# 7.6 极分解和奇异值分解

Tiao Lu

Peking University

2017.12.1

## §7.6 极分解与奇异值分解

回想一下我们在 $\mathbf{C}$ 和 $\mathcal{L}(V)$ 之间做的类比. 按照这个类比, 一个复数 $z$ 相应于一个算子 $T$, 而 $\bar{z}$ 相应于 $T^*$. 实数相应于自伴算子, 而非负数相应于 (不恰当地所谓) 正算子. $\mathbf{C}$ 的另一个重要的子集是单位圆, 它由所有满足 $|z| = 1$ 的复数 $z$ 组成. 条件 $|z| = 1$ 等价于 $\bar{z}z = 1$. 按照我们的类比, 这相应于条件 $T^*T = I$, 等价于 $T$ 是等距同构 (参见 7.36). 也就是说, $\mathbf{C}$ 中的单位圆相应于全体等距同构.

继续我们的类比, 注意到每个非零复数 $z$ 都可以写成如下形式

$$z=\left(\frac{z}{|z|}\right)|z|=\left(\frac{z}{|z|}\right)\sqrt{\bar{z}z},$$

其中第一个因子, 即 $z/|z|$, 是单位圆中的元素. 这种类比使我们猜到, 任何算子 $T\in\mathcal{L}(V)$ 都可以写成一个等距同构和 $\sqrt{T^*T}$ 的乘积. 我们现在就来证明这个猜测确实是对的.

**7.41 极分解 (Polar Decomposition)**: 如果 $T\in\mathcal{L}(V)$, 则有一个等距同构 $S\in\mathcal{L}(V)$ 使得

$$T=S\sqrt{T^*T}.$$

**证明**: 假设 $T \in \mathcal{L}(V)$. 若 $\boldsymbol{v} \in V$, 则

$$\begin{aligned}\|T\boldsymbol{v}\|^2 &= \langle T\boldsymbol{v}, T\boldsymbol{v}\rangle \\ &= \langle T^*T\boldsymbol{v}, \boldsymbol{v}\rangle \\ &= \langle \sqrt{T^*T}\sqrt{T^*T}\boldsymbol{v}, \boldsymbol{v}\rangle \\ &= \langle \sqrt{T^*T}\boldsymbol{v}, \sqrt{T^*T}\boldsymbol{v}\rangle \\ &= \|\sqrt{T^*T}\boldsymbol{v}\|^2.\end{aligned}$$

于是, 对所有 $\boldsymbol{v} \in V$ 都有

**7.42** $$\|T\boldsymbol{v}\| = \|\sqrt{T^*T}\boldsymbol{v}\|.$$

我们已经证明了,

$$\|Tv\| = \|\sqrt{T^* T}v\|, \quad \forall v \in V. \tag{1}$$

显然 $\sqrt{T^* T}$ 是一个正算子 (不恰当的称谓).

**利用 Eq. (??)**

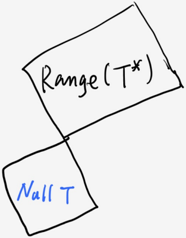
$\operatorname{null} T = \operatorname{null}\sqrt{T^* T}$

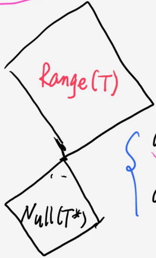
命题6.46

domain 和 codomain $V = W$

说 $\operatorname{Null} T = (\operatorname{Range}(T^*))^{\perp}$

$\operatorname{Null} T^* = (\operatorname{Range} T)^{\perp}$




从而

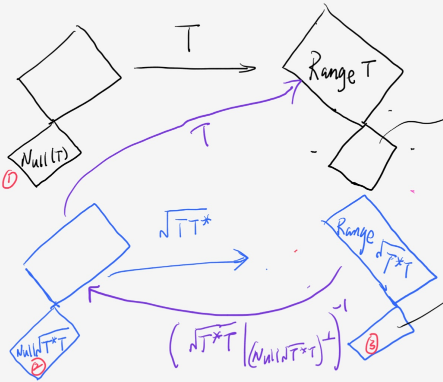
$$\begin{cases} \dim \operatorname{Null} T + \dim \operatorname{Range} T^* = \dim V \\ \dim \operatorname{Null} T^* + \dim \operatorname{Range} T = \dim V \end{cases}$$

$\Rightarrow \dim \operatorname{Range} T = \dim \operatorname{Range} T^*$

$\Rightarrow \dim \operatorname{Null} T = \dim \operatorname{Null} T^*$

由线性映射基本定理

$$\begin{cases} \dim \operatorname{Null}(T) + \dim \operatorname{Range} T = \dim V \\ \dim \operatorname{Null}(T^*) + \dim \operatorname{Range} T^* = \dim V \end{cases}$$




基 $f_1, f_2, \cdots, f_m$ (规范正交)

基 $g_1, g_2, \cdots, g_m$ (规范正交基)

下面说明为什么二者个数一样多

$\|Tv\| = \|\sqrt{T^*T}\,v\| \Rightarrow \underset{①}{\text{Null}\,T} \underset{=}{=} \underset{②}{\text{Null}\sqrt{T^*T}} \Rightarrow \dim \text{Null}\,T = \dim \text{Null}\sqrt{T^*T}$

$\Rightarrow \dim \text{Range}\,T = \dim \text{Range}\sqrt{T^*T} \Rightarrow \dim(\text{Range}\,T)^\perp = \dim(\text{Range}\sqrt{T^*T})^\perp$

定义 $Sg_i = f_i,\ i=1,\cdots,m.$

② = ③ (因为 $\sqrt{T^*T}$ 是自伴算子)

① = ② = ③

$\sqrt{T^*T}$ 作为 $(\mathrm{Null}\sqrt{T^*T})^{\perp} \longrightarrow \mathrm{Range}\sqrt{T^*T}$ 是一个双射，它的逆映射存在 记作 $\underbrace{\left[(\sqrt{T^*T})\big|_{(\mathrm{Null}\sqrt{T^*T})^{\perp}}\right]^{-1}}$ 简记 $(T^*T)^{-\frac{1}{2}}$

对于 $v \in \mathrm{Range}\sqrt{T^*T}$，$Sv = T\,(T^*T)^{-1/2}$

$$\|Sv\| = \|T\,(T^*T)^{-\frac{1}{2}}v\| = \|\sqrt{T^*T}(\sqrt{T^*T})^{-1}v\| = \|v\|$$

因 $\|Tx\| = \|\sqrt{T^*T}x\|, \forall x \in V$

$u \in \mathrm{Null}(\sqrt{T^*T})$，$Su = S\sum_{i=1}^{n} c_i g_i = \sum_{i=1}^{n} c_i S g_i = \sum_{i=1}^{n} c_i f_i$

$\|S(u+v)\|^2 = \|Su\|^2 + \|Sv\|^2 = \|u\|^2 + \|v\|^2 = \|u+v\|^2$. 由勾股定理.

例5：设 $e_1, e_2$ 是 $V$ 的规范正交基，$T \in \mathcal{L}(V)$，$Te_1 = 0$，$Te_2 = 2e_1$

$$\mathcal{M}(T) = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \qquad \mathcal{M}(T^*) = \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix},$$

$$T^*Te_1 = 0, \quad T^*Te_2 = T^*(2e_1) = 2T^*e_1 = 4e_2$$

$$\mathcal{M}(T^*T) = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \qquad \mathcal{M}(\sqrt{T^*T}) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

$\mathrm{Null}\, T = \mathrm{span}\{e_1\}$，$\mathrm{Null}\sqrt{T^*T} = \mathrm{span}(e_1)$.

$\mathrm{Range}\, T = \mathrm{span}(e_1)$ $\quad \mathrm{Range}\sqrt{T^*T} = \mathrm{span}(e_2)$

$\dim \mathrm{Range}\, T = \dim \mathrm{Range}\sqrt{T^*T}$，但 $\mathrm{Range}\, T \neq \mathrm{Range}\sqrt{T^*T}$

$\sqrt{TT^*}$ 是一个自伴算子，所以还可以观察到 $\mathrm{Null}\sqrt{T^*T} \perp \mathrm{Range}\sqrt{T^*T}$

$T$ 不是一个自伴算子，所以 $\mathrm{Null}\, T$ 和 $\mathrm{Range}\, T$ 不垂直.

我们构造一个等距同构 $S$，使得 $T = S\sqrt{T^*T}$

① $\text{Null}\sqrt{T^*T} = \text{Null}\,T$，但我们并没有在二者之间作一个一一对应，而是利用 $\dim \text{Null}\sqrt{T^*T} = \dim(\text{Range}\,T)^\perp$

$\text{Null}\sqrt{T^*T} = (\text{Range}\sqrt{T^*T})^\perp$ 对自伴算子 $\sqrt{T^*T}$ 成立

将 $(\text{Range}\sqrt{T^*T})^\perp$ 和 $(\text{Range}\,T)^\perp$ 对应起来。

在这我们这个例子中，$(\text{Range}\sqrt{T^*T})^\perp = \text{span}(e_1)$

$(\text{Range}\,T)^\perp = \text{span}(e_2)$，因此 $Se_1 = e_2$

然后：我们来看如何在 $\text{Range}\sqrt{T^*T} = \text{span}(e_2)$ 上定义 $S$.

$\sqrt{T^*T}$ 限制在 $(\text{Null}\sqrt{T^*T})^{\perp} \longrightarrow \text{Range}\sqrt{T^*T}$ 是一个可逆的线性

映射。要注意 $\text{Null}\sqrt{T^*T} = \text{Null}\,T$，故 $(\text{Null}\sqrt{T^*T})^{\perp} = (\text{Null}\,T)^{\perp}$.

$$\text{Null}\,T = \text{span}(e_1) \quad \text{Null}\sqrt{T^*T} = \text{span}(e_1)$$

$$\sqrt{T^*T} : \text{span}(e_2) \longrightarrow \text{span}(e_2) \text{ 是可逆映射}$$

$$Se_2 = T\left(\sqrt{T^*T}\right)^{-1} e_2 = T\left(\tfrac{1}{2} e_2\right) = e_1$$

这样定义的 $S$ 对任意的 $v = c_1e_1 + c_2e_2$

$$Sv = S(c_1e_1 + c_2e_2) = c_1Se_1 + c_2e_2 = c_1e_2 + c_1e_1$$

是一个等距映射。 $M(T) = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$, $M(S) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $M(\sqrt{T^*T}) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$

**定义 0.1 (Unitary operator)**

*A unitary operator is a bounded linear operator $U \in \mathcal{V}$, where $V$ is a finite-dimensional nonzero inner-product space, that satisfies $U^* U = UU^* = I$.*

- The weaker condition $U^* U = I$ defines an isometry.
- The other condition, $UU^* = I$, defines a coisometry.

# 奇异值分解

$T \in \mathcal{L}(V)$ 是一个线性算子，其中 $V$ 是一个复内积空间。既然 $T$ 的 domain 和 codomain 是同一个空间，我们一般为 domain 和 codomain 选取同一个基。此时，我们知道我们只能找到一个规范正交基，使得 $T$ 的矩阵是上三角矩阵。当 $A$ 是正规算子时，我们才能作对角化。但如果为 domain 和 codomain 选用不同的基，则对任意的算子 $T \in \mathcal{L}(V)$，我们总能找到两个规范正交基 $e_1, e_2, \ldots, e_n$ 和 $f_1, \ldots, f_n$ 使得 $\mathcal{M}(T, (e_1, \ldots, e_n), (f_1, \ldots, f_n))$ 是对角矩阵，这称作奇异值分解。

$T\in\mathcal{L}(V)$、由极分解定理知、存在一个等距映射 $S$、使得

$$T = S\sqrt{T^*T}.$$

由于 $\sqrt{T^*T}$ 是一个正算子，因此它一定是自伴算子，那么一定存在一个规范正交基 $e_1, e_2, \ldots, e_n$、使得

$$\sqrt{T^*T}\, e_j = \sigma_j e_j. \quad j=1,\ldots,n$$

且 $\sigma_j$ 是非负实数。从而 $Te_j = S\sqrt{T^*T}\,e_j = S\sigma_j e_j = \sigma_j S e_j$

如果 $f_j = Se_j,\ j=1,2,\ldots,n$ 是 $V$ 的一个规范正交基，那么我们就证明了奇异值分解定理。$\langle f_j, f_k\rangle = \langle Se_j, Se_k\rangle = \langle e_j, S^*Se_k\rangle = \langle e_j, e_k\rangle = \delta_{j,k}$

这样我们就利用 $S$ 是等距映射证明了 $f_j$ 也是规范正交基。 $\mathcal{M}(T, e_j, f_j) = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$

其中 $\sigma_1, \cdots, \sigma_n$ 称为 $T$ 的奇异值（singular value）

它是 $\sqrt{T^*T}$ 的特征值，也是 $T^*T$ 的特征值的正平方根.

矩阵的奇异值分解：我们有一个矩阵 $A \in \mathbb{C}^{n\times n}$，

我们可以认为它是算子 $T$ 关于标准基 $e_1, e_2, \cdots, e_n$ 的矩阵.

由算子的奇异值分解定理知，存在两个规范正交基 $\varphi_1, \cdots, \varphi_n$

和 $\psi_1, \cdots, \psi_n$，使得 $M(T; (\varphi_1, \cdots, \varphi_n), (\psi_1, \cdots, \psi_n)) = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix}$

问题：$M(T; (\varphi_1, \cdots, \varphi_n), (\psi_1, \cdots, \psi_n))$ 和 $A$ 之间的关系是什么？

设 $\phi_j = V_{1,j} e_1 + V_{2,j} e_2 + \cdots V_{n,j} e_n$

写成矩阵形式 $[\phi_1 \ \phi_2 \ \cdots \ \phi_n] = [e_1 \ e_2 \ \cdots e_n] \begin{bmatrix} V_{1,1} & V_{1,2} & \cdots & V_{1,n} \\ V_{2,1} & V_{2,2} & \cdots & V_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ V_{n,1} & V_{n,2} & & V_{n,n} \end{bmatrix}$

由于 $(\phi_1, \cdots, \phi_n)$ 和 $(e_1, \cdots, e_n)$ 都是规范正交基

$\|\phi_j\|^2 = \sum_j |V_{i,j}|^2 = 1$, $\langle \phi_j, \phi_k \rangle = \sum_{l=1}^{n} V_{l,j} \overline{V_{l,k}} = 0$

容易验证 $V = [V_{ij}]$ 是一个酉矩阵 (unitary matrix)

$V^* V = I$.

设 $\psi_j = \sum_{k=1}^{n} U_{k,j} e_k$, 写成矩阵形式 $[\psi_1, \cdots, \psi_n] = [e_1, \cdots, e_n] U$

容易由 $(\psi_j)$ 和 $(e_j)$ 是规范正交基验证 $U^* U = I$.

$$\underbrace{M\left(T;(\phi_1,\dots,\phi_n),(\psi_1,\dots,\psi_n)\right)}_{B} = U^* A V$$

证明上式：

$$T\phi_i = T V_{k,i} e_k = V_{k,i} T e_k = V_{k,i} A_{\ell,k} e_\ell \quad (*)$$

因 $[\psi_1, \dots, \psi_n] = [e_1, \dots, e_n] \begin{bmatrix} U_{11} & \cdots & U_{1n} \\ \vdots & \ddots & \vdots \\ U_{n1} & \cdots & U_{nn} \end{bmatrix}$

$$[e_1 \cdots e_n] = [\psi_1 \cdots \psi_n] U^{-1} = [\psi_1 \cdots \psi_n] U^*$$

$$e_\ell = U^*_{j\ell} \psi_j$$

继续(*)式，用上式代 $e_\ell$ 得 $T\phi_i = V_{k,i} A_{\ell,k} U^*_{j,\ell} \psi_j$

$$= U^*_{j,\ell} A_{\ell,k} V_{k,i} \psi_j$$

这样就证明 $B = U^* A V$，其中 $B$ 是对角阵。

上面的式子，我们也经常写成 $\forall A \in \mathbb{C}^{m\times n}$，$\exists$ 两个矩阵 $U$ 和 $V$.

使得 $A = U \Sigma V$.

其中 $\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$ 是对角阵 $\sigma_1, \dots, \sigma_n$ 是非负实数

一般来说，很多矩阵的奇异值可以按照由大到小的顺序排列

$$\sigma_1 \geq \sigma_2 \geq \cdots \sigma_j \geq \cdots$$

随着 $j$ 的增大，$\sigma_j$ 迅速的变小，我们可以只保留一个或几个奇异值，得到下面的近似 $A \approx U \begin{bmatrix} \sigma_1 & & \\ & 0 & \\ & & \ddots & \\ & & & 0 \end{bmatrix} V$.

举个具体的例子

$$A \approx \begin{bmatrix} U_{1,1} & U_{1,2} & U_{1,3} & U_{1,4} \\ U_{2,1} & U_{2,2} & U_{2,3} & U_{2,4} \\ U_{3,1} & U_{3,2} & U_{3,3} & U_{3,4} \\ U_{4,1} & U_{4,2} & U_{4,3} & U_{4,4} \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{bmatrix} \begin{bmatrix} V_{1,1} & V_{1,2} & V_{1,3} & V_{1,4} \\ V_{2,1} & V_{2,2} & V_{2,3} & V_{2,4} \\ V_{3,1} & V_{3,2} & V_{3,3} & V_{3,4} \\ V_{4,1} & V_{4,2} & V_{4,3} & V_{4,4} \end{bmatrix}$$

$$= \begin{bmatrix} U_{1,1} & U_{1,2} & U_{1,3} & U_{1,4} \\ U_{2,1} & U_{2,2} & U_{2,3} & U_{2,4} \\ U_{3,1} & U_{3,2} & U_{3,3} & U_{3,4} \\ U_{4,1} & U_{4,2} & U_{4,3} & U_{4,4} \end{bmatrix} \begin{bmatrix} \sigma_1 V_{1,1} & \sigma_1 V_{1,2} & \sigma_1 V_{1,3} & \sigma_1 V_{1,4} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$
= \begin{bmatrix}
U_{1,1}\sigma_1 V_{1,1} & U_{1,1}\sigma_1 V_{1,2} & U_{1,1}\sigma_1 V_{1,3} & U_{1,1}\sigma_1 V_{1,4} \\
U_{2,1}\sigma_1 V_{1,1} & U_{2,1}\sigma_1 V_{1,2} & U_{2,1}\sigma_1 V_{1,3} & U_{2,1}\sigma_1 V_{1,4} \\
U_{3,1}\sigma_1 V_{1,1} & U_{3,1}\sigma_1 V_{1,2} & U_{3,1}\sigma_1 V_{1,3} & U_{3,1}\sigma_1 V_{1,4} \\
U_{4,1}\sigma_1 V_{1,1} & U_{4,1}\sigma_1 V_{1,2} & U_{4,1}\sigma_1 V_{1,3} & U_{4,1}\sigma_1 V_{1,4}
\end{bmatrix}
$$

$$
= \sigma_1 \begin{bmatrix} U_{1,1} \\ U_{2,1} \\ U_{3,1} \\ U_{4,1} \end{bmatrix} \begin{bmatrix} V_{1,1} & V_{1,2} & V_{1,3} & V_{1,4} \end{bmatrix}
$$

A 共有 $n^2$ 个数、它的近似（秩1）近似只需要存储 $(2n+1)$ 个数（一个奇异值，两个向量）。

当 $n=10^6$ 时 $n^2=10^{12}$，每个复数 16个字节 $16\times10^{12}$ bytes $=$ 16T

$(2n+1) = (2\times10^6+1)\times16 \approx$ 32M。差别巨大。

# Homework

Problems 10-34.
提示见 lectureLA15hw.pdf (见 ftp)