欧氏空间

黄利兵

数学科学学院

2023 年 4 月 24 日

主要内容

导言

- 这一章讨论实内积空间, 即赋予内积的实线性空间. 有限维的实线性空间也称为欧氏空间. 在欧氏空间中, 可以讨论角度、距离等度量概念. 所以, 欧氏空间是几何背景非常明确的一种对象, 它可以看成是通常的平面、三维立体空间等在高维的推广.
- 在欧氏空间中, 可以取两两正交的单位向量构成的一组基 (称为标准正交基). 要讨论线性变换的度量性质, 最自然的方式就是考虑该线性变换在这样一组基下的矩阵.
- 在欧氏空间中, 可以讨论一大类具有特殊性质的线性变换, 如保持长度的正交变换, 特征值全为实数的对称变换等. 它们都是正规变换的特例.
- 矩阵与几何现象的对应仍然是本章的一个重要方面. 例如, 获取标准正交基的 Schmidt 正交化方法, 对应于可逆矩阵的 QR 分解; 正交投影, 对应于最小二乘法; 等等. 这些方法都有广泛的应用.

内积的定义

在中学我们已接触过几何向量的内积运算. 向量的长度与夹角等度量性质都可以通过向量的内积来表示. 在抽象的讨论中, 我们取内积作为基本的概念.

定义

设 V 是实数域 $\mathbb{R}$ 上的线性空间. V 上的二元实值函数 $(\ \cdot\ ,\ \cdot\ ) : V \times V \to \mathbb{R}$ 如果满足

- 对称性: $(\alpha, \beta) = (\beta, \alpha)$, $\forall \alpha, \beta \in V$;
- 线性性: $(k_1\alpha_1 + k_2\alpha_2, \beta) = k_1(\alpha_1, \beta) + k_2(\alpha_2, \beta)$,
  $$\forall k_1, k_2 \in \mathbb{R},\ \ \alpha_1, \alpha_2, \beta \in V;$$
- 正定性: $(\alpha, \alpha) \geq 0$, 且等号成立当且仅当 $\alpha = 0$,

则称 (α, β) 是 V 上的一个内积. 定义了内积的实线性空间称为实内积空间. 有限维的实内积空间也称为 Euclid 空间, 或欧氏空间. 无穷维的实内积空间也称为 Hilbert 空间.

注

第二个条件是说, 对每个固定的 β, 映射 $\alpha \mapsto (\alpha, \beta)$ 是线性的, 即该二元函数关于第一个分量是线性的. 结合对称性可知, 它关于第二个分量也是线性的.

欧氏空间的例子 (一)

几何向量间的内积自然满足上述三条性质, 所以三维立体空间中全体几何向量构成一个 Euclid 空间.

例

在 $\mathbb{R}^n = \mathbb{R}^{1\times n}$ 中, 对于任意

$$\alpha = (a_1, a_2, \cdots, a_n), \quad \beta = (b_1, b_2, \cdots, b_n),$$

定义

$$(\alpha, \beta) = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

可以验证这是一个内积, 称为 $\mathbb{R}^n$ 上的标准内积, $\mathbb{R}^n$ 对于这个内积为 Euclid 空间. 以后提到 $\mathbb{R}^n$ 时都是采用这个内积.

上面这个内积可用矩阵记号写为 $(\alpha, \beta) = \alpha\beta^{\mathsf{T}}$, 它可推广为

例

在实线性空间 $\mathbb{R}^{m\times n}$ 中定义 $(A, B) = \mathrm{tr}(AB^{\mathsf{T}})$, 容易验证这是 $\mathbb{R}^{m\times n}$ 上的内积. $\mathbb{R}^{m\times n}$ 对于这个内积成为一个 mn 维 Euclid 空间.

欧氏空间的例子 (二)

例

在闭区间 $[a, b]$ 上连续函数所构成的线性空间 $C[a, b]$ 中, 定义

$$(f, g) = \int_a^b f(x)g(x)\,\mathrm{d}x.$$

容易验证它满足对称性、线性性以及正定性. 因此, $C[a, b]$ 对于这个内积成为一个 Hilbert 空间.

例

考虑所有使得 $\sum_{i=1}^{\infty} a_i^2$ 收敛的无穷实数列 $(a_1, a_2, \cdots)$ 所构成的线性空间 H. 对于 H 中任意两个向量 $a = (a_1, a_2, \cdots)$, $\quad b = (b_1, b_2, \cdots)$, 定义

$$(a, b) = \sum_{i=1}^{\infty} a_i b_i.$$

容易验证这是 H 上的一个内积.

Cauchy-Buniakowski 不等式

定义

非负实数 $\sqrt{(\alpha,\alpha)}$ 称为向量 α 的长度, 记作 $|\alpha|$.

显然零向量的长度为 0, 非零向量的长度 >0. 由内积关于两个分量的线性性容易看出, $|k\alpha| = |k|\,|\alpha|$, $k \in \mathbb{R}$, $\alpha \in V$. 因此, 当 $\alpha \neq 0$ 时, $\frac{1}{|\alpha|}\alpha$ 是一个单位向量 (即长度为 1 的向量). 把 α 变成 $\frac{1}{|\alpha|}\alpha$ 称为把 α 单位化.

定理 (Cauchy-Buniakowski)

在实内积空间 V 中, 对于任意向量 α, β, 有

$$|(\alpha,\beta)| \leq |\alpha|\,|\beta|,$$

等号成立当且仅当 α, β 线性相关.

证明.

若 $\beta = 0$, 则两端均为 0. 若 $\beta \neq 0$, 则由 $(\alpha + x\beta, \alpha + x\beta) \geq 0$ 可知

$$(\beta, \beta)x^2 + 2(\alpha, \beta)x + (\alpha, \alpha) \geq 0.$$

于是判别式 ≤ 0, 即 $(\alpha, \beta)^2 \leq (\alpha, \alpha)(\beta, \beta)$.
等号成立当且仅当前面关于 x 的一元二次函数最小值为零, 即存在某个 x_0 使它等于 0, 这时 $\alpha + x_0\beta = 0$. □

将这个定理分别应用于前面的两个例子, 分别得到

例 (Cauchy 不等式)

对于任意两组实数 $a_1, \cdots, a_n$ 与 $b_1, \cdots, b_n$, 有

$$|a_1 b_1 + \cdots + a_n b_n| \leq \sqrt{a_1^2 + \cdots + a_n^2}\sqrt{b_1^2 + \cdots + b_n^2}.$$

例 (Schwarz 不等式)

对于任意 $f, g \in C[a,b]$, 有

$$\left|\int_a^b f(x)g(x)\,\mathrm{d}x\right| \le \left(\int_a^b f(x)^2\,\mathrm{d}x\right)^{\frac{1}{2}}\left(\int_a^b g(x)^2\,\mathrm{d}x\right)^{\frac{1}{2}}.$$

由于 $|(\alpha,\beta)| \le |\alpha|\,|\beta|$, 下面这个定义是合理的.

定义

在实内积空间 V 中, 非零向量 α, β 的夹角定义为 $\arccos\frac{(\alpha,\beta)}{|\alpha|\,|\beta|}$. 因此, 夹角的范围 $\in [0,\pi]$. 特别地, 当 $(\alpha,\beta)=0$ 时, α 与 β 的夹角为 $\pi/2$, 这时我们称 α 与 β 正交, 记作 $\alpha \perp \beta$. 由于 $(0,\beta)=0$, 约定 0 与任意向量正交.

思考题

(***) 取定 Euclid 空间 V 的一组基 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 和 n 个实数 $c_1, c_2, \cdots, c_n$. 证明: 存在唯一的向量 $\alpha \in V$, 使得

$$(\alpha,\alpha_i) = c_i, \quad 1 \le i \le n.$$

与 Cauchy-Buniakowski 不等式等价, 有下面的三角形不等式.

定理

在实内积空间中, 对任意向量 α, β, 有 $|\alpha+\beta| \le |\alpha| + |\beta|$.

证明.

注意

$$\begin{aligned} |\alpha+\beta|^2 &= (\alpha+\beta, \alpha+\beta) = |\alpha|^2 + 2(\alpha,\beta) + |\beta|^2 \\ &\le |\alpha|^2 + 2|\alpha|\,|\beta| + |\beta|^2 = (|\alpha| + |\beta|)^2, \end{aligned}$$

就得到 $|\alpha+\beta| \le |\alpha| + |\beta|$. □

利用上述证明中第一行的等式, 我们有如下

推论 (勾股定理)

在实内积空间中, 若向量 α 与 β 正交, 则 $|\alpha+\beta|^2 = |\alpha|^2 + |\beta|^2$.

利用数学归纳法, 还可将它推广为: 若向量 α_1, α_2, $\cdots$, α_r 两两正交, 则

$$|\alpha_1 + \alpha_2 + \cdots + \alpha_r|^2 = |\alpha_1|^2 + |\alpha_2|^2 + \cdots + |\alpha_r|^2.$$

距离

定义

在非空集合 E 上定义的二元函数 $d: E \times E \to \mathbb{R}$ 如果满足

- 对称性: $d(x, y) = d(y, x)$, $\forall x, y \in E$;
- 正定性: $d(x, y) \geq 0$, 且等号成立当且仅当 $x = y$;
- 三角形不等式: $d(x, z) \leq d(x, y) + d(y, z)$, $\forall x, y, z \in E$;

则称 d 是一个距离. 把 $d(x, y)$ 称为 x 与 y 之间的距离. 赋予距离的集合称为一个度量空间.

在实内积空间 V 中, 对于任意向量 α, β, 定义

$$d(\alpha, \beta) = |\alpha - \beta|.$$

不难验证它是一个距离.

思考题

(**) 如果定义 $d(\alpha, \beta) = \sqrt{|\alpha - \beta|}$, 它是否还是距离?

度量矩阵

设 V 是 n 维 Euclid 空间. 在 V 中取一组基 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$, 则对 V 中任意两个向量 $\alpha = a_1\varepsilon_1 + a_2\varepsilon_2 + \cdots + a_n\varepsilon_n, \quad \beta = b_1\varepsilon_1 + b_2\varepsilon_2 + \cdots + b_n\varepsilon_n$,
由内积的线性性质得

$$\begin{aligned}(\alpha, \beta) &= (a_1\varepsilon_1 + a_2\varepsilon_2 + \cdots + a_n\varepsilon_n, b_1\varepsilon_1 + b_2\varepsilon_2 + \cdots + b_n\varepsilon_n) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}(\varepsilon_i, \varepsilon_j)a_ib_j.\end{aligned}$$

定义

若 $\varepsilon_1, \cdots, \varepsilon_n$ 是 Euclid 空间 V 的一组基, 令 $a_{ij} = (\varepsilon_i, \varepsilon_j)$, $1 \leq i, j \leq n$, 我们称 $n \times n$ 矩阵 $A = (a_{ij})$ 为基 $\varepsilon_1, \cdots, \varepsilon_n$ 的度量矩阵.

例

在 $\mathbb{R}[x]_n$ 上定义内积 $(f, g) = \int_0^1 f(x)g(x)\,\mathrm{d}x$. 由于 $(x^{i-1}, x^{j-1}) = 1/(i+j-1)$, 所以基 $1, x, \cdots, x^{n-1}$ 的度量矩阵为 $\left(\frac{1}{i+j-1}\right)$.

正定矩阵

在 Euclid 空间 V 中, 设基 $\varepsilon_1, \cdots, \varepsilon_n$ 的度量矩阵为 A, 向量 α, β 在这组基下的坐标分别为

$$a = (a_1, \cdots, a_n)^{\mathsf{T}}, \quad b = (b_1, \cdots, b_n)^{\mathsf{T}}.$$

那么, 前面的讨论告诉我们, 内积 (α, β) 可用矩阵乘法表示

$$(\alpha, \beta) = a^{\mathsf{T}} A b.$$

因而, 度量矩阵完全确定了内积.

内积的性质也可翻译为度量矩阵的性质.

定义

设 A 为 n 阶实矩阵, 如果它满足

- A 是对称矩阵, 即 $A^{\mathsf{T}} = A$;
- 对任意非零列向量 $x \in \mathbb{R}^{n\times 1}$, 有 $x^{\mathsf{T}} A x > 0$;

则称 A 为 n 阶正定矩阵.

由内积的性质容易看出

命题

度量矩阵是正定的.

例

前面提到的矩阵 $\left(\frac{1}{i+j-1}\right)$ 是正定的. 因此, 对任意实数 $x_1, \cdots, x_n$, 有

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{x_i x_j}{i+j-1} \geq 0.$$

思考题

- (****) 设 $a_1, \cdots, a_n$ 为互不相同的正实数, 证明: 矩阵 $\left(\frac{1}{a_i+a_j}\right)$ 是正定矩阵.
- (*****) 设 $a_1, \cdots, a_n$ 为互不相同的正实数, 且 $p > 0$. 证明: 矩阵 $\left(\frac{1}{(a_i+a_j)^p}\right)$ 是正定矩阵.

矩阵的合同

现在，设基 $\varepsilon_1, \cdots, \varepsilon_n$ 的度量矩阵为 A. 另取新的一组基 $\eta_1, \cdots, \eta_n$, 设从原来的基到这组基的过渡矩阵为 $T = (t_1, \cdots, t_n)$, 即 η_i 在基 $\varepsilon_1, \cdots, \varepsilon_n$ 下的坐标为 t_i, $1 \leq i \leq n$. 那么,

$$(\eta_i, \eta_j) = t_i^{\mathsf{T}} A t_j.$$

可见基 $\eta_1, \cdots, \eta_n$ 的度量矩阵为 $T^{\mathsf{T}} A T$.

定义

设 $A, B \in \mathbb{R}^{n\times n}$. 如果存在可逆矩阵 T 使得 $B = T^{\mathsf{T}} A T$, 则称 A 与 B 合同.

思考题

(**) 合同是不是矩阵之间的等价关系?

上面的讨论告诉我们:

命题

Euclid 空间中不同基的度量矩阵是合同的.

标准正交基

自然的问题是: 怎样的基的度量矩阵最简单?

定义

实内积空间 V 中一组非零向量, 如果其中任意两个都正交, 就称为一个 正交向量组. n 维 Euclid 空间中, 由 n 个向量构成的正交向量组称为 V 的正交基. 在一组正交基中, 如果每个向量都是单位向量, 则称为标准正交基.

由定义可以看出, 正交基的度量矩阵是对角矩阵, 标准正交基的度量矩阵是单位矩阵. 因此, 在标准正交基下, 内积的坐标运算非常简单. 如何找标准正交基?

引理

若 $\alpha_1, \cdots, \alpha_k$ 是 Euclid 空间 V 中的正交向量组, 则它们线性无关.

证明.

若 $x_1\alpha_1 + \cdots + x_k\alpha_k = 0$, 则两端与 α_i 作内积得 $(x_i\alpha_i, \alpha_i) = 0$, 可见 $x_i = 0$, $1 \le i \le k$. $\square$

Schmidt 正交化

下面我们将证明, 可以从任意一组基出发, 采用递推的方法生成一组标准正交基.

定理

设 α_1, α_2, $\cdots$, α_n 是 Euclid 空间 V 的一组基, 则在 V 中有标准正交基 ε_1, ε_2, $\cdots$, ε_n, 使得

$$L(\alpha_1, \alpha_2, \cdots, \alpha_k) = L(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_k), \quad 1 \leq k \leq n.$$

证明.

先构造正交基 η_1, $\cdots$, η_n. 令 $\eta_1 = \alpha_1$, 则 $L(\eta_1) = L(\alpha_1)$. 假设已经得到了正交向量组 η_1, $\cdots$, η_k, 使得 $L(\eta_1, \cdots, \eta_k) = L(\alpha_1, \cdots, \alpha_k)$, 我们令

$$\eta_{k+1} = \alpha_{k+1} - \frac{(\alpha_{k+1}, \eta_1)}{(\eta_1, \eta_1)}\eta_1 - \frac{(\alpha_{k+1}, \eta_2)}{(\eta_2, \eta_2)}\eta_2 - \cdots - \frac{(\alpha_{k+1}, \eta_k)}{(\eta_k, \eta_k)}\eta_k.$$

下面验证 η_{k+1} 与 η_i 正交 $(1 \leq i \leq k)$. 事实上

证明 (续).

$$(\eta_{k+1},\eta_i)=(\alpha_{k+1},\eta_i)-\frac{(\alpha_{k+1},\eta_i)}{(\eta_i,\eta_i)}(\eta_i,\eta_i)=0.$$

因此, η_1, η_2, $\cdots$, η_{k+1} 是正交向量组, 从而线性无关. 由 η_{k+1} 的构造可知

$$L(\eta_1,\cdots,\eta_k,\eta_{k+1})=L(\alpha_1,\cdots,\alpha_k,\alpha_{k+1}).$$

再单位化, 即令 $\varepsilon_i=\frac{1}{|\eta_i|}\eta_i$, $1\le i\le n$, 则 ε_1, $\cdots$, ε_n 是符合要求的标准正交基. 上述构造标准正交基的过程称为 Schmidt 正交化. □

推论

任何一个正交向量组可扩充为一组正交基.

证明.

先将该向量组扩充为一组基, 再应用 Schmidt 正交化方法. 容易验证, 原来的正交向量组在此过程中保持不变. □

推论

若基 $\alpha_1, \cdots, \alpha_n$ 的度量矩阵为 A, 则存在可逆上三角矩阵 T, 使得 $T^{\mathrm{T}} A T = E_n$.

证明.

对基 $\alpha_1, \cdots, \alpha_n$ 应用 Schmidt 正交化方法, 得到标准正交基 $\varepsilon_1, \cdots, \varepsilon_n$. 由

$$L(\varepsilon_1, \cdots, \varepsilon_k) = L(\alpha_1, \cdots, \alpha_k), \quad 1 \le k \le n$$

可知 ε_k 可被 $\alpha_1, \cdots, \alpha_k$ 线性表出. 因此两组基之间的过渡矩阵 T 为上三角矩阵. 利用两组基的度量矩阵之间的关系即得 $T^{\mathrm{T}} A T = E_n$. □

推论

若 $A \in \mathbb{R}^{n\times n}$ 是正定矩阵, 则存在可逆上三角矩阵 T, 使得 $A = T^{\mathrm{T}} T$.

事实上, 还可要求这里的上三角矩阵 T 的对角元全为正实数.

例

将 $\mathbb{R}^4$ 的基 $\alpha_1 = (1,1,1,1)$, $\alpha_2 = (2,2,0,0)$, $\alpha_3 = (3,1,2,0)$, $\alpha_4 = (3,1,0,2)$ 变成标准正交基.

解答.

先正交化, 依次得到

$$\begin{aligned}
\eta_1 &= \alpha_1 = (1,1,1,1),\\
\eta_2 &= \alpha_2 - \frac{(\alpha_2,\eta_1)}{(\eta_1,\eta_1)}\eta_1\\
&= (1,1,-1,-1),\\
\eta_3 &= \alpha_3 - \frac{(\alpha_3,\eta_1)}{(\eta_1,\eta_1)}\eta_1 - \frac{(\alpha_3,\eta_2)}{(\eta_2,\eta_2)}\eta_2\\
&= (1,-1,1,-1),\\
\eta_4 &= \alpha_4 - \frac{(\alpha_4,\eta_1)}{(\eta_1,\eta_1)}\eta_1 - \frac{(\alpha_4,\eta_2)}{(\eta_2,\eta_2)}\eta_2 - \frac{(\alpha_4,\eta_3)}{(\eta_3,\eta_3)}\eta_3\\
&= (1,-1,-1,1).
\end{aligned}$$

解答 (续).

再单位化, 就得到标准正交基

$$\frac{1}{2}(1,1,1,1), \qquad \frac{1}{2}(1,1,-1,-1),$$
$$\frac{1}{2}(1,-1,1,-1), \qquad \frac{1}{2}(1,-1,-1,1).$$

思考题

- (**) 如果在 $\mathbb{R}[x]_4$ 上定义内积 $(f,g)=\int_0^1 f(x)g(x)\,\mathrm{d}x$, 试从基 1, x, x^2, x^3 出发, 利用 Schmidt 正交化方法求出一组标准正交基.
- (***) 若 T 为可逆实方阵, 证明存在正交矩阵 Q 和上三角矩阵 R, 使得 $T=QR$. 将可逆矩阵分解为正交矩阵与上三角矩阵的乘积, 也称为 QR 分解.

正交矩阵

由于标准正交基在 Euclid 空间中占有特殊的地位，有必要讨论从一组标准正交基到另一标准正交基的过渡矩阵. 由于两组标准正交基的度量矩阵都是 E_n, 所以过渡矩阵 T 一定满足 $T^{\mathrm{T}} E_n T = E_n$, 即 $T^{\mathrm{T}} T = E_n$.

定义

设 $Q \in \mathbb{R}^{n\times n}$. 如果 $Q^{\mathrm{T}} Q = E_n$, 则称 Q 为 n 阶 (实) 正交矩阵. 所有 n 阶实正交矩阵的集合记作 $O(n, \mathbb{R})$ 或 $O(n)$.

注

- 单位矩阵 E_n 是正交矩阵.
- 正交矩阵 Q 的逆矩阵就是它的转置，因此它也满足 $QQ^{\mathrm{T}} = E_n$. 这表明正交矩阵 Q 的逆矩阵 Q^{T} 仍是正交矩阵.
- 若 A, B 都是 n 阶正交矩阵，则 AB 也是正交矩阵. 事实上，$(AB)^{\mathrm{T}}(AB) = B^{\mathrm{T}} A^{\mathrm{T}} AB = E_n$.

上面这三条性质表明，$O(n)$ 构成一个群.

如果从分块矩阵的角度观察正交矩阵, 还可得到

- 若正交矩阵 Q 的第 i 行为 $e_i \in \mathbb{R}^{1\times n}$, $1 \le i \le n$, 则由 $QQ^{\mathrm{T}} = E_n$ 可知 $e_i e_j^{\mathrm{T}} = \delta_{ij}$, 所以 $e_1, e_2, \cdots, e_n$ 构成 $\mathbb{R}^{1\times n}$ 的标准正交基.
- 同理, 正交矩阵 Q 的 n 个列向量构成 $\mathbb{R}^{n\times 1}$ 的标准正交基.

例

当 θ 为实数时, $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ 与 $B = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$ 都是 2 阶正交矩阵.

例

矩阵 $A = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$ 和 $B = \frac{1}{7}\begin{bmatrix} 2 & 6 & 3 \\ 3 & 2 & -6 \\ 6 & -3 & 2 \end{bmatrix}$ 都是 3 阶正交矩阵.

由定义可知, 从标准正交基到标准正交基的过渡矩阵为正交矩阵.

欧氏空间的同构

同一个实线性空间可以赋予不同的内积, 从而成为不同的实内积空间. 不同的实线性空间更能得到不同的内积空间. 那么, 哪些实内积空间在本质上是一样的?

定义

设 V_1 和 V_2 是实内积空间. 如果存在线性空间同构 $\sigma : V_1 \to V_2$, 使得

$$(\sigma(\alpha), \sigma(\beta)) = (\alpha, \beta), \quad \forall \alpha, \beta \in V_1,$$

则称 V_1 与 V_2 同构, 并称 σ 为 V_1 到 V_2 的一个同构映射.

Euclid 空间之间的同构映射, 就是线性空间之间的同构映射, 并且保持内积, 因此一定把标准正交基变为标准正交基.
设 V 是 n 维 Euclid 空间, 在 V 中取一组标准正交基 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$. 考虑这样的映射 $\sigma : V \to \mathbb{R}^{n\times 1}$, 它把每个向量 α 映为它在这组基下的坐标 a. 则 σ 是从 V 到 $\mathbb{R}^{n\times 1}$ 的线性同构.

进一步, 当 α, β 的坐标分别为 a, b 时, 由于基 $\varepsilon_1, \cdots, \varepsilon_n$ 的度量矩阵为 E_n, 我们有

$$(\alpha,\beta) = a^{\mathsf{T}} E_n b = a^{\mathsf{T}} b = (\sigma(\alpha), \sigma(\beta)).$$

因此

命题

每个 n 维 Euclid 空间 V 都与 $\mathbb{R}^n$ 同构.

注意同构是 Euclid 空间之间的等价关系, 因为它满足

- 反身性: V 与 V 同构, 因为 id 是同构映射;
- 对称性: 若 V 与 W 同构, 则 W 与 V 同构. 这是因为, 若 $\sigma: V \to W$ 是同构映射, 则 $\sigma^{-1}: W \to V$ 是同构映射;
- 传递性: 若 V 与 U 同构, U 与 W 同构, 则 V 与 W 同构. 事实上, 当 $\sigma: V \to U$ 和 $\tau: U \to W$ 都是同构映射时, $\tau \circ \sigma: V \to W$ 也是同构映射.

由于 n 维 Euclid 空间总同构于 $\mathbb{R}^n$, 所以我们有

定理

两个 Euclid 空间同构的充要条件是它们的维数相同.

子空间

设 V 是一个实内积空间, W 是 V 的线性子空间. 显然 V 上指定的内积 $(\cdot\,,\,\cdot)$ 可以限制到 W 上, 从而 W 对于这个内积也成为一个实内积空间. 此时称 W 是 V 的一个子空间.

设 S 是实内积空间 V 的一个子集, V 中向量 α 如果与 S 中每个向量都正交, 则称 α 与 S 正交, 记作 $\alpha \perp S$.

进一步, 如果子集 S' 中每个向量都与 S 正交, 则称 S' 与 S 正交, 记作 $S' \perp S$.

定义

设 S 是实内积空间 V 的一个非空子集, V 中与 S 正交的所有向量构成的集合称为 S 的正交补, 记作 $S^{\perp}$, 即

$$S^{\perp} = \{\alpha \in V \mid (\alpha, \beta) = 0, \forall \beta \in S\}.$$

例

在 $\mathbb{R}^3$ 中, 若 $S = \{\mathbf{u}, \mathbf{v}\}$ 且向量 $\mathbf{u}$ 与 $\mathbf{v}$ 不共线, 则 $S^{\perp}$ 是与平面 $L(\mathbf{u}, \mathbf{v})$ 垂直的直线.

正交补

命题

设 S 是实内积空间 V 的一个非空子集, 则 $S^\perp$ 是 V 的一个子空间.

证明.

任取 $\alpha, \beta \in S^\perp$, 则 $(\alpha,\gamma)=0$, $(\beta,\gamma)=0$ 对任意 $\gamma \in S$ 成立. 从而 $(\alpha+\beta,\gamma)=0$ 也对任意 $\gamma \in S$ 成立. 这表明 $\alpha+\beta \in S^\perp$. 因此 $S^\perp$ 对加法封闭. 同理它也对数乘封闭. □

引理

在实内积空间 V 中, 如果子空间 V_1, V_2 正交, 则它们的和是直和.

证明.

如果 $\alpha \in V_1 \cap V_2$, 则 $(\alpha,\alpha)=0$, 表明 $\alpha=0$. □

如果 S 是实内积空间 V 的一个子空间, 它的正交补 $S^\perp$ 是否是 S 在 V 中的补子空间呢? 换言之, $V=S+S^\perp$ 是否成立呢?

例

考虑所有使得 $\sum_{i=1}^{\infty} a_i^2$ 收敛的无穷实数列 $(a_1, a_2, \cdots)$ 所构成的线性空间 H. 前面我们已定义了这个空间的内积.

在 H 中, 只有有限项非零的那些数列构成一个子空间 S. 容易发现, 与 S 中每个向量都正交的只有零向量, 即 $S^{\perp}$ 为零子空间.

这个例子表明, 如果 S 是无穷维子空间, 则 $V = S + S^{\perp}$ 不一定成立.

定理

如果 S 是实内积空间 V 的有限维子空间, 则 $V = S \oplus S^{\perp}$.

证明.

在 S 中取一个标准正交基 $\varepsilon_1, \cdots, \varepsilon_m$.
任取 $\alpha \in V$, 令

$$\begin{aligned} \alpha_1 &= (\alpha, \varepsilon_1)\varepsilon_1 + \cdots + (\alpha, \varepsilon_m)\varepsilon_m, \\ \alpha_2 &= \alpha - \alpha_1. \end{aligned}$$

证明 (续).

因为当 $1 \leq i \leq m$ 时

$$(\alpha_2, \varepsilon_i) = (\alpha, \varepsilon_i) - (\alpha_1, \varepsilon_i) = 0,$$

所以 α_2 与 S 中任意向量都正交, 即 $\alpha_2 \in S^{\perp}$. 结合 $\alpha_1 \in S$ 以及 $\alpha = \alpha_1 + \alpha_2$ 就证明了 $V = S + S^{\perp}$.
前面的引理保证了这个和是直和. $\square$

后面我们将主要讨论有限维的情况. 对于 n 维 Euclid 空间 V, 如果把子空间 V_1 的正交补记作 V_2, 则有 $V_1 \perp V_2$, 且 $V = V_1 + V_2$. 反之, 满足条件 $V_1 \perp V_2$ 的子空间 V_2 包含在 V_1 的正交补中, 如果它还满足 $V = V_1 + V_2$, 则其维数与 V_1 的正交补相等, 从而这时 V_2 就是 V_1 的正交补.
注意这两个条件关于 V_1, V_2 是对称的, 所以 V_2 的正交补就是 V_1. 换言之, $(V_1^{\perp})^{\perp} = V_1$.

注

如果 S 是无穷维子空间, 则我们能证明 $V = \overline{S} + S^{\perp}$, 其中 $\overline{S}$ 表示 S 的闭包.

正交投影

定义

设 W 是 n 维 Euclid 空间 V 的子空间. V 中任一向量 α 能唯一地表示成

$$\alpha = \alpha_1 + \alpha_2, \quad \alpha_1 \in W, \alpha_2 \in W^{\perp}.$$

称 α_1 为 α 在 W 上的正交投影, 也称为内射影. 从 V 到 W 的映射 $\alpha \mapsto \alpha_1$ 也称为 V 到 W 的正交投影.

如果取 W 的标准正交基 $\varepsilon_1, \cdots, \varepsilon_m$, 则在前面定理的证明中我们已经知道, α 在 W 上的正交投影为

$$\alpha_1 = (\alpha, \varepsilon_1)\varepsilon_1 + \cdots + (\alpha, \varepsilon_m)\varepsilon_m.$$

命题

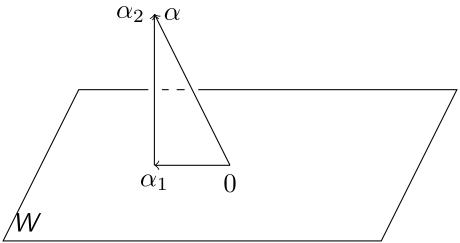
设 W 是 n 维 Euclid 空间 V 的子空间. 固定 $\alpha \in V$, 设 α 在 W 上的正交投影为 α_1, 则 α 与 W 中任意向量 β 的距离的最小值恰在 $\beta = \alpha_1$ 时取到.

证明.

由于 $\alpha_1-\beta\in W$, $\alpha-\alpha_1\in W^\perp$, 我们有 $(\alpha-\alpha_1,\alpha_1-\beta)=0$, 因而

$$|\alpha-\beta|^2=|\alpha-\alpha_1|^2+|\alpha_1-\beta|^2\geq|\alpha-\alpha_1|^2.$$

等号成立当且仅当 $\alpha_1-\beta=0$. □



思考题

(**) Schmidt 正交化与正交投影的关系是什么?

最小二乘法

下面介绍正交投影的一种重要应用：最小二乘法.
在许多实际问题中需要研究一个变量 y 与其他一些变量 $t_1, t_2, \cdots, t_n$ 之间的依赖关系. 经过实际观测和分析, 假设 y 和 $t_1, t_2, \cdots, t_n$ 之间呈线性关系

$$y = x_1 t_1 + x_2 t_2 + \cdots + x_n t_n.$$

为了确定出其中的系数 $x_1, \cdots, x_n$, 需要做多次观测. 每次观测所得的数据都导致系数 $x_1, \cdots, x_n$ 满足一个线性方程. 于是 m 组观测数据就导致线性方程组 $A\mathbf{x} = \beta$, 其中 $A \in \mathbb{R}^{m\times n}$, $\beta \in \mathbb{R}^{m\times 1}$, $\mathbf{x} = (x_1, \cdots, x_n)^{\mathrm{T}}$. 由于测量误差或其他原因, 这个方程组可能无解. 即任意 $\mathbf{x}$ 都不能使向量 $A\mathbf{x} - \beta$ 为零. 这时, 怎样找到一组合理的 "解" 呢?

定义

对于线性方程组 $A\mathbf{x} = \beta$, 如果当 $\mathbf{x} = \mathbf{x}_0$ 时, $|A\mathbf{x} - \beta|^2$ 取到最小值, 则称 $\mathbf{x}_0$ 为该方程组的最小二乘解. 求最小二乘解的问题称为最小二乘法问题.

命题

向量 $\mathbf{x}_0$ 是方程组 $A\mathbf{x} = \beta$ 的最小二乘解, 当且仅当它是线性方程组 $A^{\mathsf{T}}A\mathbf{x} = A^{\mathsf{T}}\beta$ 的解.

证明.

设 A 的列向量为 $\alpha_1, \cdots, \alpha_n$, 它们张成的子空间为 W. 注意 $A\mathbf{x} = x_1\alpha_1 + \cdots + x_n\alpha_n \in W$. 要求 $|A\mathbf{x} - \beta|^2$ 的最小值, 相当于在 W 中找离 β 距离最近的元素. 由前面的讨论, 这个距离最近的元素 $A\mathbf{x}_0$ 就是 β 在 W 上的正交投影, 即 $A\mathbf{x}_0 - \beta \in W^{\perp}$.

注意 $A\mathbf{x}_0 - \beta \in W^{\perp}$ 当且仅当

$$(A\mathbf{x}_0 - \beta, \alpha_i) = 0, \quad 1 \leq i \leq n.$$

也即 $\alpha_i^{\mathsf{T}}(A\mathbf{x}_0 - \beta) = 0$, $1 \leq i \leq n$. 它进一步等价于 $A^{\mathsf{T}}(A\mathbf{x}_0 - \beta) = 0$. □

思考题

(***) 上面的证明中已蕴含了如下事实: 对任意 $A \in \mathbb{R}^{m\times n}$, $\beta \in \mathbb{R}^{m\times 1}$, 线性方程组 $A^{\mathsf{T}}A\mathbf{x} = A^{\mathsf{T}}\beta$ 总有解. 你能用线性方程组的理论给出一个直接的证明吗?

共轭变换

引理 (表示引理)

设 V 为 n 维 Euclid 空间. 对任意线性映射 $f: V \to \mathbb{R}$, 存在唯一的向量 $\xi \in V$, 使得 $f(\alpha) = (\xi, \alpha)$ 对任意 $\alpha \in V$ 成立.

证明.

取 V 的标准正交基 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$. 设 $f(\varepsilon_i) = t_i$, $1 \le i \le n$, 则易证

$$\xi = t_1\varepsilon_1 + t_2\varepsilon_2 + \cdots + t_n\varepsilon_n$$

是满足要求的唯一向量. □

现在设 $\mathcal{A}$ 是 n 维 Euclid 空间 V 上的线性变换. 对每个固定的向量 β, 从 V 到 $\mathbb{R}$ 的映射 $\alpha \mapsto (\mathcal{A}\alpha, \beta)$ 是线性映射, 因而由表示引理可知, 存在唯一的向量 $\mathcal{A}^*\beta$, 使得

$$(\mathcal{A}\alpha, \beta) = (\alpha, \mathcal{A}^*\beta), \quad \forall \alpha \in V.$$

将 β 变为 $\mathcal{A}^*\beta$ 这个变换, 称为 $\mathcal{A}$ 的 共轭变换, 记作 $\mathcal{A}^*$.
注意等式 $(\mathcal{A}\alpha, \beta) = (\alpha, \mathcal{A}^*\beta)$ 的左端关于 β 是线性的, 易知 $\mathcal{A}^*$ 是线性变换.

共轭变换的矩阵

设 $\mathcal{A}$ 是 n 维 Euclid 空间 V 上的线性变换, $\mathcal{A}^*$ 是它的共轭变换. 任取 V 的一组基 $\alpha_1, \cdots, \alpha_n$, 设这组基的度量矩阵为 $G = (g_{ij})$, 即 $g_{ij} = (\alpha_i, \alpha_j)$, $1 \leq i, j \leq n$. 又设向量 α, β 在这组基下的坐标分别为 a, b, 变换 $\mathcal{A}$ 和 $\mathcal{A}^*$ 在这组基下的矩阵分别为 A, B, 那么

$$(\mathcal{A}\alpha, \beta) = (Aa)^{\mathsf{T}} Gb = a^{\mathsf{T}} A^{\mathsf{T}} Gb,$$
$$(\alpha, \mathcal{A}^*\beta) = a^{\mathsf{T}} GBb.$$

要使 $(\mathcal{A}\alpha, \beta) = (\alpha, \mathcal{A}^*\beta)$ 对任意 α, β 成立, 需要 $a^{\mathsf{T}} A^{\mathsf{T}} Gb = a^{\mathsf{T}} GBb$ 对任意 a, $b \in \mathbb{R}^{n\times 1}$ 成立, 即 $A^{\mathsf{T}} G = GB$. 因此,

$$B = G^{-1} A^{\mathsf{T}} G.$$

特别地, 取标准正交基时, 度量矩阵 $G = E_n$, 这时 $\mathcal{A}^*$ 与 $\mathcal{A}$ 的矩阵互为转置.

共轭变换的性质

利用矩阵与线性变换之间的对应关系，可知共轭变换有以下性质

- $(\mathcal{A}+\mathcal{B})^* = \mathcal{A}^* + \mathcal{B}^*$;
- $(k\mathcal{A})^* = k\mathcal{A}^*$;
- $(\mathcal{A}\mathcal{B})^* = \mathcal{B}^*\mathcal{A}^*$;
- $(\mathcal{A}^*)^* = \mathcal{A}$;
- 当 $\mathcal{A}$ 可逆时，$\mathcal{A}^*$ 也可逆，且 $(\mathcal{A}^*)^{-1} = (\mathcal{A}^{-1})^*$.

命题

在 n 维 Euclid 空间 V 中，线性变换 $\mathcal{A}$ 的共轭变换是 $\mathcal{A}^*$，那么，$\ker \mathcal{A}^*$ 的正交补恰好是 $\mathcal{A}V$.

证明.

由定义，$(\mathcal{A}V)^\perp = \{\beta \in V \,|\, (\mathcal{A}\alpha, \beta) = 0, \forall \alpha \in V\}$. 其中 $(\mathcal{A}\alpha, \beta) = (\alpha, \mathcal{A}^*\beta)$，所以我们有

$$(\mathcal{A}V)^\perp = \{\beta \in V \,|\, (\alpha, \mathcal{A}^*\beta) = 0, \forall \alpha \in V\} = \{\beta \in V \,|\, \mathcal{A}^*\beta = 0\},$$

这就证明了 $\mathcal{A}V$ 的正交补是 $\ker \mathcal{A}^*$. □

正规变换

定义

设 $\mathcal{A}$ 是 n 维 Euclid 空间的线性变换, 如果

$$\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A},$$

则称 $\mathcal{A}$ 为正规变换.

定义

设 $A \in \mathbb{R}^{n\times n}$, 如果

$$AA^{\mathsf{T}} = A^{\mathsf{T}}A,$$

则称 A 为正规矩阵.

由线性变换与矩阵的对应关系立即得到

引理

$\mathcal{A}$ 是正规变换当且仅当它在标准正交基下的矩阵是正规矩阵.

正规变换的重要特点是, 每个不变子空间总有不变的补子空间.

命题

设 $\mathcal{A}$ 是 n 维 Euclid 空间 V 的正规变换. 如果 W 是 $\mathcal{A}$ 的不变子空间, 则 $W^{\perp}$ 也是 $\mathcal{A}$ 的不变子空间. 进一步, W 和 $W^{\perp}$ 都是 $\mathcal{A}^*$ 的不变子空间.

证明.

将 W 的标准正交基扩充为整个空间的标准正交基, 则 $\mathcal{A}$ 在这组基下的矩阵为 $A = \begin{bmatrix} A_1 & A_3 \\ 0 & A_2 \end{bmatrix}$, 其中 A_1 是 $\mathcal{A}|_W$ 的矩阵.

由 $A^{\mathsf{T}}A = AA^{\mathsf{T}}$ 可得 $A_1^{\mathsf{T}}A_1 = A_1A_1^{\mathsf{T}} + A_3A_3^{\mathsf{T}}$, 取迹得 $\mathrm{tr}(A_3A_3^{\mathsf{T}}) = 0$. 因而 A_3 中所有元素的平方和为零, $A_3 = 0$. 这表明 $W^{\perp}$ 也是 $\mathcal{A}$ 的不变子空间. 注意 $A^{\mathsf{T}} = \mathrm{diag}(A_1^{\mathsf{T}}, A_2^{\mathsf{T}})$, 就得到 W, $W^{\perp}$ 都是 $\mathcal{A}^*$ 的不变子空间. □

从证明过程中还可得到 $A_1^{\mathsf{T}}A_1 = A_1A_1^{\mathsf{T}}$, 所以

推论

若 $\mathcal{A}$ 是 n 维 Euclid 空间的正规变换, W 是 $\mathcal{A}$ 的不变子空间, 则 $\mathcal{A}|_W$ 是 W 的正规变换, 且 $(\mathcal{A}|_W)^* = \mathcal{A}^*|_W$.

基于上面的结论, 要弄清正规变换的结构, 只需找到维数较低的不变子空间.

引理

设 $\mathcal{A}$ 是 n 维实线性空间 V 上的线性变换, 则 $\mathcal{A}$ 必有 1 维或 2 维不变子空间.

证明.

如果 $\mathcal{A}$ 有实的特征值, 则相应的特征向量张成 1 维的不变子空间. 下面假设 $\mathcal{A}$ 没有实的特征值. 取 $\mathcal{A}$ 在某组基下的矩阵 A. 把 A 看作复矩阵, 则它有虚的特征值 $a+b\sqrt{-1}$, 其中 $a,b\in\mathbb{R}$, 且 $b\neq 0$. 设相应的复特征向量为 $\mathbf{u}+\mathbf{v}\sqrt{-1}$, 其中 $\mathbf{u},\mathbf{v}\in\mathbb{R}^{n\times 1}$, 那么

$$A(\mathbf{u}+\mathbf{v}\sqrt{-1})=(a+b\sqrt{-1})(\mathbf{u}+\mathbf{v}\sqrt{-1}),$$

比较两端的实部和虚部, 可得

$$A\mathbf{u}=a\mathbf{u}-b\mathbf{v},\quad A\mathbf{v}=a\mathbf{v}+b\mathbf{u}.$$

可见, 以 $\mathbf{u}$, $\mathbf{v}$ 为坐标的两个向量所张成的子空间是 $\mathcal{A}$ 的不变子空间. □

平面上的正规变换

引理

设 $\mathcal{A}$ 是平面上的正规变换. 如果 $\mathcal{A}$ 没有实特征根, 则 $\mathcal{A}$ 是旋转和数乘变换的复合, 即它在某组标准正交基下的矩阵为

$$r \cdot \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad r > 0.$$

证明.

设 $\mathcal{A}$ 在某组标准正交基下的矩阵为 $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. 由 $A^{\mathsf{T}}A = AA^{\mathsf{T}}$ 可得

$$a^2 + b^2 = a^2 + c^2, \quad c^2 + d^2 = b^2 + d^2, \quad ac + bd = ab + cd.$$

由于 A 的特征多项式 $\lambda^2 - (a+d)\lambda + (ad - bc)$ 无实根, 其判别式 $(a-d)^2 + 4bc < 0$, 故 b 与 c 异号. 由前两个方程得 $b = -c \neq 0$, 代入第三个方程得 $a = d$. 于是矩阵 A 的复特征值为 $a \pm \sqrt{-1}b$, 它也可写为 $r(\cos\theta \pm \sqrt{-1}\sin\theta)$. □

正规变换的标准形

利用上面的结果, 我们得到

定理

设 $\mathcal{A}$ 为 n 维 Euclid 空间 V 的正规变换, 则 $\mathcal{A}$ 在某组标准正交基下的矩阵为

$$A = \mathrm{diag}(\lambda_1, \cdots, \lambda_r, A_1, \cdots, A_s),$$

其中 $A_i = r_i \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix}$, $r_i > 0$.

证明.

首先, $\mathcal{A}$ 总有 1 维或 2 维的不变子空间 V_1. 这时 $V_1^{\perp}$ 仍是不变子空间且 $\mathcal{A}|_{V_1}$ 仍是正规变换, 于是在 $V_1^{\perp}$ 中又可找到 1 维或 2 维的不变子空间. 如此继续下去, 就可将 V 分解为若干个 1 维或 2 维的不变子空间的正交直和. □

这个结果也可用矩阵来叙述.

推论

实 n 阶方阵 A 为正规矩阵, 当且仅当存在正交矩阵 Q, 使得

$$Q^{\mathsf{T}}AQ = \mathrm{diag}(\lambda_1, \cdots, \lambda_r, A_1, \cdots, A_s),$$

其中 $A_i = k_i \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix}$, $k_i > 0$.

思考题

(****) 设 $A \in \mathbb{R}^{n\times n}$ 的特征多项式的 n 个复数根分别为 $\lambda_1, \lambda_2, \cdots, \lambda_n$. 求证: A 为正规矩阵的充要条件是

$$\mathrm{tr}(AA^{\mathsf{T}}) = |\lambda_1|^2 + |\lambda_2|^2 + \cdots + |\lambda_n|^2.$$

正交变换

现在我们讨论几类常见的正规变换.

定义

设 $\mathcal{A}$ 是 n 维 Euclid 空间 V 上的线性变换. 如果 $\mathcal{A}$ 保持内积, 即

$$(\mathcal{A}\alpha, \mathcal{A}\beta) = (\alpha, \beta), \quad \forall \alpha, \beta \in V,$$

则称 $\mathcal{A}$ 为正交变换.

例

平面上的旋转是正交变换, 关于某条直线的镜面反射也是正交变换. 空间中绕某条轴的旋转是正交变换, 关于某个平面的镜面反射也是正交变换.

思考题

(***) 证明: Euclid 空间 V 上的变换 (不一定线性)$\mathcal{A}$ 如果保持内积, 则它一定是线性变换.

正交变换的刻画

正交变换可以从不同的角度来刻画.

定理

设 $\mathcal{A}$ 是 n 维 Euclid 空间 V 上的线性变换, 则以下条件等价:

(1) $\mathcal{A}$ 保持向量长度, 即 $|\mathcal{A}\alpha| = |\alpha|, \forall \alpha \in V$;

(2) $\mathcal{A}$ 保持内积, 即 $\mathcal{A}$ 是正交变换;

(3) $\mathcal{A}$ 将标准正交基变为标准正交基;

(4) $\mathcal{A}$ 在标准正交基下的矩阵是正交矩阵;

(5) $\mathcal{A}$ 可逆, 且 $\mathcal{A}^{-1} = \mathcal{A}^*$;

证明.

(1)$\Longrightarrow$(2). 极化恒等式

$$(\alpha, \beta) = \frac{1}{4}(|\alpha + \beta|^2 - |\alpha - \beta|^2)$$

表明, 内积可以用长度表示. 因此, 当线性变换 $\mathcal{A}$ 保持长度时, 它也保持内积.

证明 (续).

(2)$\Longrightarrow$(3). 如果 $\varepsilon_1, \cdots, \varepsilon_n$ 是标准正交基, 即 $(\varepsilon_i, \varepsilon_j) = \delta_{ij}$, 那么, 由于 $\mathcal{A}$ 保持内积, 我们有

$$(\mathcal{A}\varepsilon_i, \mathcal{A}\varepsilon_j) = (\varepsilon_i, \varepsilon_j) = \delta_{ij},$$

因此 $\mathcal{A}\varepsilon_1, \cdots, \mathcal{A}\varepsilon_n$ 也是标准正交基.

(3)$\Longrightarrow$(4). 显然.

(4)$\Longrightarrow$(5). 如果 $\mathcal{A}$ 在标准正交基下的矩阵是正交矩阵 A, 那么, A 可逆, 且 $A^{-1} = A^{\mathsf{T}}$. 这也就表明 $\mathcal{A}$ 可逆, 且 $\mathcal{A}^{-1} = \mathcal{A}^*$.

(5)$\Longrightarrow$(1). 利用 $\mathcal{A}^{-1} = \mathcal{A}^*$, 我们有

$$(\mathcal{A}\alpha, \mathcal{A}\alpha) = (\alpha, \mathcal{A}^*\mathcal{A}\alpha) = (\alpha, \alpha),$$

因此 $|\mathcal{A}\alpha| = |\alpha|$. □

思考题

(**) 如果 Euclid 空间 V 上的变换 (不一定是线性变换)$\mathcal{A}$ 保持向量长度, 它是否一定是正交变换?

将 V 上所有正交变换的集合记作 $O(V)$. 注意 $O(V)$ 这个集合具有如下性质:

- 若 $\mathcal{A} \in O(V)$, 则 $\mathcal{A}^{-1} \in O(V)$;
- 若 $\mathcal{A}, \mathcal{B} \in O(V)$, 则 $\mathcal{A}\mathcal{B} \in O(V)$.

这表明 $O(V)$ 构成一个群, 称为 V 上的正交群.
注意正交变换 $\mathcal{A}$ 满足 $\mathcal{A}\mathcal{A}^* = \mathrm{id}$, 所以它一定满足 $\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A}$, 即 $\mathcal{A}$ 是正规的. 为了弄清正交变换的结构, 只需弄清它的复特征值的情况.

引理

若 A 是正交矩阵, 则它的任意复特征值的模长为 1.

证明.

设 λ_0 是 A 的一个复特征值, ξ 是相应的复特征向量. 将等式 $A\xi = \lambda_0\xi$ 取共轭, 再转置, 可得

$$\overline{\xi}^{\mathsf{T}} A^{\mathsf{T}} = \overline{\lambda}_0 \overline{\xi}^{\mathsf{T}}.$$

因此我们有

$$\overline{\xi}^{\mathsf{T}}\xi = \overline{\xi}^{\mathsf{T}} A^{\mathsf{T}} A\xi = |\lambda_0|^2 \overline{\xi}^{\mathsf{T}}\xi,$$

其中 $\overline{\xi}^{\mathsf{T}}\xi \neq 0$, 所以 $|\lambda_0| = 1$. □

正交变换的标准形

利用上面的引理, 结合正规变换的性质, 我们立即得到

命题

若 $\mathcal{A}$ 是 n 维 Euclid 空间 V 上的正交变换, 则存在 V 的标准正交基, 使得 $\mathcal{A}$ 在这组基下的矩阵为

$$\operatorname{diag}(E_r, -E_s, A_1, \cdots, A_t),$$

其中 $A_j = \begin{bmatrix} \cos\theta_j & -\sin\theta_j \\ \sin\theta_j & \cos\theta_j \end{bmatrix}$, $1 \leq j \leq t$.

推论

若 $\mathcal{A}$ 是 n 维 Euclid 空间 V 上的正交变换, 则 $\det(\mathcal{A}) = \pm 1$.

通常将行列式为 1 的正交变换称为旋转或第一类的, 将行列式为 -1 的正交变换称为第二类的.

例

3 维 Euclid 空间中的第一类正交变换, 一定是绕某个轴的旋转.

对称变换

定义

设 $\mathcal{A}$ 是 n 维 Euclid 空间 V 上的线性变换, 如果 $\mathcal{A}^* = \mathcal{A}$, 则称 $\mathcal{A}$ 为对称变换, 也称为自共轭变换.

由共轭变换与矩阵的对应关系立即得到

命题

n 维 Euclid 空间 V 上的线性变换 $\mathcal{A}$ 是对称的, 当且仅当它在标准正交基下的矩阵是对称矩阵.

当 $\mathcal{A}^* = \mathcal{A}$ 时, 自然有 $\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A}$, 所以对称变换都是正规变换. 为了弄清对称变换的结构, 只需弄清它的复特征值.

引理

设 A 为 n 阶实对称矩阵, 则 A 的特征多项式的根全为实数.

证明.

设 λ_0 是特征多项式的一个复数根, 相应的复特征向量为 ξ, 即 $A\xi = \lambda_0\xi$. 于是

$$\overline{\xi}^{\mathsf{T}} A\xi = \lambda_0 \overline{\xi}^{\mathsf{T}} \xi.$$

注意左端取共轭转置仍等于它自身, 因此右端也有此性质, 从而得到 $\overline{\lambda_0} = \lambda_0$, 即 λ_0 为实数. □

利用这个引理, 结合正规变换的性质, 就得到

命题

设 $\mathcal{A}$ 是 n 维 Euclid 空间 V 上的对称变换, 则存在标准正交基, 使得 $\mathcal{A}$ 在这组基下的矩阵为对角矩阵. 进一步, V 可分解为不同特征子空间的正交直和.

推论

设 A 为 n 阶实对称矩阵, 则存在正交矩阵 T, 使得 $T^{\mathsf{T}}AT$ 为对角矩阵.

由前面的命题, 要求出这里的正交矩阵 T, 只需找到各特征子空间的标准正交基.

例

求正交矩阵 T 使得 $T^{\mathsf{T}}AT$ 为对角矩阵, 其中

$$A = \begin{bmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{bmatrix}.$$

解答

容易算得 $|\lambda E - A| = (\lambda - 1)^3(\lambda + 3)$, 因此 A 的特征值为 -3 和 1(三重).

解线性方程组 $(-3E - A)x = 0$, 可得属于 -3 的特征子空间的一组基为 $\alpha = (1, -1, -1, 1)^{\mathsf{T}}$. 单位化之后得到 $\beta = (1/2, -1/2, -1/2, 1/2)^{\mathsf{T}}$.

解线性方程组 $(E - A)x = 0$, 可得属于 1 的特征子空间的一组基为 $\alpha_1 = (1, 1, 0, 0)^{\mathsf{T}}$, $\alpha_2 = (1, 0, 1, 0)^{\mathsf{T}}$, $\alpha_3 = (-1, 0, 0, 1)^{\mathsf{T}}$. 下面从这组基出发, 利用 Schmidt 正交化方法得到标准正交基.

解答 (续)

先正交化

$$\eta_1 = \alpha_1 = (1, 1, 0, 0)^{\mathsf{T}},$$
$$\eta_2 = \alpha_2 - \frac{1}{2}\eta_1 = \frac{1}{2}(1, -1, 2, 0)^{\mathsf{T}},$$
$$\eta_3 = \alpha_3 + \frac{1}{2}\eta_1 + \frac{1}{3}\eta_2 = \frac{1}{3}(-1, 1, 1, 3)^{\mathsf{T}}.$$

再单位化, 就得到标准正交基

$$\beta_1 = \frac{1}{\sqrt{2}}(1, 1, 0, 0)^{\mathsf{T}}, \quad \beta_2 = \frac{1}{\sqrt{6}}(1, -1, 2, 0)^{\mathsf{T}}, \quad \beta_3 = \frac{1}{\sqrt{12}}(-1, 1, 1, 3)^{\mathsf{T}}.$$

因此, 只要取 $T = (\beta, \beta_1, \beta_2, \beta_3)$, 则 $T^{\mathsf{T}}AT$ 为对角矩阵 $\mathrm{diag}(-3, 1, 1, 1)$.

酉空间及其变换

定义

设 V 是复数域上的线性空间. V 上的二元复函数 $(\cdot\,,\,\cdot)$ 如果满足

- 关于第一个分量是线性的: $(k_1\alpha_1 + k_2\alpha_2, \beta) = k_1(\alpha_1, \beta) + k_2(\alpha_2, \beta)$, $\forall k_1, k_2 \in \mathbb{C}$, $\alpha_1, \alpha_2, \beta \in V$;
- Hermite 性: $(\alpha, \beta) = \overline{(\beta, \alpha)}$, $\forall \alpha, \beta \in V$;
- 正定性: (α, α) 为非负实数, 且只有在 $\alpha = 0$ 时有 $(\alpha, \alpha) = 0$;

则称 $(\cdot\,,\,\cdot)$ 为 V 上的一个内积, 具有内积的复线性空间称为复内积空间. 有限维复内积空间也称为酉空间.

例

在 $\mathbb{C}^{m\times n}$ 中定义

$$(A, B) = \mathrm{tr}(A\overline{B}^{\mathrm{T}}),$$

则 $(\cdot\,,\,\cdot)$ 是内积, $\mathbb{C}^{m\times n}$ 成为酉空间.

酉空间与 Euclid 空间在很多方面是类似的, Euclid 空间的大多数结论只需稍作修改就成为酉空间的相应结论, 证明方法也类似 (甚至更简单), 因此, 我们将只列举这些结论, 而不加证明.

- 酉空间的内积对第一个分量是线性的，但它没有对称性，所以对第二个分量不是线性的，而是半线性的，即

$$(\alpha, k_1\beta_1 + k_2\beta_2) = \overline{k_1}(\alpha, \beta_1) + \overline{k_2}(\alpha, \beta_2).$$

- 如果记 $|\alpha| = \sqrt{(\alpha, \alpha)}$ 为向量 α 的长度，则有 $|k\alpha| = |k||\alpha|$，并有如下的 Cauchy-Buniakowski 不等式

$$|(\alpha, \beta)| \leq |\alpha|\,|\beta|,$$

  等号成立当且仅当向量 α, β 线性相关. 在量子力学中，通常将 $\frac{|(\alpha,\beta)|^2}{|\alpha|^2|\beta|^2}$ 解释成概率.

- 注意 $|\alpha + \beta|^2 = |\alpha|^2 + (\alpha, \beta) + \overline{(\alpha, \beta)} + |\beta|^2 \leq |\alpha|^2 + 2|(\alpha, \beta)| + |\beta|^2$，我们仍可由 Cauchy-Buniakowski 不等式得到三角形不等式

$$|\alpha + \beta| \leq |\alpha| + |\beta|.$$

- 在酉空间 V 中，如果 $(\alpha, \beta) = 0$，则称向量 α 与 β 正交，记作 $\alpha \perp \beta$.

- 若酉空间 V 的基 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 满足

$$(\alpha_i, \alpha_j) = \delta_{ij}, \quad 1 \leq i, j \leq n,$$

则称为标准正交基. 在标准正交基下, 坐标分别为 $a, b \in \mathbb{C}^{n\times 1}$ 的两个向量的内积为 $a^{\mathsf{T}}\overline{b}$.

- 从酉空间 V 的任何一组基出发, 可以用 Schmidt 正交化方法得到标准正交基. 从标准正交基到标准正交基的过渡矩阵为酉矩阵, 即满足 $\overline{Q}^{\mathsf{T}} Q = E_n$ 的矩阵 $Q \in \mathbb{C}^{n\times n}$.
- 酉空间 V 的子空间 W 有诱导的内积, 从而自然成为酉空间. 与 W 中所有向量都正交的那些向量构成一个子空间, 称为 W 的正交补, 记作 $W^{\perp}$. 当 W 是有限维子空间时, 有 $V = W \oplus W^{\perp}$.

下面讨论 n 维酉空间 V 上的线性变换.

- 若 $\mathcal{A}$ 是酉空间 V 上的线性变换, 则其共轭变换 $\mathcal{A}^*$ 是满足条件
$$(\mathcal{A}\alpha, \beta) = (\alpha, \mathcal{A}^*\beta), \quad \forall \alpha, \beta \in V$$
的唯一线性变换. 如果 $\mathcal{A}$ 在标准正交基下的矩阵为 A, 则 $\mathcal{A}^*$ 在同一组基下的矩阵为 $\overline{A}^{\mathrm{T}}$.
- 若酉空间 V 上的线性变换 $\mathcal{A}$ 满足 $\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A}$, 则称为正规变换.
若 W 是正规变换 $\mathcal{A}$ 的不变子空间, 则 $W^\perp$ 也是 $\mathcal{A}$ 的不变子空间. 注意 $\mathcal{A}$ 总有特征向量, 因此总有 1 维不变子空间. 于是 V 可分解为 n 个 1 维不变子空间的正交直和, 即 $\mathcal{A}$ 总是可对角化的.
- 正规变换在标准正交基下的矩阵是正规矩阵, 即满足 $A\overline{A}^{\mathrm{T}} = \overline{A}^{\mathrm{T}}A$ 的矩阵 $A \in \mathbb{C}^{n\times n}$. 上面的结论用矩阵的语言来叙述也就是: 若 $A \in \mathbb{C}^{n\times n}$ 为正规矩阵, 则存在酉矩阵 Q, 使得
$$\overline{Q}^{\mathrm{T}}AQ = \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$
其中 $\lambda_1, \cdots, \lambda_n$ 为 A 的特征值. 也就是说, 正规矩阵总酉相似于对角矩阵.

与 Euclid 空间的情形类似, 我们介绍几类特殊的正规变换.

- 保持内积的线性变换称为酉变换. 酉变换 $\mathcal{A}$ 有如下几种等价的刻画:
  - 保持内积: $(\mathcal{A}\alpha, \mathcal{A}\beta) = (\alpha, \beta)$, $\forall \alpha, \beta \in V$;
  - 保持长度: $|\mathcal{A}\alpha| = |\alpha|$, $\forall \alpha \in V$;
  - $\mathcal{A}$ 将标准正交基变为标准正交基;
  - $\mathcal{A}$ 在标准正交基下的矩阵为酉矩阵.

  由于酉变换保持向量长度, 易知它的每个特征值的模长为 1. 这也是酉矩阵的性质.
- 满足 $\mathcal{A} = \mathcal{A}^*$ 的线性变换 $\mathcal{A}$ 称为 Hermite 变换. Hermite 变换在标准正交基下的矩阵为 Hermite 矩阵, 即满足 $A = \overline{A}^T$ 的矩阵 $A \in \mathbb{C}^{n\times n}$. Hermite 变换的特征值全为实数. 这也是 Hermite 矩阵的性质.

思考题

(***) 证明如下的 Schur 分解定理: 任意 n 阶复方阵总酉相似于某个上三角矩阵.

奇异值分解 (一)

当 $m \geq n$ 时, 对于 $A \in \mathbb{R}^{m\times n}$, 由于 $A^\mathsf{T}A$ 是 n 阶对称矩阵, 所以存在 n 阶正交矩阵 V 使得 $V^\mathsf{T}(A^\mathsf{T}A)V$ 为对角矩阵 $\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$.

设 $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$, 则有

$$(A\mathbf{v}_i, A\mathbf{v}_j) = \mathbf{v}_i^\mathsf{T} A^\mathsf{T} A\mathbf{v}_j = \lambda_i \delta_{ij}.$$

特别地, 当 $i = j$ 时, 由上式可得 $\lambda_i = |A\mathbf{v}_i|^2 \geq 0$.

不妨设 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. 令 $\sigma_i = \sqrt{\lambda_i}$, $1 \leq i \leq n$, 称 $\sigma_1, \sigma_2, \cdots, \sigma_n$ 为 A 的奇异值.

由于 $A\mathbf{v}_1, \cdots, A\mathbf{v}_n$ 两两正交, 因此, 存在 $\mathbb{R}^{m\times 1}$ 的标准正交基 $\mathbf{u}_1, \cdots, \mathbf{u}_m$, 使得

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i, \quad 1 \leq i \leq n.$$

如果记 $U = (\mathbf{u}_1, \cdots, \mathbf{u}_m)$, 并记 Σ 为 $m \times n$ 的对角矩阵 $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$, 则上式也可写为矩阵等式

$$AV = U\Sigma.$$

奇异值分解 (二)

定理

若 $A \in \mathbb{R}^{m\times n}$, 则存在 m 阶正交矩阵 U, n 阶正交矩阵 V, 以及对角元非负的 $m \times n$ 对角矩阵 Σ, 使得

$$A = U\Sigma V^{\mathsf{T}},$$

这样的分解称为矩阵 A 的奇异值分解.

如果奇异值中仅有 $\sigma_1, \cdots, \sigma_r$ 不等于零, 则从上述分解可得到

- A 的秩为 r, 并且 $A^{\mathsf{T}}A = V\Sigma^{\mathsf{T}}\Sigma V^{\mathsf{T}}$ 和 $AA^{\mathsf{T}} = U\Sigma\Sigma^{\mathsf{T}}U^{\mathsf{T}}$ 的秩都为 r;
- $A = \sigma_1\mathbf{u}_1\mathbf{v}_1^{\mathsf{T}} + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^{\mathsf{T}}$. 也就是说, 正交矩阵 U, V 中仅有前 r 列是必要的, 其他列对这个分解来说无关紧要.

例

矩阵 $\begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$ 的奇异值分解为

$$\begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1 \end{bmatrix}.$$

应用 (一): 最小二乘解

我们应用奇异值分解来求出线性方程组 $A\mathbf{x} = \beta$ 的最小二乘解.
首先, 将 A 写为

$$A = U\Sigma V^{\mathsf{T}},$$

则 $A\mathbf{x} = \beta$ 等价于 $\Sigma V^{\mathsf{T}}\mathbf{x} = U^{\mathsf{T}}\beta$. 只需找到 Σ 的某种 "逆" 矩阵, 就可完成求解.
对于 $m \times n$ 对角矩阵

$$\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_r, 0, \cdots, 0),$$

其中 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, 我们约定 Σ^{-} 为如下的 $n \times m$ 对角矩阵

$$\Sigma^{-} = \mathrm{diag}\Big(\frac{1}{\sigma_1}, \cdots, \frac{1}{\sigma_r}, 0, \cdots, 0\Big).$$

我们断言: $\mathbf{x}_0 = V\Sigma^{-}U^{\mathsf{T}}\beta$ 是线性方程组 $A\mathbf{x} = \beta$ 的一个最小二乘解, 即它满足 $A^{\mathsf{T}}A\mathbf{x}_0 = A^{\mathsf{T}}\beta$.
事实上, $A^{\mathsf{T}}A\mathbf{x}_0 = (V\Sigma^{\mathsf{T}}\Sigma V^{\mathsf{T}})(V\Sigma^{-}U^{\mathsf{T}}\beta) = V\Sigma^{\mathsf{T}}U^{\mathsf{T}}\beta = A^{\mathsf{T}}\beta$.

应用 (二): 图片压缩

考虑一张 1200 万像素的照片, 通常它的大小是 3000×4000 像素. 如果把每个像素点的颜色用一个整数表示, 则这张照片可以写成一个 3000×4000 的矩阵 A. 为了在电脑中保存这张照片, 我们是否总需要保存 3000×4000 个数呢?

一般地, 若 $m \times n$ 矩阵 A 的奇异值分解为 $A = U\Sigma V^T$, 则当 A 的秩为 r 时, 我们只需保存 U, V 中的 r 个列向量, 从而一共只需保存 $r(m+n+1)$ 个数. 当 r 较小时, 这将显著地减少保存的数据量.

事实上, 许多照片的奇异值具有这样的特点: 除了较大的几个奇异值以外, 大多数奇异值都是非常小的. 把这些较小的奇异值修改为零, 则我们得到了一个新的矩阵, 它的秩比原来小, 从而需要保存的数据少, 但图片的质量 (清晰度) 并没有太大的损失. 下面几个网页提供了一些具体的例子.

- 狮子 https://aaronschlegel.me/image-compression-singular-value-decomposition.html
- 拱门 https://www.frankcleary.com/svdimage/
- 老虎 http://andrew.gibiansky.com/blog/mathematics/cool-linear-algebra-singular-value-decomposition/

应用 (三): 数据恢复

让我们从一个幼稚的例子看起. 假设原有一个矩阵

$$A = \begin{bmatrix} 1 & 2 & -2 & 3 \\ -4 & 0 & 1 & 2 \\ 3 & -2 & 1 & -5 \end{bmatrix}.$$

现在, 由于某种原因, 其中的数据被干扰了, 我们看到的矩阵为

$$\widetilde{A} = \begin{bmatrix} 1.01 & 2.01 & -2 & 2.99 \\ -4.01 & 0.01 & 1.01 & 2.02 \\ 3.01 & -1.99 & 1 & -4.98 \end{bmatrix}.$$

计算可知, $\widetilde{A}$ 的奇异值分解为 $\widetilde{A} = U\Sigma V^{\mathsf{T}}$, 其中

$$U = \begin{bmatrix} -0.357 & 0.736 & 0.575 \\ -0.459 & -0.674 & 0.578 \\ 0.813 & -0.057 & 0.579 \end{bmatrix}, \quad V^{\mathsf{T}} = \begin{bmatrix} 0.513 & -0.306 & 0.139 & -0.790 \\ 0.744 & 0.360 & -0.502 & 0.255 \\ 0.271 & 0.489 & 0.819 & 0.131 \\ 0.332 & -0.733 & 0.241 & 0.542 \end{bmatrix},$$

对角矩阵

$$\Sigma = \begin{bmatrix} 7.655 & & & \\ & 4.404 & & \\ & & 0.017 & 0 \end{bmatrix}.$$

从中可以看出, $\widetilde{A}$ 的第 3 个奇异值显著地小于前 2 个, 因此 $\widetilde{A}$ 接近于某个秩为 2 的矩阵. 通过把第 3 个奇异值修改为 0, 我们可以获得一个更接近 A 的矩阵

$$\begin{bmatrix} 1.007 & 2.005 & -2.008 & 2.989 \\ -4.013 & 0.005 & 1.002 & 2.019 \\ 3.007 & -1.995 & 0.992 & -4.981 \end{bmatrix}$$

这个例子的更实际的版本, 可以是以下场景:

- 一张照片被水浸湿了一小部分, 如何在一定程度上恢复出原来的照片?
- 硬盘上的一系列文件中有一个小文件不小心误删了, 如何设法恢复?