

Lecture 11 奇异值分解、Hermite 矩阵的变分性质 — 2025.11.25

教授：邵美悦

Scribe: 路人甲

1 大纲

1. 奇异值分解（续）
2. Hermite 矩阵的变分性质

2 奇异值分解（续）

2.1 极分解

我们知道，任何一个复数 z 都可以写成极坐标形式： $z = \rho e^{i\theta} = \rho(\cos \theta + i \sin \theta)$ ，如果写成矩阵形式就是

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} := \mathbf{Q}\mathbf{H}.$$

其中， \mathbf{Q} 是一个正交矩阵， \mathbf{H} 是一个对称正定矩阵（特殊地，这里是对角阵）。这启发我们思考：对于任意一个矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ ，是否也存在类似的“极分解”？我们能否把一个矩阵 \mathbf{A} 也分解成一个伸缩部分和一个旋转部分？

定理 2.1 (极分解). 若 $\mathbf{A} \in \mathbb{C}^{m \times n}$ ，则存在列正交归一的矩阵 $\mathbf{Q} \in \mathbb{C}^{m \times n}$ 和 Hermite 半正定矩阵 $\mathbf{H} \in \mathbb{C}^{n \times n}$ ，使得 $\mathbf{A} = \mathbf{Q}\mathbf{H}$ ，并且当 \mathbf{A} 非奇异时，分解是唯一的。

证明. 非奇异方阵情形我们在上节课的作业里已经证明过了，不再赘述。对于一般情形，我们采用类似的思路。考虑 \mathbf{A} 的奇异值分解 $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ ，现在让我们来做一些代数变换：

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

于是我们定义 $\mathbf{Q} = \mathbf{U}\mathbf{V}^*$ ， $\mathbf{H} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^*$ ，则有

$$\mathbf{A} = \mathbf{Q}\mathbf{H}.$$

其中 \mathbf{H} 是 Hermite 半正定矩阵， \mathbf{Q} 是列正交归一阵。□

类似地，我们还可以写出另一种形式：

$$\mathbf{A} = (\mathbf{U}\mathbf{\Sigma}\mathbf{U}^*)(\mathbf{U}\mathbf{V}^*) = \tilde{\mathbf{H}}\mathbf{Q}.$$

可以发现，两种形式的酉因子是相同的，只是酉因子的位置不同罢了。

从几何角度看，极分解告诉我们：任何一个线性映射 \mathbf{A} 都可以看成一个旋转（或反射） \mathbf{Q} 与一个沿特征方向伸缩的 \mathbf{H} 的复合，复合的先后顺序并不重要。SVD 需要两个不同的旋转 \mathbf{U} 和 \mathbf{V}^* ，因此极分解更加精简，但代价是伸缩部分 \mathbf{H} 不再是对角阵。想象一下，假如我们有一个橡皮泥球体，极分解做的事情就是先整体旋转球体，然后沿固定方向挤压或拉伸；SVD 是先旋转到一个特定方向，挤压或拉伸，再旋转到另一个方向。

极分解有一系列优良的性质。接下来我们介绍两个应用。

考虑 Hermite 矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ ，它有谱分解 $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^*$ 。如果我们记 $\mathbf{\Lambda} = \text{diag}(\mathbf{\Lambda}_+, \mathbf{\Lambda}_-)$ ，并相应地对 \mathbf{Q} 作分块 $\mathbf{Q} = [\mathbf{Q}_+, \mathbf{Q}_-]$ ，我们有

$$\begin{aligned} \mathbf{A} &= [\mathbf{Q}_+, \mathbf{Q}_-] \begin{bmatrix} \mathbf{\Lambda}_+ & \\ & \mathbf{\Lambda}_- \end{bmatrix} \begin{bmatrix} \mathbf{Q}_+^* \\ \mathbf{Q}_-^* \end{bmatrix} \quad \text{谱分解 (EVD)} \\ &= [\mathbf{Q}_+, \mathbf{Q}_-] \begin{bmatrix} \mathbf{\Lambda}_+ & \\ & -\mathbf{\Lambda}_- \end{bmatrix} \begin{bmatrix} \mathbf{Q}_+^* \\ -\mathbf{Q}_-^* \end{bmatrix} \quad \text{SVD} \\ &= [\mathbf{Q}_+, \mathbf{Q}_-] \begin{bmatrix} \mathbf{Q}_+^* \\ -\mathbf{Q}_-^* \end{bmatrix} \left([\mathbf{Q}_+, -\mathbf{Q}_-] \begin{bmatrix} \mathbf{\Lambda}_+ & \\ & \mathbf{\Lambda}_- \end{bmatrix} \begin{bmatrix} \mathbf{Q}_+^* \\ -\mathbf{Q}_-^* \end{bmatrix} \right) \quad \text{极分解} \\ &:= \tilde{\mathbf{Q}}\mathbf{H} \end{aligned}$$

我们把 $\tilde{\mathbf{Q}}$ 展开，有

$$\tilde{\mathbf{Q}} = \mathbf{Q}_+\mathbf{Q}_+^* - \mathbf{Q}_-\mathbf{Q}_-^* \quad (1)$$

这说明 $\tilde{\mathbf{Q}}$ 可以拆分成两个正交投影的差。由于 \mathbf{Q}_+ 和 \mathbf{Q}_- 是正交的并且构成了全空间的一组正交基，因此两个投影算子的和算子应该是恒等算子，即

$$\mathbf{I} = \mathbf{Q}_+\mathbf{Q}_+^* + \mathbf{Q}_-\mathbf{Q}_-^* \quad (2)$$

将 (1) 和 (2) 联立，有

$$\begin{aligned} \mathbf{Q}_+\mathbf{Q}_+^* &= \frac{1}{2}(\tilde{\mathbf{Q}} + \mathbf{I}) \\ \mathbf{Q}_-\mathbf{Q}_-^* &= \frac{1}{2}(\mathbf{I} - \tilde{\mathbf{Q}}) \end{aligned}$$

这样，我们就利用极分解得到了分离 \mathbf{A} 的正特征子空间和负特征子空间的方法。细心的读者可以发现， $\tilde{\mathbf{Q}}$ 正是符号算子，它在正特征子空间上的作用为 $+1$ ，而在负特征子空间上的作用为 -1 。这个分解在数值计算中比直接计算谱分解要更稳定。

第二个应用是：极分解可以用于酉逼近。假如我们有一个接近正交归一阵的矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$ ，即我们有 $\mathbf{A}^* \mathbf{A} \approx \mathbf{I}$ 。现在我们想要找一个最接近 \mathbf{A} 的正交归一阵 \mathbf{Q}_1 去逼近 \mathbf{A} ，即我们要求解下面的优化问题：

$$\min_{\mathbf{Q}_1^* \mathbf{Q}_1 = \mathbf{I}} \|\mathbf{A} - \mathbf{Q}_1\|_{\text{F}}^2.$$

由于 $\|\cdot\|_{\text{F}}$ 是酉不变范数，我们考虑 \mathbf{A} 的奇异值分解 $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*$ ，我们有

$$\min_{\mathbf{Q}_1^* \mathbf{Q}_1 = \mathbf{I}} \|\mathbf{A} - \mathbf{Q}_1\|_{\text{F}}^2 = \min_{\mathbf{Q}_1^* \mathbf{Q}_1 = \mathbf{I}} \|\mathbf{U}(\mathbf{\Sigma} - \mathbf{Q}_2) \mathbf{V}^*\|_{\text{F}}^2 = \min_{\mathbf{Q}_2^* \mathbf{Q}_2 = \mathbf{I}} \|\mathbf{\Sigma} - \mathbf{Q}_2\|_{\text{F}}^2.$$

根据 Frobenius 范数的定义 $\|\mathbf{A}\|_{\text{F}}^2 = \text{tr}(\mathbf{A}^* \mathbf{A})$ ，我们有

$$\min_{\mathbf{Q}_2^* \mathbf{Q}_2 = \mathbf{I}} \|\mathbf{\Sigma} - \mathbf{Q}_2\|_{\text{F}}^2 = \min_{\mathbf{Q}_2^* \mathbf{Q}_2 = \mathbf{I}} \text{tr}((\mathbf{\Sigma} - \mathbf{Q}_2)^* (\mathbf{\Sigma} - \mathbf{Q}_2)).$$

考虑到 $\mathbf{\Sigma}^2$ 和 $\mathbf{Q}_2^* \mathbf{Q}_2$ 是常量，对优化并无影响，于是上面的优化问题等价于

$$\begin{aligned} \max_{\mathbf{Q}_2^* \mathbf{Q}_2 = \mathbf{I}} \text{tr}(\mathbf{\Sigma}^* \mathbf{Q}_2 + \mathbf{Q}_2^* \mathbf{\Sigma}) &= \max_{\mathbf{Q}_2^* \mathbf{Q}_2 = \mathbf{I}} \sum_{i=1}^n (\sigma_i q_{i,i} + \sigma_i \overline{q_{i,i}}) \\ &= \max_{\mathbf{Q}_2^* \mathbf{Q}_2 = \mathbf{I}} \sum_{i=1}^n 2\sigma_i \text{Re}(q_{i,i}) \\ &\leq \sum_{i=1}^n 2\sigma_i |q_{i,i}| \\ &\leq \sum_{i=1}^n 2\sigma_i \\ &= 2 \sum_{i=1}^n \sigma_i. \end{aligned}$$

另一方面，取 $\mathbf{Q}_2 = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}$ ，我们有

$$\max_{\mathbf{Q}_2^* \mathbf{Q}_2 = \mathbf{I}} \text{tr}(\mathbf{\Sigma}^* \mathbf{Q}_2 + \mathbf{Q}_2^* \mathbf{\Sigma}) \geq 2 \sum_{i=1}^n \sigma_i.$$

因此

$$\max_{\mathbf{Q}_2^* \mathbf{Q}_2 = \mathbf{I}} \text{tr}(\mathbf{\Sigma}^* \mathbf{Q}_2 + \mathbf{Q}_2^* \mathbf{\Sigma}) = 2 \sum_{i=1}^n \sigma_i.$$

当且仅当 $\mathbf{Q}_2 = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}$ 时等号成立。此时 $\mathbf{Q}_1 = \mathbf{U} \mathbf{Q}_2 \mathbf{V}^*$ ，它正是极分解的酉因子的精简形式！因此，这个结论告诉我们，极分解产生的酉因子是矩阵的最佳酉逼近。

2.2 Moore–Penrose 广义逆

我们知道，对于可逆矩阵 \mathbf{A} ，存在唯一的逆矩阵 \mathbf{A}^{-1} 满足：

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

然而，虽然可逆矩阵是稠密的，但在实际问题中我们却常常遇到奇异矩阵、长方矩阵、秩亏损矩阵——这些矩阵不存在逆矩阵——那么我们能否为它们定义一种“广义”的逆，使其在可逆时退化为通常的逆，同时在不可逆时保持逆矩阵的一些优良性质呢？接下来，我们就着手对“逆”这个概念进行推广。

首先我们要明确一些推广的原则。对于一个可逆线性变换 $\mathbf{A} : \mathbb{C}^n \rightarrow \mathbb{C}^n$ ，

- 它是双射；
- 它存在唯一的逆变换 \mathbf{A}^{-1} ；
- 逆变换满足 $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ 。

对于不可逆或长方矩阵，情况就复杂了。考虑矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$ ，

- 像空间 $\text{Range}(\mathbf{A}) \subset \mathbb{C}^m$ ；
- 核空间 $\text{Ker}(\mathbf{A}) \subset \mathbb{C}^n$ 。

我们希望找到一个广义逆 \mathbf{A}^\dagger ，使得它在 $\text{Range}(\mathbf{A})$ 上的行为类似于逆，而在 $\text{Ker}(\mathbf{A})$ 上的行为要有合理的定义。

Moore 和 Penrose 提出了一个精妙的定义。Moore 说，如果我们考虑 \mathbf{A} 的奇异值分解

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \Sigma_\star & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^\star$$

那么我们定义其广义逆为

$$\mathbf{A}^\dagger = \mathbf{V} \begin{bmatrix} \Sigma_\star^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^\star.$$

Penrose 则用一个方程组刻画了这个广义逆的行为：存在唯一解 $\mathbf{X} = \mathbf{A}^\dagger$ 满足

$$\mathbf{A}\mathbf{X}\mathbf{A} = \mathbf{A} \tag{3}$$

$$\mathbf{X}\mathbf{A}\mathbf{X} = \mathbf{X} \tag{4}$$

$$(\mathbf{A}\mathbf{X})^\star = \mathbf{A}\mathbf{X} \tag{5}$$

$$(\mathbf{X}\mathbf{A})^\star = \mathbf{X}\mathbf{A} \tag{6}$$

接下来我们就来证明这个结论。

证明. 存在性显然, 我们直接代入

$$\mathbf{X} = \mathbf{V} \begin{bmatrix} \Sigma_{\star}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^{\star}$$

到上面四个方程组中, 很容易就能够验证结论成立。

下面我们证明唯一性。假设还有另一个矩阵 \mathbf{Y} 满足上面的四个方程组。下面我们说明 $\mathbf{X} = \mathbf{Y}$ 。

由 (4) 我们有

$$\mathbf{X} = \mathbf{X} \mathbf{A} \mathbf{X}.$$

由 (3) 知, $\mathbf{A} = \mathbf{A} \mathbf{X} \mathbf{A} = \mathbf{A} \mathbf{Y} \mathbf{A}$, 因此

$$\mathbf{X} = \mathbf{X}(\mathbf{A} \mathbf{Y} \mathbf{A}) \mathbf{X} = \mathbf{X}(\mathbf{A} \mathbf{Y})(\mathbf{A} \mathbf{X}).$$

由 (5) 知, $(\mathbf{A} \mathbf{X})^{\star} = \mathbf{A} \mathbf{X}$, $(\mathbf{A} \mathbf{Y})^{\star} = \mathbf{A} \mathbf{Y}$, 因此

$$\mathbf{X} = \mathbf{X}(\mathbf{A} \mathbf{Y})^{\star}(\mathbf{A} \mathbf{X})^{\star} = \mathbf{X} \mathbf{Y}^{\star}(\mathbf{A}^{\star} \mathbf{X}^{\star} \mathbf{A}^{\star}).$$

再由 (3) 可知 $\mathbf{A}^{\star} \mathbf{X}^{\star} \mathbf{A}^{\star} = \mathbf{A}^{\star}$, 所以

$$\mathbf{X} = \mathbf{X} \mathbf{Y}^{\star} \mathbf{A}^{\star} = \mathbf{X}(\mathbf{A} \mathbf{Y})^{\star}.$$

由 (5) 有 $(\mathbf{A} \mathbf{Y})^{\star} = \mathbf{A} \mathbf{Y}$, 因此

$$\mathbf{X} = \mathbf{X} \mathbf{A} \mathbf{Y}.$$

另一方面, 我们从 \mathbf{Y} 出发, 也能得到类似的结论:

$$\begin{aligned} \mathbf{Y} &= \mathbf{Y} \mathbf{A} \mathbf{Y} = \mathbf{Y} \mathbf{A} \mathbf{X} \mathbf{A} \mathbf{Y} \\ &= (\mathbf{Y} \mathbf{A})^{\star}(\mathbf{X} \mathbf{A})^{\star} \mathbf{Y} \\ &= (\mathbf{A}^{\star} \mathbf{Y}^{\star} \mathbf{A}^{\star}) \mathbf{X}^{\star} \mathbf{Y} \\ &= (\mathbf{X} \mathbf{A})^{\star} \mathbf{Y} \\ &= \mathbf{X} \mathbf{A} \mathbf{Y}. \end{aligned}$$

因此 $\mathbf{X} = \mathbf{Y}$. 因此 Penrose 方程组的解唯一。 □

或许到现在, 读者还是不明白为什么要这样定义广义逆——这些方程看上去好像和我们之前提到的“原则”一点关系也没有。现在, 我们就来逐一解读这几个方程。

1. 方程 (3) 说明了 $\mathbf{A} \mathbf{A}^{\dagger}$ 在 $\text{Range}(\mathbf{A})$ 上是恒等算子——这是 $\mathbf{A} \mathbf{A}^{-1} = \mathbf{I}$ 的推广。同时, 这个方程还说明了 $\mathbf{A} \mathbf{A}^{\dagger}$ 是幂等算子。

2. 方程 (4) 说明了 $\mathbf{A}^\dagger \mathbf{A}$ 在 $\text{Range}(\mathbf{A}^\dagger)$ 上是恒等算子——这是 $\mathbf{A}^{-1} \mathbf{A} = \mathbf{I}$ 的推广。同时，这个方程还说明了 $\mathbf{A}^\dagger \mathbf{A}$ 是幂等算子。
3. 方程 (5) 说明了 $\mathbf{A} \mathbf{A}^\dagger$ 是自伴算子。结合第一条，易见 $\mathbf{A} \mathbf{A}^\dagger$ 是一个投影算子。
4. 方程 (6) 说明了 $\mathbf{A}^\dagger \mathbf{A}$ 是自伴算子。结合第二条，易见 $\mathbf{A}^\dagger \mathbf{A}$ 是一个投影算子。

既然 $\mathbf{A} \mathbf{A}^\dagger$ 和 $\mathbf{A}^\dagger \mathbf{A}$ 都是投影算子，那么它们分别是到哪个空间的投影算子呢？由于 $\mathbf{A} \mathbf{A}^\dagger$ 在 $\text{Range}(\mathbf{A})$ 上是恒等算子，显然它是到 $\text{Range}(\mathbf{A})$ 上的投影算子——不然也不会把 \mathbf{A} 原封不动地映到 \mathbf{A} 。同理， $\mathbf{A}^\dagger \mathbf{A}$ 是到 $\text{Range}(\mathbf{A}^\dagger)$ 上的投影算子。

为了更清楚地洞见这两个投影算子的行为，我们把 \mathbf{A}^\dagger 的显式解代入，有

$$\begin{aligned}\mathbf{A} \mathbf{A}^\dagger &= \mathbf{U} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^\star = \mathbf{U}_1 \mathbf{U}_1^\star \\ \mathbf{A}^\dagger \mathbf{A} &= \mathbf{V} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^\star = \mathbf{V}_1 \mathbf{V}_1^\star\end{aligned}$$

在上节课我们说过， $\mathbf{U}_1 \mathbf{U}_1^\star$ 是到 $\text{Range}(\mathbf{A})$ 的投影算子，因此 $\mathbf{A} \mathbf{A}^\dagger$ 是到 $\text{Range}(\mathbf{A})$ 的投影算子； $\mathbf{V}_1 \mathbf{V}_1^\star$ 是到 $\text{Range}(\mathbf{A}^\star)$ 的投影算子，因此 $\mathbf{A}^\dagger \mathbf{A}$ 是到 $\text{Range}(\mathbf{A}^\star)$ 的投影算子。因此， $\text{Range}(\mathbf{A}^\star) = \text{Range}(\mathbf{A}^\dagger)$ ， \mathbf{A}^\star 与 \mathbf{A}^\dagger 必然存在着某种内在联系——这一联系我们稍后再进行揭示。

于是，到核空间 $\text{Ker}(\mathbf{A})$ 的投影算子可以表示为 $\mathbf{V}_2 \mathbf{V}_2^\star = \mathbf{I} - \mathbf{A}^\dagger \mathbf{A}$ ，到 $\text{Ker}(\mathbf{A}^\star)$ 的投影算子可以表示为 $\mathbf{U}_2 \mathbf{U}_2^\star = \mathbf{I} - \mathbf{A} \mathbf{A}^\dagger$ 。这几个空间和投影算子的关系可以用一个表格表示：

空间	投影算子	等价空间
$\text{Range}(\mathbf{A})$	$\mathbf{U}_1 \mathbf{U}_1^\star = \mathbf{A} \mathbf{A}^\dagger$	$\text{Range}(\mathbf{U}_1) = \text{Range}(\mathbf{A} \mathbf{A}^\dagger)$
$\text{Range}(\mathbf{A}^\star)$	$\mathbf{V}_1 \mathbf{V}_1^\star = \mathbf{A}^\dagger \mathbf{A}$	$\text{Range}(\mathbf{V}_1) = \text{Range}(\mathbf{A}^\dagger \mathbf{A})$
$\text{Ker}(\mathbf{A})$	$\mathbf{V}_2 \mathbf{V}_2^\star = \mathbf{I} - \mathbf{A}^\dagger \mathbf{A}$	$\text{Range}(\mathbf{V}_2) = \text{Range}(\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})$
$\text{Ker}(\mathbf{A}^\star)$	$\mathbf{U}_2 \mathbf{U}_2^\star = \mathbf{I} - \mathbf{A} \mathbf{A}^\dagger$	$\text{Range}(\mathbf{U}_2) = \text{Range}(\mathbf{I} - \mathbf{A} \mathbf{A}^\dagger)$

现在，我们换一个角度来解释为什么 $\text{Range}(\mathbf{A}^\dagger) = \text{Range}(\mathbf{A}^\star)$ 。我们知道，奇异值分解有算子形式。我们利用算子形式的奇异值分解表示 \mathbf{A}^\dagger 和 \mathbf{A}^\star ，有

$$\begin{aligned}\mathbf{A}^\star &= \sum_{i=1}^r \sigma_i \mathbf{v}_i \mathbf{u}_i^\star \\ \mathbf{A}^\dagger &= \sum_{i=1}^r \sigma_i^{-1} \mathbf{v}_i \mathbf{u}_i^\star\end{aligned}$$

因此这两者张成了相同的空间。既然如此，我们能不能用 \mathbf{A} 和 \mathbf{A}^\star 来表示 \mathbf{A}^\dagger 呢？我们暂且按下不表。

最小二乘问题 所谓的最小二乘问题，就是给定 $\mathbf{A} \in \mathbb{C}^{m \times n}$ 和 $\mathbf{b} \in \mathbb{C}^n$ ，求解 $\arg \min \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$ 。如果有多个解，我们希望求 2-范数最小的解。

从代数角度看，由于 2-范数具有酉不变性，我们可以先对原式进行一些正交变换。考虑 \mathbf{A} 的奇异值分解 $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ ，我们记 $\mathbf{y} = \mathbf{V}^*\mathbf{x}$ ，于是我们只要求解下面这个优化问题：

$$\min_{\mathbf{y}} \|\mathbf{\Sigma}\mathbf{y} - \mathbf{U}^*\mathbf{b}\|_2.$$

如果记 $\mathbf{U}^*\mathbf{b} = [\mathbf{c}_1^T, \mathbf{c}_2^T]^T$ ，我们有

$$\min_{\mathbf{y}} \|\mathbf{\Sigma}\mathbf{y} - \mathbf{U}^*\mathbf{b}\|_2 = \left\| \begin{bmatrix} \mathbf{\Sigma}_* \mathbf{y}_1 - \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} \right\|_2.$$

其中 \mathbf{c}_2 是一个常量——那么我们唯一能动手脚的就是 $\mathbf{\Sigma}_* \mathbf{y}_1 - \mathbf{c}_1$ 。由于 $\mathbf{\Sigma}_*$ 非奇异，显然我们可以找到一个 \mathbf{y} 使得 $\mathbf{\Sigma}_* \mathbf{y} - \mathbf{c}_1 = \mathbf{0}$ ，即 $\mathbf{y}_1 = \mathbf{\Sigma}_*^{-1} \mathbf{c}_1$ 。若我们要求解的范数最小，那么我们就有 $\mathbf{y}_2 = \mathbf{0}$ 。于是，

$$\mathbf{y} = \mathbf{\Sigma}^\dagger \mathbf{c}.$$

所以

$$\mathbf{x}_{LS} = \mathbf{V}\mathbf{y} = \mathbf{V}\mathbf{\Sigma}^\dagger \mathbf{U}^* \mathbf{b} = \mathbf{A}^\dagger \mathbf{b}.$$

从几何角度，这个结论将变得更加显然。我们可以将 \mathbf{b} 分解为在 $\text{Range}(\mathbf{A})$ 中的分量 \mathbf{b}_\parallel 和垂直于 $\text{Range}(\mathbf{A})$ 的分量 \mathbf{b}_\perp 。根据勾股定理，显然，当 $\mathbf{A}\mathbf{x} = \mathbf{b}_\parallel$ 时， $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$ 取到最小值，如图 1 所示。由于 $\mathbf{A}\mathbf{A}^\dagger$ 是到 $\text{Range}(\mathbf{A})$ 的投影算子， $\mathbf{b}_\parallel = \mathbf{A}\mathbf{A}^\dagger \mathbf{b}$ ，亦即 $\mathbf{x} = \mathbf{A}^\dagger \mathbf{b}$ 。

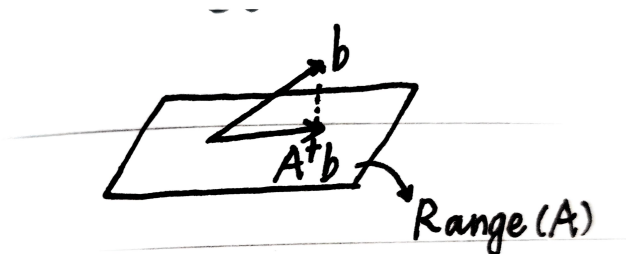


图 1: 最小二乘问题

建立起这个几何直观之后，我们还可以从另一个角度给出最小二乘解。我们需要 $\mathbf{b} - \mathbf{A}\mathbf{x}_{LS}$ 垂直于 $\text{Range}(\mathbf{A})$ ，这也就是

$$\mathbf{A}^*(\mathbf{A}\mathbf{x} - \mathbf{b}) = \mathbf{0}.$$

因此，

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

如果 \mathbf{A} 列满秩， $\mathbf{A}^* \mathbf{A}$ 就是正定矩阵， \mathbf{x} 有唯一解

$$\mathbf{x}_{LS} = (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^* \mathbf{b}.$$

对比我们之前的结果 $\mathbf{x}_{LS} = \mathbf{A}^\dagger \mathbf{b}$, 我们显然有

$$\mathbf{A}^\dagger = (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*.$$

因此, 当 \mathbf{A} 列满秩时, \mathbf{A}^\dagger 可由 \mathbf{A} 和 \mathbf{A}^* 表示。同理, 当 \mathbf{A} 行满秩时, 由于 \mathbf{A}^* 列满秩, 我们也可以得到 \mathbf{A}^\dagger 关于 \mathbf{A} 和 \mathbf{A}^* 的表达式:

$$\mathbf{A}^\dagger = \mathbf{A}^* (\mathbf{A} \mathbf{A}^*)^{-1}.$$

广义逆作为一种运算, 我们难免要研究它的运算性质。因此, 我们将以广义逆的运算性质结束本小节的讨论。

命题 2.2. $(\mathbf{A}^*)^\dagger = (\mathbf{A}^\dagger)^*$.

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

$$(\mathbf{A}^*)^\dagger = \mathbf{U} (\boldsymbol{\Sigma}^*)^\dagger \mathbf{V}^*$$

而 $(\boldsymbol{\Sigma}^*)^\dagger = (\boldsymbol{\Sigma}^\dagger)^*$ 是显然的, 因此

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

□

命题 2.3. $(\alpha \mathbf{A})^\dagger = \alpha^{-1} \mathbf{A}^\dagger, \alpha \neq 0$.

证明. 显然。

□

命题 2.4. $\mathbf{A} \in \mathbb{R}^{m \times n} \Leftrightarrow \mathbf{A}^\dagger \in \mathbb{R}^{n \times m}$.

证明. 实矩阵 \mathbf{A} 的奇异值分解可以取正交矩阵 \mathbf{U}, \mathbf{V} , 因此 $\mathbf{A}^\dagger \in \mathbb{R}^{n \times m}$ 是显然的。

□

命题 2.5. $(\mathbf{A}^\dagger)^\dagger = \mathbf{A}$.

证明. 显然。

□

命题 2.6. 若 $\mathbf{A} \in \mathbb{C}^{m \times r}$ 列满秩, $\mathbf{B} \in \mathbb{C}^{r \times n}$ 行满秩, 则 $(\mathbf{AB})^\dagger = \mathbf{B}^\dagger \mathbf{A}^\dagger$.

证明. 我们先证明一个引理: 对于列满秩矩阵 \mathbf{A} , 我们有 $\mathbf{A}^\dagger \mathbf{A} = \mathbf{I}_r$; 对于行满秩矩阵 \mathbf{B} , 我们有 $\mathbf{BB}^\dagger = \mathbf{I}_r$.

这个引理是非常显然的。由于 $\mathbf{A}^\dagger \mathbf{A} = \mathbf{V}_1 \mathbf{V}_1^*$, 且 \mathbf{A} 列满秩, 因此 \mathbf{V}_1 就是 \mathbf{V} , 因此 $\mathbf{A}^\dagger \mathbf{A} = \mathbf{I}_r$. \mathbf{B} 同理。

现在，我们就来逐一验证 Penrose 的四个方程。

方程 (3)：我们需要验证 $(AB)(B^\dagger A^\dagger)(AB) = AB$ ，由引理结论显然成立。

方程 (4)：我们需要验证 $(B^\dagger A^\dagger)(AB)(B^\dagger A^\dagger) = B^\dagger A^\dagger$ ，由引理结论显然成立。

方程 (5)：我们需要验证 $((AB)(B^\dagger A^\dagger))^* = (AB)(B^\dagger A^\dagger)$ 。由引理，我们只需证 $(AA^\dagger)^* = AA^\dagger$ ，是以结论显然。

方程 (6)：我们需要验证 $((B^\dagger A^\dagger)(AB))^* = (B^\dagger A^\dagger)(AB)$ 。由引理，我们只需证 $(B^\dagger B)^* = B^\dagger B$ ，是以结论显然。 \square

推论 2.7. $A \in \mathbb{Q}^{m \times n} \Leftrightarrow A^\dagger \in \mathbb{Q}^{n \times m}$.

证明. 若 A 列满秩或行满秩， A^\dagger 可由 A 和 A^* 表示，显然为有理矩阵。

若 A 亏秩，考虑满秩分解 $A = CR$ ，由上一个命题，我们知道 $A^\dagger = R^\dagger C^\dagger$ ，再由 R^\dagger, C^\dagger 均为有理矩阵可知 A^\dagger 也是有理矩阵。 \square

命题 2.8. 若 A 和 B 至少有一个是酉矩阵，则 $(AB)^\dagger = B^\dagger A^\dagger$ 。

证明. 不妨设 A 为酉矩阵，那么 $A^* = A^\dagger$ 。设 B 的奇异值分解为 $B = U\Sigma V^*$ ，于是

$$AB = (AU)\Sigma V^*.$$

于是

$$(AB)^\dagger = V\Sigma(AU)^* = (V\Sigma U^*)A^* = B^\dagger A^\dagger$$

B 为酉矩阵时同理。 \square

请注意，一般而言， $(AB)^\dagger \neq B^\dagger A^\dagger$ ， $(A^k)^\dagger \neq (A^\dagger)^k$ 。

命题 2.9. $(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$ 。

证明. 我们逐一验证 Penrose 方程组。

方程 (3)：

$$(A \otimes B)(A^\dagger \otimes B^\dagger)(A \otimes B) = (AA^\dagger A) \otimes (BB^\dagger B) = A \otimes B.$$

方程 (4)：

$$(A^\dagger \otimes B^\dagger)(A \otimes B)(A^\dagger \otimes B^\dagger) = (A^\dagger AA^\dagger) \otimes (B^\dagger BB^\dagger) = A^\dagger \otimes B^\dagger.$$

方程 (5)：

$$((A \otimes B)(A^\dagger \otimes B^\dagger))^* = (AA^\dagger \otimes BB^\dagger)^* = (AA^\dagger)^* \otimes (BB^\dagger)^* = AA^\dagger \otimes BB^\dagger.$$

方程 (6):

$$((A^\dagger \otimes B^\dagger)(A \otimes B))^* = (A^\dagger A \otimes B^\dagger B)^* = (A^\dagger A)^* \otimes (B^\dagger B)^* = A^\dagger A \otimes B^\dagger B.$$

□

3 Hermite 矩阵的变分性质

极分解告诉我们, 任意一个矩阵的深层次结构都是一个酉矩阵和 Hermite 半正定矩阵。Hermite 矩阵的核心特征就是它的实特征值和正交特征向量系。学过数值算法的读者都知道, 矩阵有病态良态之分。那么, 假如我们给矩阵施加一个小扰动, 它的特征值会怎么变化呢?

3.1 Rayleigh 商

让我们从一个简单的观察开始。对任意 $A \in \mathbb{C}^{n \times n}$, 它的特征方程为

$$Ax = x\lambda.$$

那么, 我们该怎么用其他量来表示 λ 呢? 考虑在两边同时左乘 x^* , 我们有

$$\lambda = \frac{x^* Ax}{x^* x}.$$

这启发我们定义下面这个量:

定义 3.1 (Rayleigh 商). 对任意矩阵 $A \in \mathbb{C}^{n \times n}$ 和非零向量 $x \in \mathbb{C}^n$, 我们定义 *Rayleigh 商*

$$R(A, x) = \frac{x^* Ax}{x^* x}.$$

显然, 若 $Ax = x\lambda$, 我们有 $R(A, x) = \lambda$.

Rayleigh 商有着清晰的几何意义和物理意义。在几何上, 它衡量了向量 x 在变换 A 下的平均伸缩率; 在振动系统中, 它代表频率; 在量子力学中, 它代表能量期望值。

你们学过量子力学吗?

—— Prof. 5eb0fe

定义 3.2 (数值域). $w(A) := \{R(A, x) : x \neq 0\}$ 称为 A 的数值域。

下面我们所有的讨论均假设 A 是 Hermite 矩阵, 并设特征值按递增顺序排列, 即 $\lambda_1(\cdot) \leq \dots \leq \lambda_n(\cdot)$, q_1, \dots, q_n 是对应的特征向量。

我们知道，Hermite 矩阵所有特征值都是实数。我们自然会好奇，它的数值域是否也是实数集的某个子集呢？正是如此。那么，Hermite 阵的数值域是否有界呢？答案同样是肯定的。现在，我们就来证明这两点。

定理 3.3 (Rayleigh 商定理). Hermite 阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 的数值域 $w(\mathbf{A}) = [\lambda_1(\mathbf{A}), \lambda_n(\mathbf{A})] \subset \mathbb{R}$.

证明. 我们先对问题做一个简化。考虑 \mathbf{A} 的谱分解 $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^*$ ，显见

$$R(\mathbf{Q}^*\mathbf{A}\mathbf{Q}, \mathbf{x}) = R(\mathbf{A}, \mathbf{Q}\mathbf{x}).$$

因此 $w(\mathbf{A}) = w(\mathbf{Q}^*\mathbf{A}\mathbf{Q}) = w(\mathbf{\Lambda})$.

由 $R(\mathbf{\Lambda}, \mathbf{x})$ 的定义易知 $w(\mathbf{\Lambda}) \in \mathbb{R}$. 不失一般性，我们设 $\|\mathbf{x}\|_2 = 1$ ，于是我们有

$$R(\mathbf{\Lambda}, \mathbf{x}) = \mathbf{x}^*\mathbf{\Lambda}\mathbf{x} = \sum_{i=1}^n \lambda_i |x_i|^2 \begin{cases} \geq \lambda_{\min} \sum_{i=1}^n |x_i|^2 = \lambda_1(\mathbf{A}) \\ \leq \lambda_{\max} \sum_{i=1}^n |x_i|^2 = \lambda_n(\mathbf{A}) \end{cases}$$

两个等号都能取到。由连续性，两个端点之间的值可以取遍，于是有 $R(\mathbf{A}, \mathbf{x}) = [\lambda_1(\mathbf{A}), \lambda_n(\mathbf{A})]$. □

另解：这是一个约束上的最值问题，可考虑 Lagrange 乘子法。

注：如果我们把 Rayleigh 商推广到广义 Rayleigh 商（即 $R(\mathbf{A}, \mathbf{B}, \mathbf{x}) = \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{B}\mathbf{x}}$ ，对应于广义特征值问题），我们也可以用 Lagrange 乘子法证明它的取值范围是 $[\lambda_{\min}, \lambda_{\max}]$. 其中 λ_{\min} 和 λ_{\max} 分别是最小和最大广义特征值。

有了这个定理，我们就能通过一个优化问题来刻画特征值的极值了。现在问题是：那些排在中间的特征值呢？

3.2 Courant–Fischer min-max 定理

为了捕捉排序在中间的特征值，我们需要引入约束优化的思想。直观上说：

- 第二大的特征值就是避开最大特征方向后能找到的最大值；
- 第三大的特征值就是避开前两个特征方向后能找到的最大值；
- 以此类推……

这一精妙的思想被表述为 Courant–Fischer min-max 定理。

定理 3.4 (Courant–Fischer min-max 定理).

$$\begin{aligned}\lambda_k(\mathbf{A}) &= \min_{\dim V=k} \max_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) \\ &= \max_{\dim V=n-k+1} \min_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x})\end{aligned}$$

我们暂且先不证明这个定理。从直观上来看，假如一个子空间 V 是 $n-1$ 维的，按我们的常识，只要 V 不包含 $\lambda_n(\mathbf{A})$ 对应的特征子空间，那么 $\max_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) = \lambda_{n-1}$ ；但倘若 V 包含 λ_n 对应的特征子空间，那么就有 $\max_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) = \lambda_n$ 。为了捕获 λ_{n-1} ，我们对所有 V 取最小值，我们就能抓到 λ_{n-1} 了。

另一方面，如果一个子空间 V 是 2 维的，按照常识，只要 V 包含 $\lambda_n(\mathbf{A}), \lambda_{n-1}(\mathbf{A})$ 对应的特征子空间，就有 $\min_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) = \lambda_{n-1}(\mathbf{A})$ ；若不包含任意一个特征空间，这个量就会比 $\lambda_{n-1}(\mathbf{A})$ 小。为了捕获 $\lambda_{n-1}(\mathbf{A})$ ，我们对所有的 V 取最大值，就能够抓到它了。

现在我们来证明 Courant–Fischer min-max 定理。

证明. 我们先证第一个等号。根据我们上面的分析，对于 $V_0 = \text{span}(\mathbf{q}_1, \dots, \mathbf{q}_k)$ ，我们应当是能够证出来 $\max_{\mathbf{x} \in V_0 \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) = \lambda_k$ 的。下面我们就来证明这个事实。

取 $\mathbf{x} = \mathbf{q}_1\alpha_1 + \dots + \mathbf{q}_k\alpha_k$ 。不妨设 $\|\mathbf{x}\|_2 = 1$ 。根据 Rayleigh 商的定义我们有

$$\begin{aligned}\mathbf{x}^* \mathbf{A} \mathbf{x} &= \boldsymbol{\alpha}^* \mathbf{Q}^* \mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}^* \mathbf{Q} \boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}^* \boldsymbol{\Lambda} \boldsymbol{\alpha} \\ &= \sum_{i=1}^k \lambda_i |\alpha_i|^2 \\ &\leq \lambda_k \sum_{i=1}^k |\alpha_i|^2 \\ &= \lambda_k.\end{aligned}$$

上式等号成立当且仅当 $\boldsymbol{\alpha} = \mathbf{e}_k$ 。因此 $\max_{\mathbf{x} \in V_0 \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) = \lambda_k$ 。因此

$$\min_{\dim V=k} \max_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) \leq \max_{\mathbf{x} \in V_0 \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) = \lambda_k. \quad (7)$$

另一方面，设 $W = \text{span}(\mathbf{q}_k, \dots, \mathbf{q}_n)$ ，则 $\dim W = n - k + 1$ 。由维数公式，

$$\dim(V \cap W) = \dim V + \dim W - \dim(V + W) = n + 1 - \dim(V + W) \geq 1.$$

因此 $\exists \mathbf{x} \in V \cap W \setminus \{\mathbf{0}\}$ 。由 $\mathbf{x} \in W$ 得

$$\max_{\mathbf{x} \in V \cap W \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) \geq \lambda_k.$$

我们将上式放大到

$$\max_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) \geq \max_{\mathbf{x} \in V \cap W \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) \geq \lambda_k.$$

因此,

$$\min_{\dim V=k} \max_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) \geq \lambda_k. \quad (8)$$

因此由 (7) 和 (8) 有

$$\min_{\dim V=k} \max_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x}) = \lambda_k(\mathbf{A}). \quad (9)$$

对于第二个等式, 我们考虑将 $-\mathbf{A}$ 代入 (9) 中。由 $\lambda_k(\mathbf{A}) = -\lambda_{n-k+1}(-\mathbf{A})$ 知

$$\lambda_k(\mathbf{A}) = \max_{\dim V=n-k+1} \min_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{x})$$

□

这是一个非常深刻的结果。这个定理告诉我们：每个特征值都可以通过某种‘层层筛选’的优化过程来刻画。基于 Courant-Fischer 定理这个强大工具，我们将建立一整套研究特征值行为的理论：

1. Cauchy 交错定理；
2. 樊氏极小化迹原理；
3. Weyl 不等式；
4. Hoffman–Wielandt 不等式。

Cauchy 交错定理 我们刚刚通过 Courant–Fischer 定理获得了 Hermite 矩阵特征值的变分刻画。现在我们来考虑一个问题：当我们从一个大的 Hermite 矩阵中“截取”一个主子矩阵时，它们的特征值之间会有怎样的关系？这个问题在实际问题中无处不在：数值计算中我们需要用子矩阵近似大矩阵的特征值；数据科学中数据降维后方差（特征值）如何变化……

定理 3.5 (Cauchy 交错定理). $\mathbf{A}^* = \mathbf{A} \in \mathbb{C}^{n \times n}$, $\mathbf{B} = \mathbf{A}(1:n-1, 1:n-1)$ 是 \mathbf{A} 的第一个 $n-1$ 阶主子式。我们有

$$\lambda_1(\mathbf{A}) \leq \lambda_1(\mathbf{B}) \leq \lambda_2(\mathbf{A}) \leq \cdots \leq \lambda_{n-1}(\mathbf{A}) \leq \lambda_{n-1}(\mathbf{B}) \leq \lambda_n(\mathbf{A}).$$

证明. 我们只需证 $\lambda_k(\mathbf{A}) \leq \lambda_k(\mathbf{B}) \leq \lambda_{k+1}(\mathbf{A})$, $\forall k = 1, \dots, n-1$.

由 Courant–Fischer min-max 定理

$$\begin{aligned}
\lambda_k(\mathbf{B}) &= \min_{\dim V=k} \max_{\mathbf{x} \in V \setminus \{\mathbf{0}\}, \|\mathbf{x}\|_2=1} \mathbf{x}^* \mathbf{B} \mathbf{x} \\
&= \min_{\dim V=k} \max_{\mathbf{x} \in V \setminus \{\mathbf{0}\}, \|\mathbf{x}\|_2=1} \begin{bmatrix} \mathbf{x} \\ 0 \end{bmatrix}^* \mathbf{A} \begin{bmatrix} \mathbf{x} \\ 0 \end{bmatrix} \\
&= \min_{\dim \tilde{V}=k} \max_{\mathbf{y} \in \tilde{V} \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{y}), \quad \tilde{V} := \left\{ \begin{bmatrix} \mathbf{x} \\ 0 \end{bmatrix} : \mathbf{x} \in V \right\} \\
&\geq \min_{\dim V=k} \max_{\mathbf{y} \in V \setminus \{\mathbf{0}\}} R(\mathbf{A}, \mathbf{y}) \\
&= \lambda_k(\mathbf{A}).
\end{aligned}$$

第二个不等式可以通过将 \mathbf{A}, \mathbf{B} 替换为 $-\mathbf{A}, -\mathbf{B}$ 代入第一个不等式得到。读者可自行验证。 \square

逸事一则：Cauchy 虽然发现了这个定理，但他本人没有成功把这个定理证出来。

Cauchy 交错定理有两个推广形式：

推论 3.6. 若 \mathbf{C} 是 \mathbf{A} 的 $n-k$ 阶主子阵，则

$$\lambda_i(\mathbf{A}) \leq \lambda_i(\mathbf{C}) \leq \lambda_{i+k}(\mathbf{A}), \quad 1 \leq i \leq n-k.$$

证明. 对 \mathbf{C} 的阶数进行向下归纳。我们已知当 \mathbf{C} 为 \mathbf{A} 的 $n-1$ 阶主子阵时 \mathbf{C} 的特征值满足题中不等式。现假设对 $m = n-1, \dots, n-k+1$ 阶主子阵 \mathbf{C} 都满足该不等式。现在我们来证 \mathbf{C} 是 $n-k$ 阶主子阵的情况。

我们一定可以找到一个 $n-k+1$ 阶主子阵 \mathbf{B} ，使得 \mathbf{C} 是 \mathbf{B} 的主子阵。由 Cauchy 交错定理，

$$\lambda_i(\mathbf{B}) \leq \lambda_i(\mathbf{C}) \leq \lambda_{i+1}(\mathbf{B}).$$

由归纳假设我们有

$$\begin{cases} \lambda_i(\mathbf{B}) \geq \lambda_i(\mathbf{A}) \\ \lambda_{i+1}(\mathbf{B}) \leq \lambda_{i+k}(\mathbf{A}) \end{cases}$$

于是

$$\lambda_i(\mathbf{A}) \leq \lambda_i(\mathbf{C}) \leq \lambda_{i+k}(\mathbf{A}).$$

\square

推论 3.7 (Poincaré 隔离定理). 若正交归一阵 \mathbf{X} 满足 $\mathbf{X}^* \mathbf{X} = \mathbf{I}_{n-m}$ ，那么 $\lambda_k(\mathbf{A}) \leq \lambda_k(\mathbf{X}^* \mathbf{A} \mathbf{X}) \leq \lambda_{k+m}(\mathbf{A})$.

证明. 将 \mathbf{X} 扩充为全空间的一组基: $\mathbf{Q} = [\mathbf{X}, \mathbf{X}_\perp]$, 则有 $\mathbf{Q}^* \mathbf{Q} = \mathbf{I}_n$.

由于相似变换不改变特征值, 我们考虑对 \mathbf{A} 作酉相似变换:

$$\mathbf{Q}^* \mathbf{A} \mathbf{Q} = \begin{bmatrix} \mathbf{X}^* \mathbf{A} \mathbf{X} & \mathbf{X}^* \mathbf{A} \mathbf{X}_\perp \\ \mathbf{X}_\perp^* \mathbf{A} \mathbf{X} & \mathbf{X}_\perp^* \mathbf{A} \mathbf{X}_\perp \end{bmatrix}$$

注意到 $\mathbf{X}^* \mathbf{A} \mathbf{X}$ 是 $\mathbf{Q}^* \mathbf{A} \mathbf{Q}$ 的 $n - m$ 阶主子阵, 于是由上一个推论我们有

$$\lambda_k(\mathbf{A}) \leq \lambda_k(\mathbf{X}^* \mathbf{A} \mathbf{X}) \leq \lambda_{k+m}(\mathbf{A}).$$

□

这个结论也有一个推论。

推论 3.8 (樊氏极小化迹原理). $\min_{\mathbf{X}^* \mathbf{X} = \mathbf{I}_k} \text{tr}(\mathbf{X}^* \mathbf{A} \mathbf{X}) = \lambda_1(\mathbf{A}) + \cdots + \lambda_k(\mathbf{A})$.

证明.

$$\text{tr}(\mathbf{X}^* \mathbf{A} \mathbf{X}) = \sum_{i=1}^k \mathbf{x}_i^* \mathbf{A} \mathbf{x}_i = \sum_{i=1}^k R(\mathbf{A}, \mathbf{x}_i).$$

由推论 3.7 知 $\lambda_i(\mathbf{X}^* \mathbf{A} \mathbf{X}) \geq \lambda_i(\mathbf{A})$, 即得待证结论。

□

Weyl 不等式

慕雙雄攜手，破宇稱守恆，啟我後學二三輩。

繼外爾規範，始強力物理，叱咤科壇六十年。

—— 丘成桐

在进行扰动分析时, 我们经常给初始数据一个向后误差去估计向前误差。向后误差一般都是加性扰动 $\mathbf{A} \rightarrow \mathbf{A} + \mathbf{E}$, 其中 \mathbf{E} 是一个小扰动矩阵。现在, 我们就来对特征值问题进行扰动分析。这也就是要估计

$$|\lambda_k(\mathbf{A} + \mathbf{E}) - \lambda_k(\mathbf{A})|.$$

Weyl 不等式很好地回答了这个问题。

定理 3.9 (Weyl 不等式).

$$\lambda_{i-j+1}(\mathbf{A}) + \lambda_j(\mathbf{B}) \leq \lambda_i(\mathbf{A} + \mathbf{B}) \leq \lambda_{i+j}(\mathbf{A}) + \lambda_{n-j}(\mathbf{B}), \quad j = 0, 1, \dots, n - i$$

第一个等号成立当且仅当

$$\begin{cases} \mathbf{Ax} = x\lambda_{i-j+1}(\mathbf{A}) \\ \mathbf{Bx} = x\lambda_j(\mathbf{B}) \\ (\mathbf{A} + \mathbf{B})\mathbf{x} = x\lambda_i(\mathbf{A} + \mathbf{B}) \end{cases}$$

第二个不等号成立当且仅当

$$\begin{cases} \mathbf{Ax} = x\lambda_{i+j}(\mathbf{A}) \\ \mathbf{Bx} = x\lambda_{n-j}(\mathbf{B}) \\ (\mathbf{A} + \mathbf{B})\mathbf{x} = x\lambda_i(\mathbf{A} + \mathbf{B}) \end{cases}$$

证明. 假设 $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, $\mathbf{z}_1, \dots, \mathbf{z}_n$ 分别是 $\mathbf{A}, \mathbf{B}, \mathbf{A} + \mathbf{B}$ 的特征向量, 分别对应于 $\lambda_1(\cdot), \dots, \lambda_n(\cdot)$.

对给定的 $i \in \{1, \dots, n\}$ 以及任意的 $j \in \{0, \dots, n-i\}$, 我们定义 $S_1 = \{\mathbf{x}_1, \dots, \mathbf{x}_{i+j}\}$, $S_2 = \{\mathbf{y}_1, \dots, \mathbf{y}_{n-j}\}$, $S_3 = \{\mathbf{z}_i, \dots, \mathbf{z}_n\}$, 则有

$$\dim S_1 + \dim S_2 + \dim S_3 = 2n + 1.$$

因此存在 $\mathbf{x} \in S_1 \cap S_2 \cap S_3$, 不妨设 $\|\mathbf{x}\|_2 = 1$.

由 $\mathbf{x} \in S_3$ 有

$$\lambda_i(\mathbf{A} + \mathbf{B}) \leq \mathbf{x}^*(\mathbf{A} + \mathbf{B})\mathbf{x} = \mathbf{x}^*\mathbf{Ax} + \mathbf{x}^*\mathbf{Bx}.$$

由 $\mathbf{x} \in S_1$ 知 $\mathbf{x}^*\mathbf{Ax} \leq \lambda_{i+j}(\mathbf{A})$; 由 $\mathbf{x} \in S_2$ 知 $\mathbf{x}^*\mathbf{Bx} \leq \lambda_{n-j}(\mathbf{B})$. 因此

$$\lambda_i(\mathbf{A} + \mathbf{B}) \leq \lambda_{i+j}(\mathbf{A}) + \lambda_{n-j}(\mathbf{B}).$$

等号成立当且仅当

$$\begin{cases} \mathbf{Ax} = x\lambda_{i+j}(\mathbf{A}) \\ \mathbf{Bx} = x\lambda_{n-j}(\mathbf{B}) \\ (\mathbf{A} + \mathbf{B})\mathbf{x} = x\lambda_i(\mathbf{A} + \mathbf{B}) \end{cases}$$

对于另一侧的不等式——想必你也能猜到——我们只需对 $-\mathbf{A}, -\mathbf{B}$ 应用上面的不等式即可得到定理中的左侧不等式了。□

这个不等式相当抽象。为了获得一个更直观的理解, 我们可以考虑它的一些重要推论。取 $j = 1$, 我们得到

$$\lambda_i(\mathbf{A}) + \lambda_1(\mathbf{B}) \leq \lambda_i(\mathbf{A} + \mathbf{B}) \leq \lambda_i(\mathbf{A}) + \lambda_n(\mathbf{B}).$$

这个不等式告诉我们, 矩阵 $\mathbf{A} + \mathbf{B}$ 的第 i 个特征值, 被“夹在”了 \mathbf{A} 的第 i 个特征值加上 \mathbf{B} 的最小和最大特征值之间。我们可以把它写成

$$\lambda_1(\mathbf{B}) \leq \lambda_i(\mathbf{A} + \mathbf{B}) - \lambda_i(\mathbf{A}) \leq \lambda_n(\mathbf{B}).$$

这意味着，第 i 个特征值的变化量被扰动矩阵 \mathbf{B} 的谱范围（最小到最大特征值）所控制。

如果我们取绝对值，并利用矩阵的谱范数（2-范数） $\|\mathbf{B}\|_2 = \max(|\lambda_1(\mathbf{B})|, |\lambda_n(\mathbf{B})|)$ ，我们可以得到一个更简洁和著名的结果：

$$|\lambda_i(\mathbf{A} + \mathbf{B}) - \lambda_i(\mathbf{A})| \leq \|\mathbf{B}\|_2.$$

这个结果非常强大。它说明，第 i 个特征值的变化幅度，绝对不会超过扰动矩阵 \mathbf{B} 的谱范数。因此，只要扰动 \mathbf{B} 很小（即其范数很小），特征值的变化就一定很小。这在数值计算和物理系统中至关重要，因为它保证了当我们的模型或数据有微小误差时，其内在的“频率”或“能量”（特征值）是稳定的。

对于原始的那个最精确的不等式，我还没找到比较好的理解方法；如果读者有更好的想法欢迎来联系我。不过，我想原定理最核心的思想就是交空间会受到两边的共同约束——因此定理的关键就是构造出 S_3 ，使得它与 S_1 和 S_2 交集非空。简单来说，Weyl 不等式通过巧妙地利用子空间的维数和交集，精确地告诉我们：两个矩阵相加后，新矩阵的特征值（能量）是如何由原始两个矩阵的特征值（能量）叠加并“交错”在一起的。我想它应该不仅仅是一个简单的上界或下界，而是揭示了特征谱之间更深刻的结构关系——但我没想明白。

Hoffman–Wielandt 不等式 Weyl 不等式告诉我们施加扰动后某个特征值的扰动情况——这是局部的视角。但有的时候我们还需要一些全局的视角——

定理 3.10 (Hoffman–Wielandt 不等式).

$$\sum_{i=1}^n |\lambda_k(\mathbf{A} + \Delta\mathbf{A}) - \lambda_k(\mathbf{A})|^2 \leq \|\Delta\mathbf{A}\|_F^2.$$

这个不等式有一个优美的几何解释——至少比 Weyl 不等式优美得多——如果我们把特征值看作 \mathbb{R}^n 中的两个点 $(\lambda_1(\mathbf{A}), \dots, \lambda_n(\mathbf{A}))$ 及 $(\lambda_1(\mathbf{A} + \Delta\mathbf{A}), \dots, \lambda_n(\mathbf{A} + \delta\mathbf{A}))$ ，那么这两个点的距离不超过矩阵之间的 Frobenius 距离。换句话说，特征值作为一个整体的移动幅度是受限的。

证明. 考虑 $\mathbf{A}(t) = \mathbf{A} + t\Delta\mathbf{A}$.

我们将不加证明地使用一个引理：对于一个解析形式的 Hermite 矩阵 $\mathbf{A}(t) = (\mathbf{A}(t))^*$ ，它存在解析形式的谱分解，即存在 $\mathbf{Q}(t), \mathbf{\Lambda}(t)$ 使得

$$\begin{cases} \mathbf{A}(t) = \mathbf{Q}(t)\mathbf{\Lambda}(t)(\mathbf{Q}(t))^* \\ (\mathbf{Q}(t))^*\mathbf{Q}(t) = \mathbf{I}_n \end{cases}$$

其中 $\mathbf{A}(t), \mathbf{Q}(t), \mathbf{\Lambda}(t)$ 的元素都是在 $t \in (a, b)$ 上的解析函数，且对于给定的 t ， $\mathbf{A}(t), \mathbf{Q}(t), \mathbf{\Lambda}(t)$ 分别是 Hermite 阵、酉矩阵、对角阵。

对 $\mathbf{Q}(t)^* \mathbf{Q}(t) = \mathbf{I}_n$ 两边求导, 得

$$\mathbf{Q}'(t)^* \mathbf{Q}(t) + \mathbf{Q}(t)^* \mathbf{Q}'(t) = \mathbf{0}.$$

这说明 $\mathbf{K}(t) = \mathbf{Q}(t)^* \mathbf{Q}'(t)$ 是反 Hermite 矩阵, 其对角元是纯虚数, 我们不妨记其对角部分 $\text{diag}(\mathbf{K}(t)) = \mathbf{iD}(t)$, 其中 $\mathbf{D}(t)$ 为实对角阵。

对 $\mathbf{\Lambda}(t) = \mathbf{Q}(t)^* \mathbf{A}(t) \mathbf{Q}(t)$ 求导得

$$\begin{aligned} \mathbf{\Lambda}'(t) &= \mathbf{Q}'(t)^* \mathbf{A}(t) \mathbf{Q}(t) + \mathbf{Q}(t)^* \mathbf{A}'(t) \mathbf{Q}(t) + \mathbf{Q}(t)^* \mathbf{A}(t) \mathbf{Q}'(t) \\ &= \mathbf{K}(t) \mathbf{\Lambda}(t) + \mathbf{Q}(t)^* \Delta \mathbf{A} \mathbf{Q}(t) + \mathbf{\Lambda}(t) \mathbf{K}(t)^*. \end{aligned}$$

由于 $\mathbf{\Lambda}(t)$ 是对角阵, 故 $\mathbf{\Lambda}'(t)$ 的非对角部分一定为零。于是

$$\begin{aligned} \mathbf{\Lambda}'(t) &= \text{diag}(\mathbf{\Lambda}'(t)) \\ &= \text{diag}(\mathbf{K}(t) \mathbf{\Lambda}(t) + \mathbf{Q}(t)^* \Delta \mathbf{A} \mathbf{Q}(t) + \mathbf{\Lambda}(t) \mathbf{K}(t)^*) \\ &= \text{diag}(\mathbf{K}(t)) \mathbf{\Lambda}(t) + \text{diag}(\mathbf{Q}(t)^* \Delta \mathbf{A} \mathbf{Q}(t)) + \mathbf{\Lambda}(t) \text{diag}(\mathbf{K}(t)^*) \\ &= \mathbf{iD}(t) \mathbf{\Lambda}(t) + \text{diag}(\mathbf{Q}(t)^* \Delta \mathbf{A} \mathbf{Q}(t)) - \mathbf{\Lambda}(t) \mathbf{iD}(t) \\ &= \text{diag}(\mathbf{Q}(t)^* \Delta \mathbf{A} \mathbf{Q}(t)). \end{aligned}$$

于是

$$\begin{aligned} \sum_{i=1}^n |\lambda_i(\mathbf{A} + \Delta \mathbf{A}) - \lambda_i(\mathbf{A})|^2 &= \|\Delta \mathbf{\Lambda}\|_{\text{F}}^2 \\ &= \|\mathbf{\Lambda}(1) - \mathbf{\Lambda}(0)\|_{\text{F}}^2 \\ &= \left\| \int_0^1 \mathbf{\Lambda}'(t) \, dt \right\|_{\text{F}}^2 \\ &\leq \left(\int_0^1 \|\mathbf{\Lambda}'(t)\|_{\text{F}} \, dt \right)^2 \\ &= \left(\int_0^1 \|\text{diag}(\mathbf{Q}(t)^* \Delta \mathbf{A} \mathbf{Q}(t))\|_{\text{F}} \, dt \right)^2 \\ &\leq \left(\int_0^1 \|\mathbf{Q}(t)^* \Delta \mathbf{A} \mathbf{Q}(t)\|_{\text{F}} \, dt \right)^2 \\ &= \left(\int_0^1 \|\Delta \mathbf{A}\|_{\text{F}} \, dt \right)^2 \\ &= \|\Delta \mathbf{A}\|_{\text{F}}^2. \end{aligned}$$

□