

Lecture 9 内积空间 — 2025.11.11

教授：邵美悦

Scribe: 路人甲

1 大纲

1. 内积空间上的线性映射
2. 奇异值分解

2 内积空间上的线性映射

2.1 Riesz 表示定理

我们考虑数域 \mathbb{F} 上的内积空间 \mathcal{V} , 并定义 \mathcal{V} 上的线性泛函 $f : \mathcal{V} \rightarrow \mathbb{F}$. 我们知道, 内积是一种线性泛函。那末, 是否所有的线性泛函都可以表示成内积的形式呢?

当然不对! 反例如: 设 $L[f] = f(0)$, 它显然不能显示地表示成内积的形式。如果我们想要强行把它写成内积的形式, 我们就要引入 δ 函数。这是一个广义函数, 并不属于我们通常讨论的函数空间。有关 δ 函数的更多内容, 请参考第五讲讲义。

不过, 大多数情况并没有那么糟糕。如果我们把内积空间限制在有限维空间上, 那么我们就会有如下定理:

定理 2.1 (Riesz 表示定理). 有限维线性空间 \mathcal{V} 上的线性泛函 f 一定具有 $f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y} \rangle$ 的形式。其中 \mathbf{y} 是由 f 唯一确定的。

证明. 设 $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ 是 \mathcal{V} 的一组标准正交基, 则对于任意的 $\mathbf{x} \in \mathcal{V}$, 都有

$$\mathbf{x} = \sum_{i=1}^n \mathbf{e}_i \alpha_i.$$

因为 f 是线性的, 所以

$$f(\mathbf{x}) = \sum_{i=1}^n f(\mathbf{e}_i) \alpha_i.$$

因此，我们只要找到一个向量 \mathbf{y} ，使得

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n f(\mathbf{e}_i) \alpha_i.$$

我们把 $\mathbf{x} = \sum_{i=1}^n \mathbf{e}_i \alpha_i$ 代入上式，得到

$$\sum_{i=1}^n \alpha_i \langle \mathbf{e}_i, \mathbf{y} \rangle = \sum_{i=1}^n f(\mathbf{e}_i) \alpha_i.$$

因此，我们只需让以下式子对任意的 i 成立：

$$\langle \mathbf{e}_i, \mathbf{y} \rangle = f(\mathbf{e}_i).$$

这也就是说， \mathbf{y} 在 \mathbf{e}_i 上的投影长度一定要是 $f(\mathbf{e}_i)$. 因此，我们设

$$\mathbf{y} = \sum_{i=1}^n \mathbf{e}_i \overline{f(\mathbf{e}_i)}.$$

这样我们就一定有

$$f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y} \rangle.$$

唯一性是显然的，因为如果存在另一个向量 \mathbf{y}' 也满足上式，那么对于任意的 \mathbf{x} ，都有

$$\langle \mathbf{x}, \mathbf{y} - \mathbf{y}' \rangle = 0.$$

这说明 $\mathbf{y} - \mathbf{y}' = \mathbf{0}$ ，即 $\mathbf{y} = \mathbf{y}'$. 因此， \mathbf{y} 由 f 唯一确定。 \square

有了这个定理，我们再来看线性泛函 f . 由 Riesz 表示定理，我们有

$$\begin{aligned} f(\mathbf{x}) &= \left\langle \sum_{i=1}^n \mathbf{e}_i \alpha_i, \sum_{i=1}^n \mathbf{e}_i \overline{f(\mathbf{e}_i)} \right\rangle \\ &= \sum_{i=1}^n \alpha_i f(\mathbf{e}_i) \langle \mathbf{e}_i, \mathbf{e}_i \rangle \\ &= \sum_{i=1}^n \alpha_i f(\mathbf{e}_i) \\ &= \langle \boldsymbol{\alpha}, \overline{f(\mathbf{e})} \rangle. \end{aligned}$$

这一事实说明：对于任何一个线性泛函，它的内积计算可以简化为坐标向量与基向量的像的内积计算。这强调了在系数空间 \mathbb{C}^n 中，线性泛函的行为与标准内积是类似的。

对偶空间 在第五讲中，我们提到过 \mathcal{V} 上有界线性泛函的全体构成一个线性空间。这个空间被称为 \mathcal{V} 的对偶空间，记作 \mathcal{V}^* . Riesz 表示定理告诉我们，有限维内积空间 \mathcal{V} 上的线性泛函等价于一个向量 $\mathbf{y} \in \mathcal{V}$ ，因此 $\mathcal{V}^* = \mathcal{V}$.

2.2 伴随算子

接下来我们将刚刚的讨论进一步推广。设有线性变换 $T : \mathcal{V} \rightarrow \mathcal{V}$, 固定一个 \mathbf{v} , 我们考虑线性泛函

$$f(\mathbf{x}) = \langle T(\mathbf{x}), \mathbf{v} \rangle.$$

根据 Riesz 表示定理, 存在一个由 f 唯一确定的向量, 我们姑且称它为 $T^*(\mathbf{v})$, 使得

$$f(\mathbf{x}) = \langle \mathbf{x}, T^*(\mathbf{v}) \rangle.$$

这也就是说, 对于任意的 $\mathbf{x}, \mathbf{v} \in \mathcal{V}$, 都有

$$\langle T(\mathbf{x}), \mathbf{v} \rangle = \langle \mathbf{x}, T^*(\mathbf{v}) \rangle.$$

这个恒等式揭示了一个深刻的事实: 对于给定的线性变换 T , 总是存在一个线性变换 T^* , 使得上式恒成立。接下来, 我们就来研究这个线性变换 T^* .

定义 2.2 (伴随算子). 设 \mathcal{V} 是数域 \mathbb{F} 上的内积空间, $T : \mathcal{V} \rightarrow \mathcal{V}$ 是一个线性变换。定义线性变换 $T^* : \mathcal{V} \rightarrow \mathcal{V}$, 使得对于任意的 $\mathbf{x}, \mathbf{v} \in \mathcal{V}$, 都有

$$\langle T(\mathbf{x}), \mathbf{v} \rangle = \langle \mathbf{x}, T^*(\mathbf{v}) \rangle.$$

则称 T^* 为 T 的伴随算子。

伴随算子是一种“翻转”的艺术。从形式上来看, 伴随算子 T^* 可以理解为在内积意义下将变换 T 从内积的第一个位置“转移”到第二个位置的操作。这种转移不仅仅是形式上的, 更深层次地反映了线性变换在内积空间中的几何性质。它回答了这样一个问题: 为了保持内积等式成立, 当我在内积的一个位置应用变换 T 时, 我需要在另一个位置应用什么样的变换来补偿这种变化。那么, 这个伴随算子是唯一的吗?

定理 2.3. 伴随算子是唯一的。伴随算子的伴随算子是原算子本身。

证明. 设 S 和 S' 都是 T 的伴随算子, 则对于任意的 $\mathbf{x}, \mathbf{v} \in \mathcal{V}$, 都有

$$\langle \mathbf{x}, S(\mathbf{v}) \rangle = \langle T(\mathbf{x}), \mathbf{v} \rangle = \langle \mathbf{x}, S'(\mathbf{v}) \rangle.$$

因此,

$$\langle \mathbf{x}, S(\mathbf{v}) - S'(\mathbf{v}) \rangle = 0.$$

这说明 $S(\mathbf{v}) - S'(\mathbf{v}) = \mathbf{0}$, 即 $S = S'$. 因此, 伴随算子是唯一的。

接下来, 我们证明伴随算子的伴随算子是原算子本身。设 T^* 是 T 的伴随算子, 则对于任意的 $\mathbf{x}, \mathbf{v} \in \mathcal{V}$, 都有

$$\langle T(\mathbf{x}), \mathbf{v} \rangle = \langle \mathbf{x}, T^*(\mathbf{v}) \rangle.$$

由内积的 Hermite 对称性，我们立刻有

$$\langle T^*(\mathbf{x}), \mathbf{v} \rangle = \langle \mathbf{x}, T(\mathbf{v}) \rangle.$$

这说明 T 是 T^* 的伴随算子。由伴随算子的唯一性，我们有 $(T^*)^* = T$.

另证：设 T 在标准正交基下的表示矩阵为 $\tilde{\mathbf{T}}$, T^* 在标准正交基下的表示矩阵为 $\tilde{\mathbf{G}}$. 再设 \mathbf{x} 在标准正交基下的坐标向量为 $\tilde{\mathbf{x}}$, \mathbf{v} 在标准正交基下的坐标向量为 $\tilde{\mathbf{v}}$. 则由伴随算子的定义，我们有

$$\tilde{\mathbf{v}}^* \tilde{\mathbf{T}} \tilde{\mathbf{x}} = \tilde{\mathbf{v}}^* \tilde{\mathbf{G}}^* \tilde{\mathbf{x}}.$$

因此 $\tilde{\mathbf{T}} = \tilde{\mathbf{G}}^*$. 由此可见, $(T^*)^* = T$. □

伴随算子 T^* 的引入，源于我们将 Riesz 表示定理从一个静态的线性泛函，动态地应用于由 T 生成的一族线性泛函 $\langle T(\mathbf{x}), \mathbf{v} \rangle$ 上。其定义方程 $\langle T(\mathbf{x}), \mathbf{v} \rangle = \langle \mathbf{x}, T^*(\mathbf{v}) \rangle$ 是连接线性变换与内积结构的桥梁。我们可以从伴随算子的视角，重新审视我们之前学过的知识。

自伴算子 我们已经清楚地了解了 Hermite 矩阵的代数性质。现在我们从几何的角度来重新认识 Hermite 矩阵。

定义 2.4 (自伴算子). 伴随算子是其本身的算子称为自伴算子。

显然，由于伴随算子对应于转置共轭，若原算子在标准正交基下的表示矩阵为 \mathbf{A} ，我们有 $\mathbf{A}^* = \mathbf{A}$ ，因此自伴算子在标准正交基下的表示矩阵是 Hermite 矩阵。我们再来仔细看看它的矩阵结构。由于 Hermite 矩阵可对角化，考虑 \mathbf{A} 的谱分解 $\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^*$ ，其中 \mathbf{Q} 是酉矩阵， Λ 是实对角阵。那么

$$T(\mathbf{E}\mathbf{Q}) = \mathbf{E}(\mathbf{Q}\Lambda\mathbf{Q}^*)\mathbf{Q} = (\mathbf{E}\mathbf{Q})\Lambda,$$

这说明：对于自伴算子，我们可以找到一组标准正交基，让它的表示矩阵是一个很简单的矩阵——实对角阵。

这一结论具有重要的几何意义：自伴算子在某个合适的标准正交坐标系下，其作用仅仅是在各个坐标方向上进行伸缩变换（伸缩系数为实数）。这为我们理解自伴算子的几何本质提供了清晰的图像。

正规算子 自伴算子在特定的基上的表示矩阵是实对角阵。这引出了一个更深层次的问题：除了自伴算子，还有哪些线性算子在某个标准正交基下对角化？

我们先来观察自伴算子，观察一下它有什么重要结构是与可对角化密切相关的。我们观察的原则是：只利用可对角化这个性质，而不利用自伴算子的 Hermite 性质。假设 T 是一个自伴算子，并

且它在某个标准正交基 \mathbf{E} 下的表示矩阵为对角阵 Λ . 对于任意的向量 \mathbf{x}, \mathbf{y} , 假设它们在基 \mathbf{E} 下的坐标向量分别为 $\boldsymbol{\alpha}, \boldsymbol{\beta}$, 则

$$\begin{aligned}\langle T(\mathbf{x}), T(\mathbf{y}) \rangle &= \langle \mathbf{E}\Lambda\boldsymbol{\alpha}, \mathbf{E}\Lambda\boldsymbol{\beta} \rangle \\ &= (\Lambda\boldsymbol{\beta})^*(\Lambda\boldsymbol{\alpha}) \\ &= \boldsymbol{\beta}^*\Lambda^*\Lambda\boldsymbol{\alpha} \\ &= \boldsymbol{\beta}^*\Lambda\Lambda^*\boldsymbol{\alpha} \\ &= \langle \mathbf{E}\Lambda^*\boldsymbol{\alpha}, \mathbf{E}\Lambda^*\boldsymbol{\beta} \rangle \\ &= \langle T^*(\mathbf{x}), T^*(\mathbf{y}) \rangle.\end{aligned}$$

上面的第一个等号是由线性映射的基本性质得到的, 第二个等号是因为两个向量的内积等价于它们在标准正交基下的坐标向量的内积, 第四个等号是因为对角阵的乘法可交换, 第五个等号则是反向利用了两个向量的内积等价于它们在标准正交基下的坐标向量的内积。

我们在上面的推导中并没有利用自伴算子的 Hermite 性质, 我们仅仅利用了可酉相似于复对角阵这一条性质。因此, 至少这个等式是对任何可对角化的线性算子成立的。这是一个重要的发现。我们为满足这个性质的线性算子起一个名字: 正规算子。

定义 2.5 (正规算子). 设 \mathcal{V} 是数域 \mathbb{F} 上的内积空间, $T : \mathcal{V} \rightarrow \mathcal{V}$ 是一个线性变换。如果对于任意的 $\mathbf{x}, \mathbf{y} \in \mathcal{V}$, 都有

$$\langle T(\mathbf{x}), T(\mathbf{y}) \rangle = \langle T^*(\mathbf{x}), T^*(\mathbf{y}) \rangle,$$

则称 T 为正规算子。

接下来, 我们自然会好奇一个问题: 正规算子是否都可对角化? 也就是说, 满足这个性质的线性算子, 是否都可以找到一个标准正交基, 使得它在该基下的表示矩阵是对角阵?

定理 2.6 (谱分解定理). 正规算子 \Leftrightarrow 可酉对角化。

证明. 充分性即定义的正向部分, 我们已经证明过了。下面我们证明必要性。

为方便起见, 我们在证明中使用矩阵表示法。设 T 在标准正交基下的表示矩阵为 \mathbf{A} . 则 T^* 在标准正交基下的表示矩阵为 \mathbf{A}^* . 由正规算子的定义, 我们有

$$\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}.$$

我们把矩阵 \mathbf{A} 称为 正规矩阵。由于任意复方阵都可以酉上三角化, 因此存在酉矩阵 \mathbf{Q} , 使得

$$\mathbf{A} = \mathbf{Q}\mathbf{T}\mathbf{Q}^*,$$

其中 \mathbf{T} 是上三角阵。代入上式, 得到

$$\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T}.$$

下面，我们利用归纳法证明：如果上三角阵 \mathbf{T} 满足 $\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T}$ ，则 \mathbf{T} 是对角阵。

当矩阵阶数为 1 时，结论显然成立。假设当矩阵阶数为 $k - 1$ 时结论成立，下面我们证明当矩阵阶数为 k 时结论也成立。设

$$\mathbf{T} = \begin{bmatrix} t_{1,1} & \mathbf{T}_{1,2} \\ \mathbf{0} & \mathbf{T}_{2,2} \end{bmatrix}.$$

由 $\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T}$ 得

$$\begin{bmatrix} t_{1,1} & \mathbf{T}_{1,2} \\ \mathbf{0} & \mathbf{T}_{2,2} \end{bmatrix} \begin{bmatrix} t_{1,1}^* & \mathbf{0} \\ \mathbf{T}_{1,2}^* & \mathbf{T}_{2,2}^* \end{bmatrix} = \begin{bmatrix} t_{1,1}^* & \mathbf{0} \\ \mathbf{T}_{1,2}^* & \mathbf{T}_{2,2}^* \end{bmatrix} \begin{bmatrix} t_{1,1} & \mathbf{T}_{1,2} \\ \mathbf{0} & \mathbf{T}_{2,2} \end{bmatrix}$$

我们把左右两边结果的左上角块进行比较，可得

$$t_{1,1}t_{1,1}^* + \mathbf{T}_{1,2}\mathbf{T}_{1,2}^* = t_{1,1}^*t_{1,1}.$$

亦即

$$\mathbf{T}_{1,2}\mathbf{T}_{1,2}^* = \mathbf{0}.$$

两边取迹，得到

$$\text{tr}(\mathbf{T}_{1,2}\mathbf{T}_{1,2}^*) = \|\mathbf{T}_{1,2}\|_F^2 = 0.$$

因此， $\mathbf{T}_{1,2} = \mathbf{0}$. 这说明 \mathbf{T} 是分块对角阵：

$$\mathbf{T} = \begin{bmatrix} t_{1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{2,2} \end{bmatrix}.$$

由归纳假设， $\mathbf{T}_{2,2}$ 是对角阵，因此 \mathbf{T} 也是对角阵。 \square

正规算子包含了内积空间中最重要的几类算子，读者可自行验证：

- 自伴算子： $T^* = T$ ；
- 反自伴算子： $T^* = -T$ ；
- 正交算子： $T^*T = I$.

正规算子的几何意义在于：存在一组相互垂直的方向（标准正交基），算子在这组方向上的作用仅仅是伸缩变换。这与自伴算子的几何图像相似，只是现在伸缩因子可以是复数（而自伴算子的特征值必须是实数）。

对于一个实正规算子，有时候我们并不满足于它可酉对角化——我们希望它能够相似于一个实对角阵。很可惜的是，由于相似变换保持特征值，如果原算子有复特征值，那么它是怎么也不可能相似于实对角阵的——我们只能把要求再降低一些，它能不能正交相似于一个分块对角阵，每个块至多不超过 2 阶呢？我们说，这样是可以的。这样的对角阵我们管它叫拟对角阵。

定理 2.7. 欧氏空间上的实正规算子可以正交拟对角化（即存在一组标准正交基使得正规算子的表示矩阵为拟对角阵，其对角块要么是实数，要么是某个复数的实矩阵表示）。

证明. 我们从复对角阵出发。由于特征值是共轭成对出现的，因此我们可以通过行列交换将共轭的特征值放在一起。这个过程是酉变换，因此不会破坏酉矩阵的结构。

接下来，我们只需证明：对于每一个共轭特征值对 $\lambda = a + bi$ 和 $\bar{\lambda} = a - bi$ ，它们构成的对角块相类似于复数 $a + bi$ 的实矩阵表示，亦即我们需要证

$$\mathbf{M} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \sim \begin{bmatrix} a + bi & 0 \\ 0 & a - bi \end{bmatrix} = \mathbf{D}.$$

首先，注意到 $\det(\mathbf{M}) = \det(\mathbf{D})$ 以及 $\text{tr}(\mathbf{M}) = \text{tr}(\mathbf{D})$ ，因此显然它们的特征值是相同的。因此 \mathbf{M} 可以酉对角化，得到的对角阵一定就是 \mathbf{D} .

对每个共轭特征值对都进行上述操作，我们就得到了一个拟对角阵。现在我们只需说明这个相似变换的过程是正交变换即可。我们考虑 $\mathbf{AQ} = \mathbf{QD}$. 我们按 \mathbf{D} 的对角分块来划分 \mathbf{Q} :

$$\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_k],$$

其中每个 \mathbf{Q}_i 对应于 \mathbf{D} 的一个对角块。则对于每个 i ，都有

$$\mathbf{AQ}_i = \mathbf{Q}_i \mathbf{D}_i.$$

若 \mathbf{D}_i 是一阶块，则 \mathbf{Q}_i 是 \mathbf{A} 的实特征向量，因此 \mathbf{Q}_i 可以取为实单位向量。若 \mathbf{D}_i 是二阶块，设 $\mathbf{v} = \mathbf{u} + i\mathbf{w}$ 是特征值 $\lambda = a + bi$ 对应的复特征向量，其中 \mathbf{u} 和 \mathbf{w} 是实向量。由于 \mathbf{A} 是实正规算子，复特征向量 \mathbf{v} 和 $\bar{\mathbf{v}} = \mathbf{u} - i\mathbf{w}$ 在复内积下正交：

$$\langle \mathbf{v}, \bar{\mathbf{v}} \rangle = \mathbf{v}^* \bar{\mathbf{v}} = 0.$$

计算内积：

$$\langle \mathbf{v}, \bar{\mathbf{v}} \rangle = (\mathbf{u} + i\mathbf{w})^* (\mathbf{u} - i\mathbf{w}) = \mathbf{u}^T \mathbf{u} - i\mathbf{u}^T \mathbf{w} - i\mathbf{w}^T \mathbf{u} + i^2 \mathbf{w}^T \mathbf{w} = \|\mathbf{u}\|^2 - \|\mathbf{w}\|^2 - 2i(\mathbf{u} \cdot \mathbf{w}).$$

令实部和虚部为零：

$$\|\mathbf{u}\|^2 - \|\mathbf{w}\|^2 = 0, \quad \mathbf{u} \cdot \mathbf{w} = 0.$$

因此， $\|\mathbf{u}\| = \|\mathbf{w}\|$ 且 $\mathbf{u} \perp \mathbf{w}$ ，即 \mathbf{u} 和 \mathbf{w} 是正交且等长的实向量。

现在，构造标准正交基：

$$\mathbf{q}_1 = \frac{\mathbf{u}}{\|\mathbf{u}\|}, \quad \mathbf{q}_2 = \frac{\mathbf{w}}{\|\mathbf{w}\|} = \frac{\mathbf{w}}{\|\mathbf{u}\|}.$$

则 \mathbf{q}_1 和 \mathbf{q}_2 是标准正交的，且 $\mathbf{Q}_i = [\mathbf{q}_1, \mathbf{q}_2]$ 是实正交矩阵。

由特征方程 $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ 可得:

$$\mathbf{A}(\mathbf{u} + i\mathbf{w}) = (a + bi)(\mathbf{u} + i\mathbf{w}) = a\mathbf{u} - b\mathbf{w} + i(b\mathbf{u} + a\mathbf{w}).$$

分离实部和虚部:

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

现在计算 $\mathbf{A}\mathbf{q}_1$ 和 $\mathbf{A}\mathbf{q}_2$:

$$\mathbf{A}\mathbf{q}_1 = \frac{1}{\|\mathbf{u}\|}\mathbf{A}\mathbf{u} = \frac{1}{\|\mathbf{u}\|}(a\mathbf{u} - b\mathbf{w}) = a\mathbf{q}_1 - b\mathbf{q}_2,$$

$$\mathbf{A}\mathbf{q}_2 = \frac{1}{\|\mathbf{w}\|}\mathbf{A}\mathbf{w} = \frac{1}{\|\mathbf{u}\|}(b\mathbf{u} + a\mathbf{w}) = b\mathbf{q}_1 + a\mathbf{q}_2.$$

因此,

$$\mathbf{A}\mathbf{Q}_i = \mathbf{A}[\mathbf{q}_1, \mathbf{q}_2] = [\mathbf{A}\mathbf{q}_1, \mathbf{A}\mathbf{q}_2] = [\mathbf{q}_1, \mathbf{q}_2] \begin{bmatrix} a & b \\ -b & a \end{bmatrix} = \mathbf{Q}_i \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

这表明对于二阶块, 存在实正交矩阵 \mathbf{Q}_i 使得 $\mathbf{A}\mathbf{Q}_i = \mathbf{Q}_i \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$. \square

投影算子 我们在自伴算子的基础上进一步细化讨论。自伴算子 \mathbf{P} 满足 $\mathbf{P} = \mathbf{P}^*$. 假如它还满足幂等性, 即 $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^*$, 则称 \mathbf{P} 为投影算子。那么, 投影算子有什么几何意义呢? 我们考虑投影算子 \mathbf{P} 在某个标准正交基下的表示矩阵。由于投影算子是自伴算子, 因此它可以酉对角化。又因为投影算子是幂等的, 因此它的特征值只能是 0 或 1. 这说明, 投影算子在某个标准正交基下的表示矩阵是一个对角阵, 对角线上只有 0 和 1. 这说明投影算子实际上就是将空间分解为两个正交子空间——一个特征值为 1 的子空间和一个特征值为 0 的子空间。投影算子在第一个子空间上保持向量不变, 而在第二个子空间上把向量映射为零。这种作用其实就等价于将内积空间 \mathcal{V} 上任意的一个向量 \mathbf{x} 投影到第一个子空间上。这也是这个算子被称为投影算子的原因。

接下来我们再仔细考察它的矩阵结构。假设它在标准正交基下的表示矩阵为 \mathbf{P} , 并设它的秩为 r , 因此它有满秩分解 $\mathbf{P} = \mathbf{W}\mathbf{W}^*$, 其中 \mathbf{W} 是一个 $n \times r$ 的矩阵。接下来我们就来解释一下为什么能够得到这个分解。假设 \mathbf{P} 可以谱分解为 $\mathbf{P} = \mathbf{Q}\Lambda\mathbf{P}^*$, 其中 $\Lambda = \text{diag}(\mathbf{0}, \mathbf{I}_r)$. 我们对 \mathbf{Q} 进行分块:

$$\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2], \quad \mathbf{Q}_1 \in \mathbb{C}^{n \times (n-r)}, \quad \mathbf{Q}_2 \in \mathbb{C}^{n \times r}.$$

则有

$$\mathbf{P} = \mathbf{Q}\Lambda\mathbf{Q}^* = \mathbf{Q}_2\mathbf{Q}_2^*.$$

因此, 我们可以取 $\mathbf{W} = \mathbf{Q}_2$. 按满秩分解的几何意义, 我们知道 $\text{Range}(\mathbf{W}) = \text{Range}(\mathbf{P})$. 于是, 我们还可以以另一种方式来理解投影算子: 对 \mathcal{V} 上任意一个向量 \mathbf{x} , 我们先对 \mathbf{x} 作用 \mathbf{W}^* , 得到 \mathbf{x} 在子空间 $\text{Range}(\mathbf{W})$ 上的坐标向量; 然后再对该坐标向量作用 \mathbf{W} , 将其映射回子空间 $\text{Range}(\mathbf{W})$. 这样, 我们就得到了 \mathbf{x} 在子空间 $\text{Range}(\mathbf{W})$ 上的投影。

对于这个满秩分解，我们还有一个结论是： $\mathbf{W}^* \mathbf{W} = \mathbf{I}_r$. 这个结论也是显然的——因为 \mathbf{W} 的列向量构成了子空间 $\text{Range}(\mathbf{W})$ 的一组标准正交基，因此它们两两正交，自然有 $\mathbf{W}^* \mathbf{W} = \mathbf{I}_r$.

如果我们记 $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_r]$, 则投影算子 \mathbf{P} 可以写成

$$\mathbf{P} = \sum_{i=1}^r \mathbf{w}_i \mathbf{w}_i^*.$$

我们可以这样理解这个式子：子空间 $\text{Range}(\mathbf{W})$ 可以分解为 r 个方向的子空间的直和

$$\text{Range}(\mathbf{W}) = \text{span}(\mathbf{w}_1) \oplus \text{span}(\mathbf{w}_2) \oplus \dots \oplus \text{span}(\mathbf{w}_r).$$

因此，到 $\text{Range}(\mathbf{W})$ 上的投影可以分解为到每个方向子空间上的投影之和，而每个方向子空间上的投影正好由 $\mathbf{w}_i \mathbf{w}_i^*$ 给出。这里的 r 也被称为投影算子的正交秩，它和我们之前所学习的行秩、列秩、矩阵秩是一样的。我们也没必要去区分它们。

3 奇异值分解

接下来，我们考虑更一般的线性映射 $T : \mathcal{V} \rightarrow \mathcal{W}$. 一个自然的问题是：我们能否找到 \mathcal{V} 和 \mathcal{W} 的标准正交基，使得 T 在这两组基上的表示矩阵具有最简单的形式？

虽然 T 本身不是一个算子，但是我们仍然可以定义它的伴随算子 $T^* : \mathcal{W} \rightarrow \mathcal{V}$ ，使得对于任意的 $\mathbf{v} \in \mathcal{V}$ 和 $\mathbf{w} \in \mathcal{W}$ ，都有

$$\langle T(\mathbf{v}), \mathbf{w} \rangle_{\mathcal{W}} = \langle \mathbf{v}, T^*(\mathbf{w}) \rangle_{\mathcal{V}}.$$

一个关键的观察是：算子 $T^* T : \mathcal{V} \rightarrow \mathcal{V}$ 是 \mathcal{V} 上的一个自伴算子，而算子 $T T^* : \mathcal{W} \rightarrow \mathcal{W}$ 则是 \mathcal{W} 上的一个自伴算子。我们只对前者进行证明，后者的证明类似。

$$\langle T^* T(\mathbf{v}_1), \mathbf{v}_2 \rangle_{\mathcal{V}} = \langle T(\mathbf{v}_1), T(\mathbf{v}_2) \rangle_{\mathcal{W}} = \langle \mathbf{v}_1, T^* T(\mathbf{v}_2) \rangle_{\mathcal{V}}.$$

因此， $T^* T$ 是自伴算子。同理， $T T^*$ 也是自伴算子。由此可见，我们可以利用自伴算子的性质来研究更一般的线性映射 $T : \mathcal{V} \rightarrow \mathcal{W}$. 进一步地，我们可以验证这两个算子是半正定自伴算子。我们只证明前者。对任意的 $\mathbf{v} \in \mathcal{V}$ ，都有

$$\langle T^* T(\mathbf{v}), \mathbf{v} \rangle_{\mathcal{V}} = \langle T(\mathbf{v}), T(\mathbf{v}) \rangle_{\mathcal{W}} = \|T(\mathbf{v})\|_{\mathcal{W}}^2 \geq 0.$$

另一方面，取 \mathcal{V} 的一组基 \mathbf{E} ，对任意向量 $\mathbf{v} \in \mathcal{V}$ ，假设其在基 \mathbf{E} 下的坐标向量为 $\boldsymbol{\alpha}$ ，并设 $T^* T$ 在 \mathbf{E} 上的表示矩阵为 \mathbf{A} ，则

$$\langle T^* T(\mathbf{v}), \mathbf{v} \rangle_{\mathcal{V}} = \boldsymbol{\alpha}^* \mathbf{A} \boldsymbol{\alpha}.$$

因此 $\boldsymbol{\alpha}^* \mathbf{A} \boldsymbol{\alpha} \geq 0$ ，因此 \mathbf{A} 是半正定矩阵。由此可见， $T^* T$ 是半正定自伴算子。同理， $T T^*$ 也是半正定自伴算子。

在之前的作业中，我们已经证明过 \mathbf{XY} 和 \mathbf{YX} 有相同的非零特征值（计重数），因此，我们合理地猜想： T^*T 和 TT^* 也有着相同的非零特征值。事实上，这个猜想也很好证。取 \mathcal{V} 的一组基 \mathbf{E} ($\text{rank}(\mathbf{E}) = n$) 和 \mathcal{W} 的一组基 \mathbf{F} ($\text{rank}(\mathbf{F}) = m$)，设 T 在基 \mathbf{E} 和 \mathbf{F} 上的表示矩阵为 $\mathbf{A} \in \mathbb{C}^{m \times n}$ 。于是， TT^* 在基 \mathbf{F} 上的表示矩阵为 \mathbf{AA}^* ，而 T^*T 在基 \mathbf{E} 上的表示矩阵为 $\mathbf{A}^*\mathbf{A}$ 。由之前的结论可知， \mathbf{AA}^* 和 $\mathbf{A}^*\mathbf{A}$ 有相同的非零特征值（计重数）。因此， T^*T 和 TT^* 也有相同的非零特征值（计重数）。

综合以上的讨论，我们知道：算子 $T^*T : \mathcal{V} \rightarrow \mathcal{V}$ 和算子 $TT^* : \mathcal{W} \rightarrow \mathcal{W}$ 都是半正定自伴算子，并且它们有相同的非零特征值（计重数）。我们记这些非零特征值为

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0,$$

其中 r 是这两个算子的秩。由自伴算子的谱分解定理，我们有

$$\begin{aligned} \mathbf{A}^*\mathbf{A} &= \mathbf{V} \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_r^2 \\ & & & & \mathbf{0} \end{bmatrix} \mathbf{V}^*, \\ \mathbf{AA}^* &= \mathbf{U} \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_r^2 \\ & & & & \mathbf{0} \end{bmatrix} \mathbf{U}^*, \end{aligned}$$

那么，由上面这两个式子，我们自然地能够猜想： \mathbf{A} 是否可以写成

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*, \quad \Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \\ & & & & \mathbf{0} \end{bmatrix} \in \mathbb{C}^{m \times n}.$$

这就是奇异值分解 (Singular Value Decomposition, SVD) 的基本形式与几何动机（至少几何动机我是这样理解的）。在证明它之前，我们先来看看如果有这个分解，我们能得到什么结论。假设存在奇异值分解 $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$ ，则有

$$T(\mathbf{EV}) = (\mathbf{FU})\Sigma.$$

这也就是说，我们可以在线性空间 \mathcal{V} 和 \mathcal{W} 中分别找到一组标准正交基，使得 T 在这两组基下的表示矩阵是一个长方形的对角阵——只有对角线上有非零元素，且这些非零元素都是实数且非负的。

接下来，我们就来证明奇异值分解的存在性。

定理 3.1 (矩阵版本的奇异值分解). 设矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 则存在酉矩阵 $\mathbf{U} \in \mathbb{C}^{m \times m}$ 和酉矩阵 $\mathbf{V} \in \mathbb{C}^{n \times n}$, 使得

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*, \quad \Sigma = \begin{bmatrix} \Sigma_* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{C}^{m \times n},$$

其中 $\Sigma_* = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ 是一个对角矩阵, 且 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$.

定理 3.2 (映射版本的奇异值分解). 设 $T : \mathcal{V} \rightarrow \mathcal{W}$ 是有限维内积空间之间的线性映射, $\dim \mathcal{V} = n$, $\dim \mathcal{W} = m$, $\dim \text{Range}(T) = r$, 则存在

- \mathcal{V} 的一组标准正交基 $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$;
- \mathcal{W} 的一组标准正交基 $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$;
- 正实数 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$,

使得对任意 $\mathbf{x} \in \mathcal{V}$, 都有

$$T(\mathbf{x}) = \sum_{i=1}^r \sigma_i \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{u}_i.$$

证明. 由于我们刚刚做了比较多的几何铺垫, 这里我们直接给出映射版本的证明。稍后我们再补充邵老师课上讲的矩阵版本的证明。

设 $T^*T : \mathcal{V} \rightarrow \mathcal{V}$ 的非零特征值 (计重数) 为 $\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2$, 对应的单位特征向量为 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$. 则由自伴算子的谱分解定理可知, $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ 可以扩充为 \mathcal{V} 的一组标准正交基 $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$. 其中 $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ 是零特征值对应的单位特征向量。

接下来我们定义 $\mathbf{u}_i = T(\mathbf{v}_i)/\sigma_i$, 则

$$\begin{aligned} \langle \mathbf{u}_i, \mathbf{u}_j \rangle &= \frac{1}{\sigma_i \sigma_j} \langle T(\mathbf{v}_i), T(\mathbf{v}_j) \rangle \\ &= \frac{1}{\sigma_i \sigma_j} \langle \mathbf{v}_i, T^*T(\mathbf{v}_j) \rangle \\ &= \frac{\sigma_j^2}{\sigma_i \sigma_j} \langle \mathbf{v}_i, \mathbf{v}_j \rangle \\ &= \delta_{i,j}, \quad i, j = 1, 2, \dots, r. \end{aligned}$$

因此, $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ 是 \mathcal{W} 中的一组标准正交向量组。我们可以将它扩充为 \mathcal{W} 的一组标准正交基 $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$.

接下来, 我们就来验证奇异值分解的表达式。首先, 我们要先看看映射对基 $\mathbf{v}_1, \dots, \mathbf{v}_n$ 是怎么作用的。对于 $i = 1, 2, \dots, r$, 有

$$T(\mathbf{v}_i) = \sigma_i \mathbf{u}_i.$$

而对于 $i = r + 1, \dots, n$, 我们自然猜想它会被映射到零向量。我们可以通过范数计算来证明这一点:

$$\begin{aligned}\langle T(\mathbf{v}_i), T(\mathbf{v}_i) \rangle &= \langle \mathbf{v}_i, T^*T(\mathbf{v}_i) \rangle \\ &= \langle \mathbf{v}_i, \mathbf{0} \rangle \\ &= 0, \quad \forall i \in \{r + 1, \dots, n\}.\end{aligned}$$

因此 $T(\mathbf{v}_i) = \mathbf{0}$.

现在, 我们考虑映射 T 对任意向量 $\mathbf{x} \in \mathcal{V}$ 的作用。设 \mathbf{x} 在基 $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ 下的坐标向量为 $\boldsymbol{\alpha}$, 则

$$\begin{aligned}T(\mathbf{x}) &= T\left(\sum_{i=1}^n \alpha_i \mathbf{v}_i\right) \\ &= \sum_{i=1}^n \alpha_i T(\mathbf{v}_i) \\ &= \sum_{i=1}^r \alpha_i \sigma_i \mathbf{u}_i + \sum_{i=r+1}^n \alpha_i \mathbf{0} \\ &= \sum_{i=1}^r \sigma_i \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{u}_i.\end{aligned}$$

这正是我们要证明的分解形式。

接下来我们考虑把矩阵写成矩阵形式。由上面的证明我们很容易得到

$$T(\mathbf{V}) = \mathbf{U}\boldsymbol{\Sigma}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{C}^{m \times n}.$$

这也就是说, 我们能够在 \mathcal{V} 和 \mathcal{W} 中分别找到一组标准正交基, 使得 T 在这两组基下的表示矩阵对角阵。因此, 若 \mathbf{A} 是 T 在某组标准正交基下的表示矩阵, 则有

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*.$$

□

接下来我们给出一种更巧妙的证明方法, 也是邵老师课上讲的方法——这种证法需要一些注意力。

证明. 考虑 $\|\mathbf{A}\|_2$ 的 2-范数:

$$\|\mathbf{A}\|_2 = \max_{\|\mathbf{v}\|_2=1} \|\mathbf{A}\mathbf{v}\|_2.$$

显然存在某个单位向量 \mathbf{v}_1 , 使得 $\|\mathbf{A}\mathbf{v}_1\|_2 = \|\mathbf{A}\|_2$.

下面我们设

$$\mathbf{u}_1 = \frac{\mathbf{A}\mathbf{v}_1}{\|\mathbf{A}\mathbf{v}_1\|_2} = \frac{\mathbf{A}\mathbf{v}_1}{\|\mathbf{A}\|_2},$$

则 \mathbf{u}_1 是 $\mathbf{A}\mathbf{v}_1$ 方向上的单位向量。

由基扩张定理，我们可以找到 \mathcal{V} 的一组基 $\mathbf{V} = [\mathbf{v}_1, \star]$ 和 \mathcal{W} 的一组基 $\mathbf{U} = [\mathbf{u}_1, \star]$ 。考虑

$$\mathbf{AV} = \mathbf{U} \begin{bmatrix} \|\mathbf{A}\|_2 & \star \\ \mathbf{0} & \star \end{bmatrix}$$

我们的目标是把上面的矩阵变成对角阵。我们考虑使用归纳法。因此，我们只需要证明上面的 \star 是零向量即可。

考虑

$$\begin{aligned} \|\mathbf{A}\|_2 &= \|\mathbf{U}^* \mathbf{AV}\|_2 \\ &= \left\| \begin{bmatrix} \|\mathbf{A}\|_2 & \star \\ \mathbf{0} & \star \end{bmatrix} \right\|_2 \\ &= \max_{\|\mathbf{x}\|_2=1} \left\| \begin{bmatrix} \|\mathbf{A}\|_2 & \star \\ \mathbf{0} & \star \end{bmatrix} \mathbf{x} \right\|_2 \\ &= \max_{\|\mathbf{y}\|_2=1} \left\| \begin{bmatrix} \|\mathbf{A}\|_2 & \mathbf{0} \\ \star & \star \end{bmatrix} \mathbf{y} \right\|_2 \\ &\geq \left\| \begin{bmatrix} \|\mathbf{A}\|_2 & \mathbf{0} \\ \star & \star \end{bmatrix} \mathbf{e}_1 \right\|_2 \\ &= \left\| \begin{bmatrix} \|\mathbf{A}\|_2 & \star \end{bmatrix} \right\|_2 \\ &\geq \|\mathbf{A}\|_2. \end{aligned}$$

第四个等号成立的原因是：矩阵的 2-范数等于其转置的 2-范数。由上面的推导我们可以看出，等号成立的唯一可能性是 $\star = \mathbf{0}$ 。这样，我们就可以把矩阵降阶了。继续使用归纳法，我们就可以完成奇异值分解的证明。 \square

接下来，我们可以导出奇异值分解的精简形式。

定理 3.3. 若 $\mathbf{A} \in \mathbb{C}^{m \times n}$ ，并且 $m \geq n$, $\text{rank}(\mathbf{A}) = r$ ，那么存在正交归一的矩阵 $\mathbf{U} \in \mathbb{C}^{m \times r}$, $\mathbf{V} \in \mathbb{C}^{n \times r}$ ，以及正定对角阵 $\Sigma \in \mathbb{R}^{r \times r}$ 使得

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^*.$$

若 \mathbf{A} 是实矩阵则 \mathbf{U}, \mathbf{V} 也可取实矩阵。

奇异值分解还有第三种形式：

定理 3.4. 若 $\mathbf{A} \in \mathbb{C}^{m \times n}$ ，并且 $r = \text{rank}(\mathbf{A})$ ，那么存在正交归一的向量组 $\mathbf{u}_1, \dots, \mathbf{u}_r \in \mathbb{C}^m$ 和 $\mathbf{v}_1, \dots, \mathbf{v}_r \in \mathbb{C}^n$ ，以及正实数

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0,$$

使得

$$\mathbf{A} = \sum_{i=1}^r \mathbf{u}_i \sigma_i \mathbf{v}_i^*.$$

这个形式实际上就是把矩阵 \mathbf{A} 分解为 r 个秩一矩阵的和。由于 \mathbf{u}_i 和 \mathbf{v}_i 是正交归一的向量组，我们把 r 称为正交秩——在线性代数中它和矩阵的秩是一样的。

为方便描述，我们规定一些术语：

1. 奇异值：我们称 $\sigma_1, \dots, \sigma_r, 0, \dots, 0$ 为 \mathbf{A} 的奇异值。习惯上常常把奇异值按降序排列，把第 k 大的奇异值记为 $\sigma_k(\mathbf{A})$ 。
2. 奇异向量：我们把 \mathbf{u}_k 称为左奇异向量，把 \mathbf{v}_k 称为右奇异向量。

接下来，我们来讨论一下奇异值和奇异向量之间的关系。

定理 3.5.

$$\begin{cases} \mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i, \\ \mathbf{A}^* \mathbf{u}_i = \sigma_i \mathbf{v}_i. \end{cases}$$

证明. 这一个结论我们在映射版本的证明中已经证明过了。现在我们从矩阵的视角来看看这个结论。由奇异值分解可知

$$\mathbf{A}\mathbf{V} = \mathbf{U}\Sigma$$

$$\mathbf{A}^* \mathbf{U} = \mathbf{V}\Sigma.$$

两边取第 i 列即得结论。 \square

不难发现，上面这个结论的直接推论是 \mathbf{v}_i 是矩阵 $\mathbf{A}^*\mathbf{A}$ 的特征向量，且对应的特征值为 σ_i^2 ；同理， \mathbf{u}_i 是矩阵 $\mathbf{A}\mathbf{A}^*$ 的特征向量，且对应的特征值也为 σ_i^2 . 这再次体现了代数观点和几何观点的统一——奇异值分解既可以从矩阵本身的结构出发，得到一系列几何结论；也可以从几何角度出发，得到矩阵的分解形式。

当然，上面这个结论还有一个等价的形式：

$$\begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^* & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix} = \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix} \sigma_i.$$

由于 $\|\cdot\|_2$ 和 $\|\cdot\|_F$ 具有酉不变性，因此

$$\begin{aligned} \|\mathbf{A}\|_2 &= \|\Sigma\|_2 = \sigma_1 \\ \|\mathbf{A}\|_F &= \|\Sigma\|_F = \sqrt{\sum_{i=1}^n \sigma_i^2} \end{aligned}$$

此外，我们还可以定义一个新的范数 (nuclear norm):

$$\|\mathbf{A}\|_{\star} = \sum_{i=1}^n \sigma_i.$$

它为什么是范数呢？正定性和齐次性显然，三角不等式的证明比较麻烦，需要用到 Weyl 不等式，邵老师应该是下次课讲。因此在这里我们就不展开讲了。

最后，我们以著名的线性代数基本定理来结束本讲的内容。

定理 3.6 (线性代数基本定理). 若 \mathbf{A} 的奇异值分解为

$$\mathbf{A} = [\mathbf{U}_1, \mathbf{U}_2] \begin{bmatrix} \Sigma_{\star} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^{\star} \\ \mathbf{V}_2^{\star} \end{bmatrix} = \mathbf{U}_1 \Sigma_{\star} \mathbf{V}_1^{\star},$$

其中 Σ_{\star} 是正定对角阵，那么 $\mathbf{U}_1 \mathbf{U}_1^{\star}$, $\mathbf{U}_2 \mathbf{U}_2^{\star}$, $\mathbf{V}_1 \mathbf{V}_1^{\star}$, $\mathbf{V}_2 \mathbf{V}_2^{\star}$ 都是投影算子，并且

- $\text{Range}(\mathbf{A}) = \text{Range}(\mathbf{U}_1 \mathbf{U}_1^{\star})$;
- $\text{Range}(\mathbf{A}^{\star}) = \text{Range}(\mathbf{V}_1 \mathbf{V}_1^{\star})$;
- $\text{Ker}(\mathbf{A}) = \text{Range}(\mathbf{A}^{\star})^{\perp} = \text{Range}(\mathbf{V}_2 \mathbf{V}_2^{\star})$;
- $\text{Ker}(\mathbf{A}^{\star}) = \text{Range}(\mathbf{A})^{\perp} = \text{Range}(\mathbf{U}_2 \mathbf{U}_2^{\star})$.

证明. 我们只证明第一个结论，其他结论的证明类似。由奇异值分解可知

$$\mathbf{A} = \mathbf{U}_1 \Sigma_{\star} \mathbf{V}_1^{\star}.$$

因此，

$$\text{Range}(\mathbf{A}) \subseteq \text{Range}(\mathbf{U}_1).$$

另一方面，取任意的 $\mathbf{y} \in \text{Range}(\mathbf{U}_1)$ ，则存在某个向量 \mathbf{x} 使得

$$\mathbf{y} = \mathbf{U}_1 \mathbf{x} = \mathbf{A} (\mathbf{V}_1 \Sigma_{\star}^{-1} \mathbf{x}) \in \text{Range}(\mathbf{A}).$$

因此， $\text{Range}(\mathbf{U}_1) \subseteq \text{Range}(\mathbf{A})$. 综上所述，我们有

$$\text{Range}(\mathbf{A}) = \text{Range}(\mathbf{U}_1).$$

□

这个定理将矩阵的四个基本子空间与 SVD 中的正交基精确对应起来：

1. $\text{Range}(\mathbf{A})$ 由 \mathbf{U}_1 的列向量张成，是 \mathbf{A} 的列空间；

2. $\text{Range}(\mathbf{A}^*)$ 由 \mathbf{V}_1 的列向量张成, 是 \mathbf{A} 的行空间;
3. $\text{Ker}(\mathbf{A})$ 由 \mathbf{V}_2 的列向量张成, 是 \mathbf{A} 的零空间;
4. $\text{Ker}(\mathbf{A}^*)$ 由 \mathbf{U}_2 的列向量张成, 是 \mathbf{A} 的左零空间。

由线性代数基本定理, 我们可以给出秩-零化度定理的几何版本的描述:

$$\begin{aligned}\text{Range}(\mathbf{A}) \oplus \text{Ker}(\mathbf{A}) &= \text{Range}(\mathbf{U}_1) \oplus \text{Range}(\mathbf{V}_2) = \mathbb{C}^n, \\ \text{Range}(\mathbf{A}^*) \oplus \text{Ker}(\mathbf{A}^*) &= \text{Range}(\mathbf{V}_1) \oplus \text{Range}(\mathbf{U}_2) = \mathbb{C}^m.\end{aligned}$$

定理中的四个投影算子实际上就是将空间 \mathbb{C}^n 和 \mathbb{C}^m 分解为这两个直和的投影算子。这些投影算子在最小二乘问题中有着重要的应用。

Gilbert Strang 曾高度评价奇异值分解, 称其为“线性代数中最伟大的发现之一”。*Matrix Computations* 一书的作者、有着 *Prof. SVD* 美誉的 Gene H. Golub 教授也曾说过: “奇异值分解是数值线性代数中最重要的工具之一。”