\begin{bmatrix}k_1\\k_2\\\vdots\\k_s\end{bmatrix}$$

映射成复数域上的具体方程

$$G(\{\alpha_i\})k=G(\{\alpha_i\},\beta),$$

而此时 $G(\{\alpha_i\})$ 是对角非奇异矩阵

$$G(\{\alpha_i\})=diag(\|\alpha_i\|^2,i=1,2,\ldots,s).$$

例 1.3.2 (三角函数系作为平方可积函数空间中的正交向量组与函数的 Fourier 展开)

1.3.1Gram-Schmidt 正交化方法

下面给出一种将内积空间V中线性无关的向量组$\beta_1,\beta_2,\ldots,\beta_s$递归地改造成正交组$\alpha_1,\alpha_2,\ldots,\alpha_s$的方法, 称为 Gram-Schmidt 正交化方法.

$$\alpha_1=\beta_1,$$

$$\alpha_j=\beta_j-\left(\alpha_1 k_{1,j}+\alpha_2 k_{2,j}+\cdots+\alpha_{j-1}k_{j-1,j}\right), j=2,3,\ldots,s.$$

这里, 所有系数$k_{ij}, i<j$, 按如下方式递归地计算

$$k_{ij}=\frac{\langle\alpha_i,\beta_j\rangle}{\langle\alpha_i,\alpha_i\rangle}, i<j\ .$$

易知, 这样计算系数就保证了$\alpha_1,\alpha_2,\ldots,\alpha_s$的正交性.

定理 1.3.3 (Gram-Schmidt 正交化方法保持子空间序列)

$$\mathrm{span}\{\alpha_1,\alpha_2,\ldots,\alpha_t\}=\mathrm{span}\{\beta_1,\beta_2,\ldots,\beta_t\},\ t=1,2,\ldots,s\ .$$

记上三角矩阵

$$R_t=\begin{bmatrix}1 & k_{1,2} & \cdots & k_{1,t-1} & k_{1,t}\\ & 1 & & \vdots & k_{2,t}\\ & & \ddots & k_{t-2,t-1} & \vdots\\ & & & 1 & k_{t-1,t}\\ 0 & & & & 1\end{bmatrix}\in\mathbb{C}^{t\times t}, t=1,2,\ldots,$$

则易知K_t可逆, 且K_t^{-1}也是对角线为 1 的上三角矩阵.

定理 1.3.4 (Gram-Schmidt 正交化方法的抽象矩阵形式)

对$t=1,2,\ldots,$

$$\begin{bmatrix}\beta_1 & \cdots & \beta_t\end{bmatrix}=\begin{bmatrix}\alpha_1 & \cdots & \alpha_t\end{bmatrix}R_t,$$

$$\begin{bmatrix}\alpha_1 & \cdots & \alpha_t\end{bmatrix}=\begin{bmatrix}\beta_1 & \cdots & \beta_t\end{bmatrix}R_t^{-1}.$$

由此矩阵公式亦可看出定理 1.3.3.

显然, 任一正交向量组$\alpha_1,\alpha_2,\ldots,\alpha_s$可单位化成标准正交组$\tilde{\alpha}_1,\tilde{\alpha}_2,\ldots,\tilde{\alpha}_s$如下

$$\tilde{\alpha}_j=\alpha_j\frac{1}{\|\alpha_j\|},1,2,\ldots.$$

写成矩阵形式

$$\begin{bmatrix}\tilde{\alpha}_1 & \cdots & \tilde{\alpha}_t\end{bmatrix}=\begin{bmatrix}\alpha_1 & \cdots & \alpha_t\end{bmatrix}\begin{bmatrix}\frac{1}{\|\alpha_1\|} & & \\ & \ddots & \\ & & \frac{1}{\|\alpha_t\|}\end{bmatrix}.$$

进而有

$$\begin{bmatrix}\tilde{\alpha}_1 & \cdots & \tilde{\alpha}_t\end{bmatrix}=\begin{bmatrix}\beta_1 & \cdots & \beta_t\end{bmatrix}R_t^{-1}\begin{bmatrix}\frac{1}{\|\alpha_1\|} & & \\ & \ddots & \\ & & \frac{1}{\|\alpha_t\|}\end{bmatrix}.$$

这里, 复上三角矩阵

$$R_t^{-1}\begin{bmatrix} \frac{1}{\|\alpha_1\|} & & \\ & \ddots & \\ & & \frac{1}{\|\alpha_t\|} \end{bmatrix} = \begin{bmatrix} \frac{1}{\|\alpha_1\|} & \cdots & * \\ & \ddots & \vdots \\ 0 & & \frac{1}{\|\alpha_t\|} \end{bmatrix} =: H_t$$

的对角线是正的实数. 于是

$$[\widetilde{\alpha}_1 \quad \cdots \quad \widetilde{\alpha}_t] = [\beta_1 \quad \cdots \quad \beta_t]H_t,$$

即任一列满秩的抽象矩阵(即线性无关的向量组所拼成的矩阵), 都可以右乘以一个正实数对角线的上三角矩阵, 得到一个标准正交组所拼成的抽象矩阵. 这就是 Gram-Schmidt 标准正交化的抽象矩阵表述.

定理 1.3.5 (矩阵的正交-三角分解)
设 $A \in \mathbb{C}^{n\times r}$ 列满秩. 则存在唯一的一对矩阵 $Q \in \mathbb{C}^{n\times r}$ 和 $R \in \mathbb{C}^{r\times r}$ 满足:
(1) $A = QR$;
(2) $Q^H Q = I_r$, 或者说, Q 的列向量组是标准酉空间 $\mathbb{C}^n$ 中的标准正交组;
(3) R 是对角线元素为正实数的上三角矩阵.
特别地, 若 A 为非奇异方阵, 则相应的 Q 为酉矩阵.

例 1.3.3 (多项式函数空间中, 单项式函数组的 Gram-Schmidt 正交化)

1.3.2 酉矩阵

本小节研究标准酉空间 $\mathbb{C}^n$, 结果自然也包括了标准欧几里得空间 $\mathbb{R}^n$ 的情形. 回忆向量 $x, y \in \mathbb{C}^n$ 的标准内积

$$\langle x, y\rangle = \bar{x}_1 y_1 + \cdots + \bar{x}_n y_n = x^{\mathrm{H}} y.$$

定义 1.3.2 (酉矩阵作为标准酉空间中的标准正交基的基矩阵)
标准酉空间中的任一标准正交基所拼成的矩阵称为一个复正交矩阵, 常称为酉矩阵. 换言之, 复数域上的一个方阵称为酉矩阵, 如果其列向量组构成标准酉空间的一个标准正交基.

相应地, 有实正交矩阵.

按此定义及定理 1.3.1, 可知酉矩阵有如下等价定义

定义 1.3.3 (用矩阵的逆来定义酉矩阵)
矩阵 $U \in \mathbb{C}^{n\times n}$ 称为酉矩阵, 如果 $U^{-1} = U^{\mathrm{H}}$, 即

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

矩阵的求逆运算 $(\cdot)^{-1}$ 相比于矩阵的求共轭转置运算 $(\cdot)^{\mathrm{H}}$, 一般说来, 复杂程度相去天渊. 这就从矩阵运算的角度说明酉矩阵之好, 也说明了酉矩阵在计算数学中的重要地位.

定义 1.3.4 (酉相似)
设 $A, B \in \mathbb{C}^{n\times n}$. 若有酉矩阵 $U \in \mathbb{C}^{n\times n}$ 使得

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

则称 A 与 B 酉相似.

酉相似的几何意义是 A 作为线性变换在标准正交基 U 下的矩阵表示为 B.

定理 1.3.6 (酉矩阵保持度量)
设 $U \in \mathbb{C}^{n\times n}$, 则下列等价
(1) U 为酉矩阵, 即 U 的列向量组是标准酉空间 $\mathbb{C}^n$ 中的标准正交基.
(2) U^{H} 为酉矩阵, 即 U 的行向量组的共轭转置是标准酉空间 $\mathbb{C}^n$ 中的标准正交基.

(3) $U^{\mathrm{H}}U = I_n$.

(4) $UU^{\mathrm{H}} = I_n$.

(5) (保内积) 对任意 $x, y \in \mathbb{C}^n$,

$\langle Ux, Uy \rangle = \langle x, y \rangle$.

(6) (保长度) 对任意 $x \in \mathbb{C}^n$,

$\|Ux\| = \|x\|$.

例 1.3.4 (实正交矩阵的结构, Givens 旋转, Householder 镜像)

1.4 正交投影与最佳逼近

问题 1.4.1 (最佳逼近问题)

设V为复内积空间, 向量$\beta\in V$, W为V的一个子空间. 求$x^*\in W$, 使得

$d(\beta,x^*)=\min\{d(\beta,x):x\in W\}$. (1.4.1)

即

$d(\beta,x^*)\le d(\beta,x)$, 对任意$x\in W$.

或者写成优化问题的通常形式:

$\min\|x-\beta\|$,

s.t. $x\in W$.

我们关心解x^*的存在性, 唯一性, 及其具体求法.

许多应用科学问题可抽象为此问题模式.

- W为一维子空间的情形.

$W=\operatorname{span}\{\alpha\}$, $\alpha\ne 0$.

解法一: 化为函数极值问题(以实内积空为例来说明)

W中的元素可有如下的参数化表示:

$W=\{\alpha k:k\in\mathbb{R}\}$

可将问题化为一单变量函数

$k\mapsto f(k)=(d(\beta,\alpha k))^2$

的求最小值问题. 由

$f(k)=(d(\beta,\alpha k))^2=\|\beta-\alpha k\|^2=\langle\beta-\alpha k,\beta-\alpha k\rangle=\langle\alpha,\alpha\rangle k^2-2\langle\alpha,\beta\rangle k+\langle\beta,\beta\rangle$

是关于k的二次函数, 可知其有唯一的最小值, 最小值点为

$k^*=-\dfrac{-2\langle\alpha,\beta\rangle}{2\langle\alpha,\alpha\rangle}=\dfrac{\langle\alpha,\beta\rangle}{\langle\alpha,\alpha\rangle}=\langle\alpha,\alpha\rangle^{-1}\langle\alpha,\beta\rangle$.

于是问题(1.4.1)的唯一解为

$x^*=\alpha\langle\alpha,\alpha\rangle^{-1}\langle\alpha,\beta\rangle$ (1.4.2)

用内积的抽象矩阵乘法运算表达为

$x^*=\alpha\left(\alpha^{\mathrm T}\alpha\right)^{-1}\alpha^{\mathrm T}\beta$

解法二: 用勾股定理直接定系数

若$\beta-\alpha k^*\perp W$, 则对任意k, 注意到$(\beta-\alpha k)-(\beta-\alpha k^*)=\alpha(k^*-k)\in W$, 由勾股定理得

$\|\beta-\alpha k\|^2=\|\beta-\alpha k^*\|^2+\|\alpha(k^*-k)\|^2\ge\|\beta-\alpha k^*\|^2$

对任意k都成立. 因此, 使得$\beta-\alpha k^*\perp W$的αk^*为所求. 又$\beta-\alpha k^*\perp W$等价于$\beta-\alpha k^*\perp\alpha$, 于是

$\langle\alpha,\beta-\alpha k\rangle=0$ 即

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

$k^*=\langle\alpha,\alpha\rangle^{-1}\langle\alpha,\beta\rangle$

- W为有限维子空间的情形.

设$\alpha_1,\alpha_2,\ldots,\alpha_m$为$W$的一组基.

记

$$x=\alpha_1k_1+\alpha_2k_2+\cdots\alpha_mk_m=\begin{bmatrix}\alpha_1 & \cdots & \alpha_m\end{bmatrix}\begin{bmatrix}k_1\\ \vdots\\ k_m\end{bmatrix}$$

$\beta-x\perp W \Leftrightarrow \alpha_i\perp\beta-x, i=1,\ldots,m$

于是,

$\langle \alpha_i, \beta - x \rangle = 0, i = 1, \ldots, m$

写成矩阵形式,

$$\begin{bmatrix} \langle \alpha_1, \alpha_1 \rangle & \cdots & \langle \alpha_1, \alpha_m \rangle \\ \vdots & \vdots & \vdots \\ \langle \alpha_m, \alpha_1 \rangle & \cdots & \langle \alpha_m, \alpha_m \rangle \end{bmatrix} \begin{bmatrix} k_1 \\ \vdots \\ k_m \end{bmatrix} = \begin{bmatrix} \langle \alpha_1, \beta \rangle \\ \vdots \\ \langle \alpha_m, \beta \rangle \end{bmatrix}$$

由 $\alpha_1, \alpha_2, \ldots, \alpha_m$ 线性无关, 及定理 1.1.2, 知 Gram 矩阵

$$G(\alpha_1, \alpha_2, \ldots, \alpha_m) = \begin{bmatrix} \langle \alpha_1, \alpha_1 \rangle & \cdots & \langle \alpha_1, \alpha_m \rangle \\ \vdots & \vdots & \vdots \\ \langle \alpha_m, \alpha_1 \rangle & \cdots & \langle \alpha_m, \alpha_m \rangle \end{bmatrix}$$

非奇异, 因此存在唯一

$$\begin{bmatrix} k_1 \\ \vdots \\ k_m \end{bmatrix}^* = \begin{bmatrix} \langle \alpha_1, \alpha_1 \rangle & \cdots & \langle \alpha_1, \alpha_m \rangle \\ \vdots & \vdots & \vdots \\ \langle \alpha_m, \alpha_1 \rangle & \cdots & \langle \alpha_m, \alpha_m \rangle \end{bmatrix}^{-1} \begin{bmatrix} \langle \alpha_1, \beta \rangle \\ \vdots \\ \langle \alpha_m, \beta \rangle \end{bmatrix}.$$

于是

$$\begin{aligned} x^* &= \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix} \begin{bmatrix} k_1 \\ \vdots \\ k_m \end{bmatrix}^* \\ &= \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix} G(\alpha_1, \alpha_2, \ldots, \alpha_m)^{-1} \begin{bmatrix} \langle \alpha_1, \beta \rangle \\ \vdots \\ \langle \alpha_m, \beta \rangle \end{bmatrix} \\ &= \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix} G(\alpha_1, \alpha_2, \ldots, \alpha_m)^{-1} G(\alpha_1, \alpha_2, \ldots, \alpha_m; \beta). \end{aligned}$$

用交互 Gram 矩阵的抽象矩阵乘法运算表达

$$x^* = \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix} \left(\begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix}^{\mathrm{H}} \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix} \right)^{-1} \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix}^{\mathrm{H}} \beta.$$

进一步, 将 W 和它的基矩阵 $\begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix}$ 用同一符号 W 表示, 可得更紧凑的表示

$$x^* = W\left(W^{\mathrm{H}} W\right)^{-1} W^{\mathrm{H}} \beta. \quad (1.4.3)$$

注意, 这里 $\left(W^{\mathrm{H}} W\right)^{-1} \in \mathbb{C}^{m \times m}$ 及 $W^{\mathrm{H}} \beta \in \mathbb{C}^{m \times 1}$, 因而 $\left(W^{\mathrm{H}} W\right)^{-1} W^{\mathrm{H}} \beta \in \mathbb{C}^{m \times 1}$ 是通常的数字矩阵. 而抽象矩阵 W 和"具体"数字矩阵 $\left(W^{\mathrm{H}} W\right)^{-1} W^{\mathrm{H}} \beta$ 之间的乘法运算, 是线性空间 V 中的加法和数乘法(数乘法的数写在右边)运算的矩阵形式表达. 固定子空间 W, 公式(1.4.3)决定了一个从 V 到 W 的映射

$$\mathcal{P}_W : \beta \mapsto x^* = W\left(W^{\mathrm{H}} W\right)^{-1} W^{\mathrm{H}} \beta.$$

可形式地记

$$\mathcal{P}_W(\cdot) = W\left(W^{\mathrm{H}} W\right)^{-1} W^{\mathrm{H}}(\cdot).$$

这是向子空间 W 的正交投影算子.

注 1.4.1 (行体制形式的正交投影算子)

此时, 抽象向量被类比成行数组, 数乘法的数写在抽象向量的左边, 向量组按行拼成抽象矩阵. 故

$$W = \operatorname{span}\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$$

的相应的抽象基矩阵, 亦记为 W, 为

$$W = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix}.$$

$$\mathcal{P}^W:\beta\mapsto x^*=G(\beta;\{\alpha_i\})G(\{\alpha_i\})^{-1}\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_m\end{bmatrix}.$$
$$=\beta W^{\mathrm{H}}\left(WW^{\mathrm{H}}\right)^{-1}W$$
可形式地记
$$\mathcal{P}^W(\cdot)=(\cdot)W^{\mathrm{H}}\left(WW^{\mathrm{H}}\right)^{-1}W.$$
这是向子空间W的正交投影算子.

例 1.4.1 (标准酉空间$\mathbb{C}^n$中, 向子空间$\mathrm{im}A$的正交投影矩阵)
这里, A是列满秩的. 记矩阵
$$P_A=A\left(A^{\mathrm{H}}A\right)^{-1}A^{\mathrm{H}}.$$
则
$$x^*=P_A\beta.$$
对固定的A, 线性映射$\beta\mapsto x^*$是从酉空间$V=\mathbb{C}^n$到子空间$W=\mathrm{im}A$的投影算子, 它由投影矩阵P_A决定.

例 1.4.2 (最小二乘数据拟合)
对“因变量”y和p个 “自变量”$x_1,x_2,\ldots,x_p$进行n次观测, 得数据如下表:

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

为方便起见，补充常值变量$x_0\equiv 1$作为“自变量”. 要找“经验公式”
$$y=a_0+a_1x_1+a_2x_2+\cdots+a_px_p$$
$$=1a_0+x_1a_1+x_2a_2+\cdots+x_pa_p$$
$$=\begin{bmatrix}1 & x_1 & x_2 & \cdots & x_p\end{bmatrix}\begin{bmatrix}a_0\\ a_1\\ a_2\\ \vdots\\ a_p\end{bmatrix},$$
即求出“参数”$a_0,a_1,\ldots,a_p$，使得这个函数关系能最好地“拟合”那n组观测数据.
最小二乘准则的几何解释
设
$$V=\mathbb{R}^n,$$
$$Y=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix},$$

$$\alpha_0=\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix},\alpha_1=\begin{bmatrix}x_{11}\\x_{21}\\\vdots\\x_{n1}\end{bmatrix},\alpha_2=\begin{bmatrix}x_{12}\\x_{22}\\\vdots\\x_{n2}\end{bmatrix},\cdots,\alpha_p=\begin{bmatrix}x_{1p}\\x_{2p}\\\vdots\\x_{np}\end{bmatrix},$$

$$X=\begin{bmatrix}\alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_p\end{bmatrix}\in\mathbb{R}^{n\times(p+1)}$$

$$W=\mathrm{im}X$$

$$k=\begin{bmatrix}a_0\\a_1\\a_2\\\vdots\\a_p\end{bmatrix}.$$

于是, 最小二乘估计

$$\hat{k}=\left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}Y.$$

例 1.4.3 (多项式拟合)

例 1.4.4 (概率论中的回归分析)

$V=\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$,

$X_1,X_2,\ldots,X_m$ 及 Y 是 $m+1$ 个有二阶矩的随机变量, 即 $X_1,X_2,\ldots,X_m,Y\in V$

$W=\mathrm{span}\{1,X_1,X_2,\ldots,X_m\}$

现将 Y 向子空间 W 正交投影.

注意到

$$\begin{aligned}W&=\mathrm{span}\{1,X_1,X_2,\ldots,X_m\}\\&=\mathrm{span}\{1,X_1-E(X_1),X_2-E(X_2),\ldots,X_m-E(X_m)\},\end{aligned}$$

多个随机变量 $X_1,X_2,\ldots,X_m$ 放在一起, 被非常不幸地俗称为一个“m-维随机向量”, 并被记成

$$X=\begin{bmatrix}X_1\\X_2\\\vdots\\X_m\end{bmatrix}$$

的形式. 其实, 每个随机变量 X_i 作为向量空间 $\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$ 中的元素, 从向量空间的观点看, 就已经是一个抽象向量. 因而多个随机变量 $X_1,X_2,\ldots,X_m$ 从向量空间的观点看, 其正确的名称是向量空间 $\mathcal{L}^2(\Omega,\mathcal{B},P(\cdot))$ 中的向量组, 记号

$$X=\begin{bmatrix}X_1\\X_2\\\vdots\\X_m\end{bmatrix}$$

其实是向量组按行体制(数乘法的数写在左边)拼成的抽象矩阵.

W 是抽象矩阵

$$\begin{bmatrix} 1 \\ X \end{bmatrix} = \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}$$

的行向量组张成的子空间，也是

$$\begin{bmatrix} 1 \\ X - E(X) \end{bmatrix} = \begin{bmatrix} 1 \\ X_1 - E(X_1) \\ X_2 - E(X_2) \\ \vdots \\ X_m - E(X_m) \end{bmatrix}$$

的行向量组张成的子空间. 于是，Y 向子空间 W 的正交投影为

$$\begin{aligned}
\mathcal{P}^W(Y) &= YW^{\mathrm{T}}\left(WW^{\mathrm{T}}\right)^{-1}W \\
&= E\left\{Y\begin{bmatrix} 1 \\ X - E(X) \end{bmatrix}^{\mathrm{T}}\right\}\left(E\left\{\begin{bmatrix} 1 \\ X - E(X) \end{bmatrix}\begin{bmatrix} 1 \\ X - E(X) \end{bmatrix}^{\mathrm{T}}\right\}\right)^{-1}\begin{bmatrix} 1 \\ X - E(X) \end{bmatrix}.
\end{aligned}$$

注意到

$$\begin{aligned}
E\left\{Y\begin{bmatrix} 1 \\ X - E(X) \end{bmatrix}^{\mathrm{T}}\right\} &= \left[E(Y) \quad E\left\{Y\left(X - E(X)\right)^{\mathrm{T}}\right\}\right] \\
&= \left[E(Y) \quad E\left\{\left(Y - E(Y)\right)\left(X - E(X)\right)^{\mathrm{T}}\right\}\right] \\
&= \begin{bmatrix} E(Y) & \underbrace{\mathrm{cov}(Y, X)}_{=:\Sigma_{Y,X}} \end{bmatrix},
\end{aligned}$$

$$E\left\{\begin{bmatrix} 1 \\ X - E(X) \end{bmatrix}\begin{bmatrix} 1 \\ X - E(X) \end{bmatrix}^{\mathrm{T}}\right\} = \begin{bmatrix} 1 & 0 \\ 0 & \underbrace{\mathrm{cov}(X, X)}_{=:\Sigma_X} \end{bmatrix},$$

于是

$$\begin{aligned}
\mathcal{P}^W(Y) &= YW^{\mathrm{T}}\left(WW^{\mathrm{T}}\right)^{-1}W \\
&= \left[E(Y) \quad \Sigma_{Y,X}\right]\begin{bmatrix} 1 & 0 \\ 0 & (\Sigma_X)^{-1} \end{bmatrix}\begin{bmatrix} 1 \\ X - E(X) \end{bmatrix} \\
&= \left[E(Y) \quad \Sigma_{Y,X}\left(\Sigma_X\right)^{-1}\right]\begin{bmatrix} 1 \\ X - E(X) \end{bmatrix} \\
&= E(Y) + \Sigma_{Y,X}\left(\Sigma_X\right)^{-1}(X - E(X)).
\end{aligned}$$

关于

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}$$

的另外一个正确的名称是向量值的随机变量，即可测映射 $X(\cdot): \Omega \to \mathbb{R}^m$.

1.4.4 子空间由正交基给出时的投影算子公式

设 $\alpha_1,\alpha_2,\ldots,\alpha_m$ 为 W 的一个标准正交基. 为减少符号, 直接记

$W=\begin{bmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_m\end{bmatrix}$.

则投影算子

$$\begin{aligned}\mathcal{P}_W(\cdot) &= W\left(W^{\mathrm{H}}W\right)^{-1}W^{\mathrm{H}}(\cdot)\\ &= W\left(I_m\right)^{-1}W^{\mathrm{H}}(\cdot)\\ &= WW^{\mathrm{H}}(\cdot).\end{aligned}$$

从而

$$\begin{aligned}\mathcal{P}_W(\beta) &= WW^{\mathrm{H}}\beta\\ &= WG(\{\alpha_i\},\beta)\\ &= \begin{bmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_m\end{bmatrix}\begin{bmatrix}\langle\alpha_1,\beta\rangle\\ \langle\alpha_2,\beta\rangle\\ \vdots\\ \langle\alpha_m,\beta\rangle\end{bmatrix}\\ &= \alpha_1\langle\alpha_1,\beta\rangle+\alpha_2\langle\alpha_2,\beta\rangle+\cdots+\alpha_m\langle\alpha_m,\beta\rangle.\end{aligned}$$

例 1.4.5 (标准酉空间中子空间由正交基给出时的投影矩阵公式)

$W\in\mathbb{C}^{n\times m}$ 的列向量组为标准正交组, 则向子空间 $\mathrm{im}W$ 的投影矩阵为

$P_W=WW^{\mathrm{H}}$.

这是一个“具体”的 n 阶复方阵.

例 1.4.6 (Fourier 展开作为正交投影)

1.5 正规矩阵

定理 1.5.1 (Schur 定理)

任一复方阵均可酉相似于上三角矩阵.

证明: (只需把第一章中的证明所用的基底取成标准正交基即可. 鉴于此定理的重要性, 这里依然详细写出)

对 n 用数学归纳法. $n=1$ 时, 显然. 假设 $n \le m$ 时命题成立, 下证 $n=m+1$ 时也成立.

设 $A \in \mathbb{C}^{(m+1)\times(m+1)}$. 由于 $m+1$ 次复特征多项式 $|\lambda I - A|$ 在复数域 $\mathbb{C}$ 中总有根, 任取其一根, 记为 λ_1, 相应的一个特征向量记为 p_1, 这里可使 p_1 的长度 $\|p_1\|=1$, 则

$Ap_1 = p_1\lambda_1$. (1.5.1)

将 p_1 扩充成 $\mathbb{C}^{(m+1)}$ 的一个标准正交基如下:

$p_1, q_2, \dots, q_{m+1}$,

则

$$A\underbrace{\begin{bmatrix} p_1 & q_2 & \cdots & q_{m+1} \end{bmatrix}}_{R} = \begin{bmatrix} p_1 & q_2 & \cdots & q_{m+1} \end{bmatrix} \underbrace{\begin{bmatrix} \lambda_1 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & * & \cdots & * \end{bmatrix}}_{\begin{bmatrix} \lambda_1 & \alpha \\ 0 & B \end{bmatrix}}. \quad (1.5.2)$$

由 $B \in \mathbb{C}^{m\times m}$ 及归纳假设, 存在 m 阶酉矩阵 Q 使得

$$Q^{\mathrm{H}} B Q = \begin{bmatrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_{m+1} \end{bmatrix}.$$

于是,

$$\begin{bmatrix} \lambda_1 & \alpha \\ 0 & B \end{bmatrix} = \left[\begin{array}{c|c} \lambda_1 & \alpha \\ \hline 0 & Q\begin{bmatrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_{m+1} \end{bmatrix} Q^{-1} \end{array}\right]$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \left[\begin{array}{c|ccc} \lambda_1 & & \alpha Q & \\ \hline & \lambda_2 & & * \\ 0 & & \ddots & \\ & 0 & & \lambda_{m+1} \end{array}\right] \begin{bmatrix} 1 & 0 \\ 0 & Q^{-1} \end{bmatrix}.$$

由此及(1.5.2), 有

$$R^{-1} A R = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \left[\begin{array}{c|ccc} \lambda_1 & & \alpha Q & \\ \hline & \lambda_2 & & * \\ 0 & & \ddots & \\ & 0 & & \lambda_{m+1} \end{array}\right] \begin{bmatrix} 1 & 0 \\ 0 & Q^{-1} \end{bmatrix},$$

从而

$$\left(R\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}\right)^{-1} A \left(R\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}\right) = \left[\begin{array}{c|ccc} \lambda_1 & & \alpha Q & \\ \hline & \lambda_2 & & * \\ 0 & & \ddots & \\ & 0 & & \lambda_{m+1} \end{array}\right]$$

即为上三角矩阵. 这里, $R\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$ 为酉矩阵. 证毕.

证法二: 已知对任意 $A \in \mathbb{C}^{n\times n}$, 存在非奇异矩阵 $P \in \mathbb{C}^{n\times n}$ 使得

$P^{-1}AP = J$
为上三角矩阵. 现对 P 的列向量组实施 Gram-schmidt 标准正交化, 即对 P 作正交-三角分解如下
$P = UT$,
其中 U 为酉矩阵, T 是(正实数对角线的)上三角矩阵. 于是
$U^{-1}AU = TJT^{-1}$.
易知, TJT^{-1} 为上三角矩阵.

定义 1.5.1 (正规矩阵)
一个复方阵若与它的共轭转置矩阵可交换, 则称为正规矩阵. 即 $A \in \mathbb{C}^{n\times n}$ 称为正规矩阵, 若
$AA^{\mathrm{H}} = A^{\mathrm{H}}A$.

例 1.5.1 (正规矩阵的典型例子)
Hermit 矩阵(对称矩阵); 反 Hermit 矩阵(反对称矩阵); 酉矩阵(正交矩阵)

推论 1.5.1 (矩阵正规性的酉相似不变性)
设 A 是正规矩阵, 且 A 酉相似于 B, 则 B 也是正规矩阵.

证明: 由已知, 有
$U^{\mathrm{H}}AU = B$.
于是

$$\begin{aligned} BB^{\mathrm{H}} &= \left(U^{\mathrm{H}}AU\right)\left(U^{\mathrm{H}}AU\right)^{\mathrm{H}} \\ &= \left(U^{\mathrm{H}}AU\right)\left(U^{\mathrm{H}}A^{\mathrm{H}}U\right) \\ &= U^{\mathrm{H}}AA^{\mathrm{H}}U \\ &= U^{\mathrm{H}}A^{\mathrm{H}}AU \\ &= U^{\mathrm{H}}A^{\mathrm{H}}UU^{\mathrm{H}}AU \\ &= B^{\mathrm{H}}B. \end{aligned}$$

推论 1.5.2 (上三角的正规矩阵是对角矩阵)
上三角矩阵若还是正规矩阵, 则必为对角矩阵.

定理 1.5.2 (正规矩阵与酉相似对角化)
$A \in \mathbb{C}^{n\times n}$ 可酉相似对角化, 当且仅当 A 是正规矩阵.

回忆, 此定理等价于说, n 阶复方阵有 n 个相互正交的特征向量的充要条件, 是该矩阵为正规矩阵.

定理 1.5.3 (三种特殊正规矩阵的特征值)
设 A 是正规矩阵, 则
(1) A 是 Hermit 矩阵的充要条件是 A 的特征值是实数.
(2) A 是反 Hermit 矩阵的充要条件是 A 的特征值的实部为零.
(2) A 是酉矩阵的充要条件是 A 的特征值的模等于 1.

1.5.3 正规矩阵酉相似对角化的具体求法

设 A 是 n 阶正规矩阵. 由定理 1.5.2 知
(1) A 的代数重数等于 A 的几何重数. 也就是说, 特征根 c 是特征多项式 $\left|\lambda I_n - A\right|$ 的 r 重根, 等价于齐次线性方程组
$\left(cI_n - A\right)x = 0$

的解空间$\ker(cI_n - A)$(称为c的特征子空间)的维数是r；或者说相应于r重特征根c，恰有r个线性无关的特征向量. 更简洁地

$r = \dim(\ker(cI_n - A)) = n - \operatorname{rank}(cI_n - A)$.

(2) 相应于不同特征值的特征子空间相互正交. 设λ_1，λ_2是两个不同的特征值, 则

$\ker(\lambda_1 I_n - A) \perp \ker(\lambda_2 I_n - A)$.

问题: 已知n阶正规矩阵A，求对角矩阵Λ及酉矩阵U，使得$U^{\mathrm{H}} AU = \Lambda$.

解法:

第一步. 求出$|\lambda I_n - A|$的所有不同的根, 记为$\lambda_1, \ldots, \lambda_p$.

第二步. 对每个λ_i，求齐次线性方程组$(\lambda_i I_n - A)x = 0$的一个基础解系, 记为

$v_1^i, \ldots, v_{r_i}^i$.

第三步. 对向量组$u_1^i, \ldots, u_{r_i}^i$实施 Gram-Schmidt 标准正交化, 得

$u_1^i, \ldots, u_{r_i}^i$.

第四步. 拼出Λ及U如下:

$\Lambda = \operatorname{diag}\{\lambda_i I_{r_i}, i = 1, \ldots, p\}$,

$U = \begin{bmatrix} v_1^1 \cdots v_{r_1}^1 & \cdots & v_1^i \cdots v_{r_i}^i & \cdots & v_1^p \cdots v_{r_p}^p \end{bmatrix}$.

1.6 Hermit 矩阵与 Hermit 二次型

考虑标准酉空间$\mathbb{C}^n$及标准欧几里得空间$\mathbb{R}^n$.

定理 1.6.1 (Hermit 矩阵的内积刻画)
设$A \in \mathbb{C}^{n\times n}$. 则下列条件等价
(1) $A^{\mathrm{H}} = A$, 即A为 Hermit 矩阵.
(2) 对任意的$x, y \in \mathbb{C}^n$, $\langle Ax, y\rangle = \langle x, Ay\rangle$.
(3) 对任意的$x \in \mathbb{C}^n$, $\langle Ax, x\rangle$是实数.

证明: $(1)\Rightarrow(2)$及$(1)\Rightarrow(3)$ 显然.
$(2)\Rightarrow(1)$ 对任意的i, j, 分别取x, y为单位矩阵I_n的第i, j列e_i, e_j, 则
$\langle Ae_i, e_j\rangle = e_i^{\mathrm{H}} A^{\mathrm{H}} e_j = \overline{a}_{ji}$,
$\langle e_i, Ae_j\rangle = e_i^{\mathrm{H}} A e_j = a_{ij}$.
因此, $a_{ij} = \overline{a}_{ji}$. 于是$A^{\mathrm{H}} = A$.
$(3)\Rightarrow(1)$ 对任意的$x, y \in \mathbb{C}^n$, 由(3)知$\langle Ax, x\rangle$, $\langle Ay, y\rangle$, $\langle A(x+y), (x+y)\rangle$及$\langle A(x+y\mathrm{i}), (x+y\mathrm{i})\rangle$均为实数, 这里$\mathrm{i} = \sqrt{-1}$为虚数单位. 又
$\langle A(x+y), (x+y)\rangle = \langle Ax, x\rangle + \langle Ay, y\rangle + \langle Ax, y\rangle + \langle Ay, x\rangle$,

$$
\begin{aligned}
\langle A(x+y\mathrm{i}), (x+y\mathrm{i})\rangle &= \langle Ax, x\rangle + \langle A(y\mathrm{i}), (y\mathrm{i})\rangle + \langle Ax, y\mathrm{i}\rangle + \langle Ay\mathrm{i}, x\rangle \\
&= \langle Ax, x\rangle - \mathrm{i}\langle Ay, y\rangle\mathrm{i} + \langle Ax, y\rangle\mathrm{i} - \mathrm{i}\langle Ay, x\rangle \\
&= \langle Ax, x\rangle + \langle Ay, y\rangle + \mathrm{i}\big(\langle Ax, y\rangle - \langle Ay, x\rangle\big),
\end{aligned}
$$

故$\langle Ax, y\rangle + \langle Ay, x\rangle = x^{\mathrm{H}}A^{\mathrm{H}}y + y^{\mathrm{H}}A^{\mathrm{H}}x$是实数, $\langle Ax, y\rangle - \langle Ay, x\rangle = x^{\mathrm{H}}A^{\mathrm{H}}y - y^{\mathrm{H}}A^{\mathrm{H}}x$是纯虚数. 记$B = A^{\mathrm{H}} = \left[b_{ij}\right]_{n\times n}$, 对任意的$i, j$, 分别取$x, y$为单位矩阵$I_n$的第$i, j$列$e_i, e_j$, 即知$b_{ij} + b_{ji}$为实数, $b_{ij} - b_{ji}$为纯虚数. 从而
$\mathrm{Im}(b_{ij}) = -\mathrm{Im}(b_{ji})$,
$\mathrm{Re}(b_{ij}) = \mathrm{Re}(b_{ji})$,
此即$b_{ij} = \overline{b}_{ji}$. 于是$B = B^{\mathrm{H}}$.

定义 1.6.1 (Hermit 二次型)
给定n阶 Hermit 矩阵A. 映射$x \mapsto \langle Ax, x\rangle = x^{\mathrm{H}}Ax$确定从$\mathbb{C}^n$到$\mathbb{R}$的函数, 称为 Hermit 二次型. 写出分量形式, 有

$$x^{\mathrm{H}}Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\overline{x}_i x_j .$$

注 1.6.1 (复变量的多元二次齐次多项式何时是 Hermit 二次型?)
由定理 1.6.1, 复变量的多元二次齐次多项式是 Hermit 二次型, 当且仅当该多项式的取值恒为实数.
(多元二次齐次多项式可以像实数一样写成二次型, 但问题出在共轭运算!)

定义 1.6.2 (Hermit 矩阵的合同)
设A和B是两个n阶 Hermit 矩阵, P是n阶非奇异矩阵. 若有
$P^{\mathrm{H}}AP = B$,
则称A和B关于P合同. 此时也简称A与B合同, 并称P为相应的合同矩阵或合同变换矩阵.

注 1.6.2 (Hermit 矩阵的酉相似也是合同)

注 1.6.3 (Hermit 二次型在坐标变换下的动态) Hermit 二次型与坐标变换两个概念的结合，导致 Hermit 矩阵的合同概念. 向量 $x \in \mathbb{C}^n$ 在 $\mathbb{C}^n$ 的标准基 I_n 下的坐标就是 x 自己: $x = I_n x$. 它在 $\mathbb{C}^n$ 的新的基底 P 下的坐标 $\widetilde{x}$ 满足: $x = P\widetilde{x}$. 于是

$$x^{\mathrm{H}} Ax = (P\widetilde{x})^{\mathrm{H}} A(P\widetilde{x}) = \widetilde{x}^{\mathrm{H}} \underbrace{\left(P^{\mathrm{H}} AP\right)}_{=:B}\widetilde{x}.$$

由 Hermit 是正规矩阵，可找到酉矩阵 U，将 Hermit 二次型 $x^{\mathrm{H}} Ax$ 通过正交坐标变换 $x = U\widetilde{x}$ 化为“平方和”形式

$$x^{\mathrm{H}} Ax = \widetilde{x}^{\mathrm{H}} (U^{\mathrm{H}} AU)\widetilde{x} = \lambda_1 \bar{x}_1 x_1 + \lambda_2 \bar{x}_2 x_2 + \cdots + \lambda_n \bar{x}_n x_n = \lambda_1 |x_1|^2 + \lambda_2 |x_2|^2 + \cdots + \lambda_n |x_n|^2,$$

这里，$\lambda_1, \lambda_2, \ldots, \lambda_n$ 是 A 的 n 个特征值.

1.6.1 Hermit 正定矩阵与 Hermit 正定二次型

定义 1.6.3 (正定矩阵与正定二次型)
给定 Hermit 矩阵 A，如果对任意 $x \in \mathbb{C}^n$，$x \neq 0$，有 $x^{\mathrm{H}} Ax > 0\,(\geq 0)$，则称 $x^{\mathrm{H}} Ax$ 为 Hermit 正定(半正定)二次型，A 为 Hermit 正定(半正定)矩阵.

这里，我们只补充大学阶段通常的线性代数课程未包含的内容.

定理 1.6.2 (正定性的 Sylvester 行列式判别法)
n 阶 Hermit 矩阵 A 是正定的，当且仅当它的 n 个顺序主子式都大于零.
证明: 用数学归纳法. $n = 1$ 时，结论显然. 现假设 $n = m - 1$ 时结论成立，往证 $n = m$ 时结论成立. 将 A 作如下分块

$$A = \begin{bmatrix} A_1 & a \\ a^{\mathrm{H}} & a_{m,m} \end{bmatrix}.$$

必要性. 已知 A 是正定的. 对任意 $z \in \mathbb{C}^{m-1}, z \neq 0$，令

$$x = \begin{bmatrix} z \\ 0 \end{bmatrix},$$

则 $x \in \mathbb{C}^m, x \neq 0$. 于是

$$x^{\mathrm{H}} Ax = \begin{bmatrix} z \\ 0 \end{bmatrix}^{\mathrm{H}} \begin{bmatrix} A_1 & a \\ a^{\mathrm{H}} & a_{m,m} \end{bmatrix} \begin{bmatrix} z \\ 0 \end{bmatrix} = z^{\mathrm{H}} A_1 z > 0.$$

亦即 A_1 是 $m-1$ 阶 Hermit 正定的. 由归纳假设，A_1 的 $m-1$ 个顺序主子式大于零，也就是 A 的前 $m-1$ 个顺序主子式大于零. 这蕴含了 A_1 为非奇异的. 又

$$\begin{bmatrix} I_{m-1} & 0 \\ -a^{\mathrm{H}} A_1^{-1} & 1 \end{bmatrix} \begin{bmatrix} A_1 & a \\ a^{\mathrm{H}} & a_{m,m} \end{bmatrix} \begin{bmatrix} I_{m-1} & -A_1^{-1} a \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} A_1 & 0 \\ 0 & a_{m,m} - a^{\mathrm{H}} A_1^{-1} a \end{bmatrix}, \quad (1.6.1.1)$$

作坐标变换

$$x = \begin{bmatrix} I_{m-1} & -A_1^{-1} a \\ 0 & 1 \end{bmatrix} y,$$

并记

$$y = \begin{bmatrix} w \\ y_m \end{bmatrix},$$

则

$$x^{\mathrm{H}}Ax=\begin{bmatrix} w \\ y_m \end{bmatrix}^{\mathrm{H}}\begin{bmatrix} A_1 & 0 \\ 0 & a_{m,m}-a^{\mathrm{H}}A_1^{-1}a \end{bmatrix}\begin{bmatrix} w \\ y_m \end{bmatrix}=w^{\mathrm{H}}A_1w+(a_{m,m}-a^{\mathrm{H}}A_1^{-1}a)|y_m|^2 .(1.6.1.2)$$

A 的正定保证了 $(a_{m,m}-a^{\mathrm{H}}A_1^{-1}a)>0$. 又由(1.6.1.1)及行列式性质有

$$|A|=|A_1|(a_{m,m}-a^{\mathrm{H}}A_1^{-1}a) .(1.6.1.3)$$

因此 $|A|>0$.

充分性. 已知 A 的 m 个顺序主子式都大于零. 这蕴含了 $m-1$ 阶 Hermit 矩阵 A_1 为非奇异的, 从而式(1.6.1.1), (1.6.1.2)及(1.6.1.3)都成立. 由 A 的 m 阶顺序主子式即 $|A|>0$, $m-1$ 阶顺序主子式即 $|A_1|>0$ 大于零, 及(1.6.1.3)知

$$(a_{m,m}-a^{\mathrm{H}}A_1^{-1}a)>0 .$$

再由归纳假设, 及(1.6.1.2), 知 A 为正定. 证毕

注 1.6.4 (半正定性的 Sylvester 行列式判别法)
Hermit 矩阵 A 是半正定的, 当且仅当它的所有各阶主子式都非负.
这里要注意, 是说的所有各阶主子式(共 $\mathrm{C}_n^1+\mathrm{C}_n^2+\cdots+\mathrm{C}_n^n=2^n-1$ 个)的非负性, 而不只是顺序主子式(共 n 个)的非负性. 例如

$$\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$$

的两个顺序主子式都非负, 但它不是半正定的. 这里, 它的 $2^2-1=3$ 个 主子式中, 有两个为 0, 一个为 -1.

注 1.6.5 (正定性的判别不须求特征根)
虽然 Hermit 矩阵正定的充要条件是它的特征值都是正数, 但 Sylvester 行列式判别法说明正定性的判定不需求特征多项式的根, 只需计算矩阵的行列式.

定理 1.6.3 (正定矩阵开平方)
设矩阵 A 是 Hermit 正定矩阵. 则存在唯一的 Hermit 正定矩阵 B 使得

$$A=B^2 .$$

(唯一性有趣, 见史荣昌 p.127)

定理 1.6.4 (Gram 矩阵与 Hermit 半正定矩阵)
Hermit 矩阵是半正定的充要条件是它可以写成一个复内积空间中某个向量组的 Gram 矩阵.

1.6.2 Hermit 矩阵特征值的极值刻画

由定理 1.5.3, Hermit 矩阵 A 的 n 个特征值 $\lambda_1,\lambda_2,\ldots,\lambda_n$ 都是实数. 不妨设

$$\lambda_1\le\lambda_2\le\cdots\le\lambda_n .$$

又回忆定理 1.6.1 (3), $\langle Ax,x\rangle=x^{\mathrm{H}}Ax$ 总是实数. 我们有

定理 1.6.5 (Hermit 矩阵特征值的极值刻画)

$$\lambda_{\min}(A)=\lambda_1=\min_{x\in\mathbb{C}^n,x\ne0}\frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x}=\min_{x\in\mathbb{C}^n,\|x\|=1}x^{\mathrm{H}}Ax ,$$

$$\lambda_{\max}(A)=\lambda_n=\max_{x\in\mathbb{C}^n,x\ne0}\frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x}=\max_{x\in\mathbb{C}^n,\|x\|=1}x^{\mathrm{H}}Ax .$$

注 1.6.6 (Hermit 矩阵特征值与 Rayleigh 商)

从 $\mathbb{C}^n \setminus \{0\}$ 到 $\mathbb{R}$ 的函数

$$x \mapsto R(x) = \frac{x^{\mathrm{H}} A x}{x^{\mathrm{H}} x}$$

称为 Hermit 矩阵 A 的 Rayleigh 商. A 的其他特征值 $\lambda_2, \ldots, \lambda_{n-1}$ 也可通过 A 的 Rayleigh 商用极值来刻画. 对此我们不在详述.

注 1.6.7 (极值刻画将特征多项式求根与二次型求极值联系起来)

1.6.3 矩阵的奇异值分解

推论 1.6.1 (标准酉空间中的 Gram 矩阵)

设 $A \in \mathbb{C}^{m\times n}$, $\mathrm{rank}(A) = r$. 则

(1) $A^{\mathrm{H}} A$ 和 AA^{H} 分别是 n 阶和 m 阶 Hermit 半正定矩阵.

(2) $\mathrm{rank}(A^{\mathrm{H}} A) = \mathrm{rank}(AA^{\mathrm{H}}) = r$.

证明: 注意到 n 阶方阵 $A^{\mathrm{H}} A$ 是 A 的 n 个列向量作成的标准酉空间 $\mathbb{C}^m$ 中的向量组的 Gram 矩阵, 本引理结论其实是定理 1.1.2 的一种具体情况.

定理 1.6.6 (矩阵的奇异值分解)

设 $A \in \mathbb{C}^{m\times n}$. 则存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V, 使得

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

其中 $r = \mathrm{rank}(A)$, r 阶对角矩阵

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

$\sigma_i > 0$, $i = 1, \ldots, r$.

证明: 注意到 n 阶方阵 $A^{\mathrm{H}} A$ 是 A 的 n 个列向量作成的标准酉空间 $\mathbb{C}^m$ 中的向量组的 Gram 矩阵, 从而由定理 1.1.2, 它是 Hermit 半正定矩阵, 且 $\mathrm{rank}(A^{\mathrm{H}} A) = \mathrm{rank}(A) = r$.

推论 1.6.2 ($A^{\mathrm{H}} A$ 与 AA^{H} 有相同的非零特征值)

注 1.6.8 (奇异值分解的几何意义)

由 $A \in \mathbb{C}^{m\times n}$ 决定的从 $\mathbb{C}^n$ 到 $\mathbb{C}^m$ 的线性映射 $x \mapsto y = Ax$, 在 $\mathbb{C}^n$ 中的标准正交基 V (入口基)和 $\mathbb{C}^m$ 中的标准正交基 U (出口基)下的矩阵表示为

$$\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}.$$

以下讨论奇异值的几何意义和极值刻画.

定义 1.6.4 (高维空间中的椭球面) (复的也完全可以!)

设 S 是 n 阶 Hermit 正定矩阵, c 是大于 0 的实数. 则 $\mathbb{C}^n$ 中的点集

$$\left\{ x \mid x \in \mathbb{C}^n, x^{\mathrm{H}} S x = c \right\}$$

称为 $\mathbb{C}^n$ 中的一个椭球面.

$$\left\{x \mid x \in \mathbb{C}^n, x^{\mathrm{H}} S x \le c\right\}$$
称为椭球体.

定理 1.6.7 (线性映射把椭球映射成椭球)
设 A 是 n 阶非奇异矩阵. 则任一椭球面在映射 $x \mapsto y = Ax$ 下的像也是椭球.
证明: 椭球
$$\left\{x \mid x \in \mathbb{C}^n, x^{\mathrm{H}} S x = c\right\}$$
的像为
$$\begin{aligned} &\left\{y \mid y = Ax, x \in \mathbb{C}^n, x^{\mathrm{H}} S x = c\right\} \\ &= \left\{y \mid y \in \mathbb{C}^n, (A^{-1}y)^{\mathrm{H}} S (A^{-1}y) = c\right\} \\ &= \left\{y \mid y \in \mathbb{C}^n, y^{\mathrm{H}}\left((A^{-1})^{\mathrm{H}} S A^{-1}\right) y = c\right\}. \end{aligned}$$
这是相应于 Hermit 正定矩阵 $(A^{-1})^{\mathrm{H}} S A^{-1}$ 的椭球.

设 A 是 n 阶非奇异矩阵, 记 A 的奇异值分解为
$$U^{\mathrm{H}} A V = \Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}.$$
现计算单位球
$$\left\{x \mid x \in \mathbb{C}^n, \|x\| = 1\right\}$$
在 A 下的像. 由
$$\|x\|^2 = \left\|A^{-1}y\right\|^2 = \left\|(U\Sigma V^{\mathrm{H}})^{-1} y\right\|^2 = \left\|V\Sigma^{-1}(U^{\mathrm{H}} y)\right\|^2 = \left\|\Sigma^{-1}\widetilde{y}\right\|^2$$
$$= \frac{|\widetilde{y}_1|^2}{\sigma_1{}^2} + \frac{|\widetilde{y}_2|^2}{\sigma_2{}^2} + \cdots + \frac{|\widetilde{y}_n|^2}{\sigma_n{}^2},$$
这里 $\widetilde{y} = U^{\mathrm{H}} y$, 即 $y = U\widetilde{y}$. 也就是说, 在新的标准正交基 U 下, 单位球的像的方程是标准的椭球面方程
$$\frac{|\widetilde{y}_1|^2}{\sigma_1{}^2} + \frac{|\widetilde{y}_2|^2}{\sigma_2{}^2} + \cdots + \frac{|\widetilde{y}_n|^2}{\sigma_n{}^2} = 1 .(1.6.3.1)$$
由于
$$U\Sigma = AV, \quad U = AV\Sigma^{-1} = \begin{bmatrix} Av_1 \frac{1}{\sigma_1} & Av_2 \frac{1}{\sigma_2} & \cdots & Av_n \frac{1}{\sigma_n} \end{bmatrix},$$
可知,
$$\sigma_i = \|Av_i\|, \quad i = 1,2,\ldots,n.$$
换言之, 标准正交出口基 $U = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$ 就是标准正交入口基 $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ 在 A 下的像的单位化.
因此, 若 σ_1 是最大奇异值, 则 v_1 就是单位球面上被 A 映的最远的向量: 映射成到椭球面(1.6.3.1)最长的半轴 $Av_1 = u_1\sigma_1$.
关于最大奇异值的这个极值解释亦可由定理 1.6.5 直接得到. 其他奇异值极值解释也可由定理 1.6.5 得到.