MATH10003.01 高等线性代数 Fall 2025

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：$z=\rho \mathrm{e}^{\mathrm{i}\theta}=\rho(\cos\theta+\mathrm{i}\sin\theta)$，如果写成矩阵形式就是

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}=\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix}:=\boldsymbol{QH}.$$

其中，$\boldsymbol{Q}$ 是一个正交矩阵，$\boldsymbol{H}$ 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 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$，是否也存在类似的“极分解”？我们能否把一个矩阵 $\boldsymbol{A}$ 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 $\boldsymbol{A}\in\mathbb{C}^{m\times n}$，则存在列正交归一的矩阵 $\boldsymbol{Q}\in\mathbb{C}^{m\times n}$ 和 Hermite 半正定矩阵 $\boldsymbol{H}\in\mathbb{C}^{n\times n}$，使得 $\boldsymbol{A}=\boldsymbol{QH}$，并且当 $\boldsymbol{A}$ 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 $\boldsymbol{A}$ 的奇异值分解 $\boldsymbol{A}=\boldsymbol{U\Sigma V}^\star$，现在让我们来做一些代数变换：

$$\boldsymbol{A}=\boldsymbol{U\Sigma V}^\star=(\boldsymbol{UV}^\star)(\boldsymbol{V\Sigma V}^\star)$$

于是我们定义 $\boldsymbol{Q}=\boldsymbol{UV}^\star$，$\boldsymbol{H}=\boldsymbol{V\Sigma V}^\star$，则有

$$\boldsymbol{A}=\boldsymbol{QH}.$$

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

类似地，我们还可以写出另一种形式：

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

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 $\boldsymbol{A}$ 都可以看成一个旋转（或反射） $\boldsymbol{Q}$ 与一个沿特征方向伸缩的 $\boldsymbol{H}$ 的复合，复合的先后顺序并不重要。SVD 需要两个不同的旋转 $\boldsymbol{U}$ 和 $\boldsymbol{V}^\star$，因此极分解更加精简，但代价是伸缩部分 $\boldsymbol{H}$ 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 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，它有谱分解 $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^\star$. 如果我们记 $\boldsymbol{\Lambda} = \operatorname{diag}(\boldsymbol{\Lambda}_+, \boldsymbol{\Lambda}_-)$，并相应地对 $\boldsymbol{Q}$ 作分块 $\boldsymbol{Q} = [\boldsymbol{Q}_+, \boldsymbol{Q}_-]$，我们有

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

我们把 $\tilde{\boldsymbol{Q}}$ 展开，有

$$\tilde{\boldsymbol{Q}} = \boldsymbol{Q}_+\boldsymbol{Q}_+^\star - \boldsymbol{Q}_-\boldsymbol{Q}_-^\star \tag{1}$$

这说明 $\tilde{\boldsymbol{Q}}$ 可以拆分成两个正交投影的差。由于 $\boldsymbol{Q}_+$ 和 $\boldsymbol{Q}_-$ 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

$$\boldsymbol{I} = \boldsymbol{Q}_+\boldsymbol{Q}_+^\star + \boldsymbol{Q}_-\boldsymbol{Q}_-^\star \tag{2}$$

将 (1) 和 (2) 联立，有

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

这样，我们就利用极分解得到了分离 $\boldsymbol{A}$ 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，$\tilde{\boldsymbol{Q}}$ 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 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$，即我们有 $\boldsymbol{A}^\star\boldsymbol{A} \approx \boldsymbol{I}$. 现在我们想要找一个最接近 $\boldsymbol{A}$ 的正交归一阵 $\boldsymbol{Q}_1$ 去逼近 $\boldsymbol{A}$，即我们想要求解下面的优化问题：

$$\min_{\boldsymbol{Q}_1^\star\boldsymbol{Q}_1=\boldsymbol{I}} \|\boldsymbol{A}-\boldsymbol{Q}_1\|_{\mathrm{F}}^2.$$

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

$$\min_{\boldsymbol{Q}_1^\star\boldsymbol{Q}_1=\boldsymbol{I}} \|\boldsymbol{A}-\boldsymbol{Q}_1\|_{\mathrm{F}}^2 = \min_{\boldsymbol{Q}_1^\star\boldsymbol{Q}_1=\boldsymbol{I}} \|\boldsymbol{U}(\boldsymbol{\Sigma}-\boldsymbol{Q}_2)\boldsymbol{V}^\star\|_{\mathrm{F}}^2 = \min_{\boldsymbol{Q}_2^\star\boldsymbol{Q}_2=\boldsymbol{I}} \|\boldsymbol{\Sigma}-\boldsymbol{Q}_2\|_{\mathrm{F}}^2.$$

根据 Frobenius 范数的定义 $\|\boldsymbol{A}\|_{\mathrm{F}}^2=\mathrm{tr}(\boldsymbol{A}^\star\boldsymbol{A})$，我们有

$$\min_{\boldsymbol{Q}_2^\star\boldsymbol{Q}_2=\boldsymbol{I}} \|\boldsymbol{\Sigma}-\boldsymbol{Q}_2\|_{\mathrm{F}}^2 = \min_{\boldsymbol{Q}_2^\star\boldsymbol{Q}_2=\boldsymbol{I}} \mathrm{tr}((\boldsymbol{\Sigma}-\boldsymbol{Q}_2)^\star(\boldsymbol{\Sigma}-\boldsymbol{Q}_2)).$$

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

$$\begin{aligned}
\max_{\boldsymbol{Q}_2^\star\boldsymbol{Q}_2=\boldsymbol{I}} \mathrm{tr}(\boldsymbol{\Sigma}^\star\boldsymbol{Q}_2+\boldsymbol{Q}_2^\star\boldsymbol{\Sigma}) &= \max_{\boldsymbol{Q}_2^\star\boldsymbol{Q}_2=\boldsymbol{I}} \sum_{i=1}^n (\sigma_i q_{i,i}+\sigma_i\overline{q_{i,i}}) \\
&= \max_{\boldsymbol{Q}_2^\star\boldsymbol{Q}_2=\boldsymbol{I}} \sum_{i=1}^n 2\sigma_i \mathrm{Re}(q_{i,i}) \\
&\le \sum_{i=1}^n 2\sigma_i|q_{i,i}| \\
&\le \sum_{i=1}^n 2\sigma_i \\
&= 2\sum_{i=1}^n \sigma_i.
\end{aligned}$$

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

$$\max_{\boldsymbol{Q}_2^\star\boldsymbol{Q}_2=\boldsymbol{I}} \mathrm{tr}(\boldsymbol{\Sigma}^\star\boldsymbol{Q}_2+\boldsymbol{Q}_2^\star\boldsymbol{\Sigma}) \ge 2\sum_{i=1}^n 2\sigma_i.$$

因此

$$\max_{\boldsymbol{Q}_2^\star\boldsymbol{Q}_2=\boldsymbol{I}} \mathrm{tr}(\boldsymbol{\Sigma}^\star\boldsymbol{Q}_2+\boldsymbol{Q}_2^\star\boldsymbol{\Sigma}) = 2\sum_{i=1}^n 2\sigma_i.$$

当且仅当 $\boldsymbol{Q}_2=\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}$ 时等号成立。此时 $\boldsymbol{Q}_1=\boldsymbol{U}\boldsymbol{Q}_2\boldsymbol{V}^\star$，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 $\boldsymbol{A}$，存在唯一的逆矩阵 $\boldsymbol{A}^{-1}$ 满足：

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

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 $\boldsymbol{A} : \mathbb{C}^n \to \mathbb{C}^n$，

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

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

- 像空间 $\mathrm{Range}(\boldsymbol{A}) \subset \mathbb{C}^m$；
- 核空间 $\mathrm{Ker}(\boldsymbol{A}) \subset \mathbb{C}^n$.

我们希望找到一个广义逆 $\boldsymbol{A}^\dagger$，使得它在 $\mathrm{Range}(\boldsymbol{A})$ 上的行为类似于逆，而在 $\mathrm{Ker}(\boldsymbol{A})$ 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 $\boldsymbol{A}$ 的奇异值分解

$$\boldsymbol{A} = \boldsymbol{U}\begin{bmatrix}\boldsymbol{\Sigma}_\star & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\boldsymbol{V}^\star$$

那么我们定义其广义逆为

$$\boldsymbol{A}^\dagger = \boldsymbol{V}\begin{bmatrix}\boldsymbol{\Sigma}_\star^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\boldsymbol{U}^\star.$$

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

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

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

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

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

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

$$\boldsymbol{X} = \boldsymbol{V}\begin{bmatrix}\boldsymbol{\Sigma}_\star^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\boldsymbol{U}^\star$$

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 $\boldsymbol{Y}$ 满足上面的四个方程组。下面我们说明 $\boldsymbol{X} = \boldsymbol{Y}$.

由 (4) 我们有

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

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

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

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

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

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

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

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

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

另一方面，我们从 $\boldsymbol{Y}$ 出发，也能得到类似的结论：

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

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

2. 方程 (4) 说明了 $\boldsymbol{A}^\dagger\boldsymbol{A}$ 在 Range($\boldsymbol{A}^\dagger$) 上是恒等算子——这是 $\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{I}$ 的推广。同时，这个方程还说明了 $\boldsymbol{A}^\dagger\boldsymbol{A}$ 是幂等算子。

3. 方程 (5) 说明了 $\boldsymbol{A}\boldsymbol{A}^\dagger$ 是自伴算子。结合第一条，易见 $\boldsymbol{A}\boldsymbol{A}^\dagger$ 是一个投影算子。

4. 方程 (6) 说明了 $\boldsymbol{A}^\dagger\boldsymbol{A}$ 是自伴算子。结合第二条，易见 $\boldsymbol{A}^\dagger\boldsymbol{A}$ 是一个投影算子。

既然 $\boldsymbol{A}\boldsymbol{A}^\dagger$ 和 $\boldsymbol{A}^\dagger\boldsymbol{A}$ 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 $\boldsymbol{A}\boldsymbol{A}^\dagger$ 在 Range($\boldsymbol{A}$) 上是恒等算子，显然它是到 Range($\boldsymbol{A}$) 上的投影算子——不然也不会把 $\boldsymbol{A}$ 原封不动地映到 $\boldsymbol{A}$. 同理，$\boldsymbol{A}^\dagger\boldsymbol{A}$ 是到 Range($\boldsymbol{A}^\dagger$) 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 $\boldsymbol{A}^\dagger$ 的显式解代入，有

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

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

在上节课我们说过，$\boldsymbol{U}_1\boldsymbol{U}_1^\star$ 是到 Range($\boldsymbol{A}$) 的投影算子，因此 $\boldsymbol{A}\boldsymbol{A}^\dagger$ 是到 Range($\boldsymbol{A}$) 的投影算子；$\boldsymbol{V}_1\boldsymbol{V}_1^\star$ 是到 Range($\boldsymbol{A}^\star$) 的投影算子，因此 $\boldsymbol{A}^\dagger\boldsymbol{A}$ 是到 Range($\boldsymbol{A}^\star$) 的投影算子。因此，Range($\boldsymbol{A}^\star$) = Range($\boldsymbol{A}^\dagger$)，$\boldsymbol{A}^\star$ 与 $\boldsymbol{A}^\dagger$ 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 Ker($\boldsymbol{A}$) 的投影算子可以表示为 $\boldsymbol{V}_2\boldsymbol{V}_2^\star=\boldsymbol{I}-\boldsymbol{A}^\dagger\boldsymbol{A}$，到 Ker($\boldsymbol{A}^\star$) 的投影算子可以表示为 $\boldsymbol{U}_2\boldsymbol{U}_2^\star=\boldsymbol{I}-\boldsymbol{A}\boldsymbol{A}^\dagger$. 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
Range($\boldsymbol{A}$)	$\boldsymbol{U}_1\boldsymbol{U}_1^\star=\boldsymbol{A}\boldsymbol{A}^\dagger$	Range($\boldsymbol{U}_1$) = Range($\boldsymbol{A}\boldsymbol{A}^\dagger$)
Range($\boldsymbol{A}^\star$)	$\boldsymbol{V}_1\boldsymbol{V}_1^\star=\boldsymbol{A}^\dagger\boldsymbol{A}$	Range($\boldsymbol{V}_1$) = Range($\boldsymbol{A}^\dagger\boldsymbol{A}$)
Ker($\boldsymbol{A}$)	$\boldsymbol{V}_2\boldsymbol{V}_2^\star=\boldsymbol{I}-\boldsymbol{A}^\dagger\boldsymbol{A}$	Range($\boldsymbol{V}_2$) = Range($\boldsymbol{I}-\boldsymbol{A}^\dagger\boldsymbol{A}$)
Ker($\boldsymbol{A}^\star$)	$\boldsymbol{U}_2\boldsymbol{U}_2^\star=\boldsymbol{I}-\boldsymbol{A}\boldsymbol{A}^\dagger$	Range($\boldsymbol{U}_2$) = Range($\boldsymbol{I}-\boldsymbol{A}\boldsymbol{A}^\dagger$)

现在，我们换一个角度来解释为什么 Range($\boldsymbol{A}^\dagger$) = Range($\boldsymbol{A}^\star$). 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 $\boldsymbol{A}^\dagger$ 和 $\boldsymbol{A}^\star$，有

$$\boldsymbol{A}^\star=\sum_{i=1}^{r}\sigma_i\boldsymbol{v}_i\boldsymbol{u}_i^\star$$

$$\boldsymbol{A}^\dagger=\sum_{i=1}^{r}\sigma_i^{-1}\boldsymbol{v}_i\boldsymbol{u}_i^\star$$

因此这两者张成了相同的空间。既然如此，我们能不能用 $\boldsymbol{A}$ 和 $\boldsymbol{A}^\star$ 来表示 $\boldsymbol{A}^\dagger$ 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 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ 和 $\boldsymbol{b} \in \mathbb{C}^n$，求解 $\arg\min \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$. 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 $\boldsymbol{A}$ 的奇异值分解 $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\star$，我们记 $\boldsymbol{y} = \boldsymbol{V}^\star\boldsymbol{x}$，于是我们只要求解下面这个优化问题：

$$\min_{\boldsymbol{y}} \|\boldsymbol{\Sigma}\boldsymbol{y} - \boldsymbol{U}^\star\boldsymbol{b}\|_2.$$

如果记 $\boldsymbol{U}^\star\boldsymbol{b} = [\boldsymbol{c}_1^{\mathrm{T}}, \boldsymbol{c}_2^{\mathrm{T}}]^{\mathrm{T}}$，我们有

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

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

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

所以

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

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

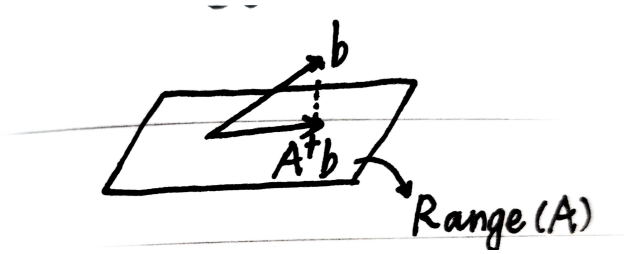

图 1: 最小二乘问题

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 $\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_{LS}$ 垂直于 Range($\boldsymbol{A}$)，这也就是

$$\boldsymbol{A}^\star(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{0}.$$

因此，

$$\boldsymbol{A}^\star\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^\star\boldsymbol{b}.$$

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

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

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

$$\boldsymbol{A}^\dagger = (\boldsymbol{A}^\star \boldsymbol{A})^{-1} \boldsymbol{A}^\star.$$

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

$$\boldsymbol{A}^\dagger = \boldsymbol{A}^\star (\boldsymbol{A}\boldsymbol{A}^\star)^{-1}.$$

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. $(\boldsymbol{A}^\star)^\dagger = (\boldsymbol{A}^\dagger)^\star$.

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

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

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

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

□

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

证明. 显然。 □

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

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

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

证明. 显然。 □

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

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

这个引理是非常显然的。由于 $\boldsymbol{A}^\dagger \boldsymbol{A} = \boldsymbol{V}_1 \boldsymbol{V}_1^\star$，且 $\boldsymbol{A}$ 列满秩，因此 $\boldsymbol{V}_1$ 就是 $\boldsymbol{V}$，因此 $\boldsymbol{A}^\dagger \boldsymbol{A} = \boldsymbol{I}_r$. $\boldsymbol{B}$ 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 $(\boldsymbol{AB})(\boldsymbol{B}^\dagger\boldsymbol{A}^\dagger)(\boldsymbol{AB}) = \boldsymbol{AB}$，由引理结论显然成立。

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

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

方程 (6)：我们需要验证 $((\boldsymbol{B}^\dagger\boldsymbol{A}^\dagger)(\boldsymbol{AB}))^\star = (\boldsymbol{B}^\dagger\boldsymbol{A}^\dagger)(\boldsymbol{AB})$. 由引理，我们只需证 $(\boldsymbol{B}^\dagger\boldsymbol{B})^\star = \boldsymbol{B}^\dagger\boldsymbol{B}$，是以结论显然。 □

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

证明. 若 $\boldsymbol{A}$ 列满秩或行满秩，$\boldsymbol{A}^\dagger$ 可由 $\boldsymbol{A}$ 和 $\boldsymbol{A}^\star$ 表示，显然为有理矩阵。

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

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

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

$$\boldsymbol{AB} = (\boldsymbol{AU})\boldsymbol{\Sigma V}^\star.$$

于是

$$(\boldsymbol{AB})^\dagger = \boldsymbol{V\Sigma}(\boldsymbol{AU})^\star = (\boldsymbol{V\Sigma U}^\star)\boldsymbol{A}^\star = \boldsymbol{B}^\dagger\boldsymbol{A}^\dagger$$

$\boldsymbol{B}$ 为酉矩阵时同理。 □

请注意，一般而言，$(\boldsymbol{AB})^\dagger \neq \boldsymbol{B}^\dagger\boldsymbol{A}^\dagger$，$(\boldsymbol{A}^k)^\dagger \neq (\boldsymbol{A}^\dagger)^k$.

命题 2.9. $(\boldsymbol{A}\otimes\boldsymbol{B})^\dagger = \boldsymbol{A}^\dagger \otimes \boldsymbol{B}^\dagger$.

证明. 我们逐一验证 Penrose 方程组。

方程 (3)：

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

方程 (4)：

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

方程 (5)：

$$((\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{A}^\dagger\otimes\boldsymbol{B}^\dagger))^\star = (\boldsymbol{AA}^\dagger\otimes\boldsymbol{BB}^\dagger)^\star = (\boldsymbol{AA}^\dagger)^\star\otimes(\boldsymbol{BB}^\dagger)^\star = \boldsymbol{AA}^\dagger\otimes\boldsymbol{BB}^\dagger.$$

方程 (6)：

$$((\boldsymbol{A}^\dagger \otimes \boldsymbol{B}^\dagger)(\boldsymbol{A} \otimes \boldsymbol{B}))^\star = (\boldsymbol{A}^\dagger \boldsymbol{A} \otimes \boldsymbol{B}^\dagger \boldsymbol{B})^\star = (\boldsymbol{A}^\dagger \boldsymbol{A})^\star \otimes (\boldsymbol{B}^\dagger \boldsymbol{B})^\star = \boldsymbol{A}^\dagger \boldsymbol{A} \otimes \boldsymbol{B}^\dagger \boldsymbol{B}.$$

□

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 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$，它的特征方程为

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}\lambda.$$

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 $\boldsymbol{x}^\star$，我们有

$$\lambda = \frac{\boldsymbol{x}^\star \boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^\star \boldsymbol{x}}.$$

这启发我们定义下面这个量：

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

$$R(\boldsymbol{A}, \boldsymbol{x}) = \frac{\boldsymbol{x}^\star \boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^\star \boldsymbol{x}}.$$

显然，若 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}\lambda$，我们有 $R(\boldsymbol{A}, \boldsymbol{x}) = \lambda$.

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 $\boldsymbol{x}$ 在变换 $\boldsymbol{A}$ 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)**.** $w(\boldsymbol{A}) := \{R(\boldsymbol{A}, \boldsymbol{x}) : \boldsymbol{x} \neq 0\}$ *称为 $\boldsymbol{A}$ 的数值域。*

下面我们所有的讨论均假设 $\boldsymbol{A}$ 是 Hermite 矩阵，并设特征值按递增顺序排列，即 $\lambda_1(\cdot) \leq \cdots \leq \lambda_n(\cdot)$，$\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ 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 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 的数值域 $w(\boldsymbol{A}) = [\lambda_1(\boldsymbol{A}), \lambda_n(\boldsymbol{A})] \subset \mathbb{R}$.*

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

$$R(\boldsymbol{Q}^\star\boldsymbol{A}\boldsymbol{Q}, \boldsymbol{x}) = R(\boldsymbol{A}, \boldsymbol{Q}\boldsymbol{x}).$$

因此 $w(\boldsymbol{A}) = w(\boldsymbol{Q}^\star\boldsymbol{A}\boldsymbol{Q}) = w(\boldsymbol{\Lambda})$.

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

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

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 $R(\boldsymbol{A}, \boldsymbol{x}) = [\lambda_1(\boldsymbol{A}), \lambda_n(\boldsymbol{A})]$. □

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 $R(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{x}) = \frac{\boldsymbol{x}^\star\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^\star\boldsymbol{B}\boldsymbol{x}}$，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 $[\lambda_1(\boldsymbol{A}), \lambda_n(A)]$.

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)**.**

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

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 $\lambda_n(\boldsymbol{A})$ 对应的特征子空间，那么 $\max_{\boldsymbol{x}\in V\setminus\{0\}} R(\boldsymbol{A},\boldsymbol{x}) = \lambda_{n-1}$；但倘若 V 包含 λ_n 对应的特征子空间，那么就有 $\max_{\boldsymbol{x}\in V\setminus\{0\}} R(\boldsymbol{A},\boldsymbol{x}) = \lambda_n$. 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 $\lambda_n(\boldsymbol{A}), \lambda_{n-1}(\boldsymbol{A})$ 对应的特征子空间，就有 $\min_{\{}\, \boldsymbol{x} \in V\setminus\{\boldsymbol{0}\}\} R(\boldsymbol{A},\boldsymbol{x}) = \lambda_{n-1}(\boldsymbol{A})$；若不包含任意一个特征空间，这个量就会比 $\lambda_{n-1}(\boldsymbol{A})$ 小。为了捕获 $\lambda_{n-1}(\boldsymbol{A})$，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 $V_0 = \operatorname{span}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k)$，我们应当是能够证出来 $\max_{\boldsymbol{x}\in V_0\setminus\{\boldsymbol{0}\}} R(\boldsymbol{A},\boldsymbol{x}) = \lambda_k$ 的。下面我们就来证明这个事实。

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

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

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

$$\min_{\dim V = k} \max_{\boldsymbol{x}\in V\setminus\{\boldsymbol{0}\}} R(\boldsymbol{A},\boldsymbol{x}) \le \max_{\boldsymbol{x}\in V_0\setminus\{\boldsymbol{0}\}} R(\boldsymbol{A},\boldsymbol{x}) = \lambda_k. \tag{7}$$

另一方面，设 $W = \operatorname{span}(\boldsymbol{q}_k, \ldots, q_n)$，则 $\dim W = n-k+1$. 由维数公式，

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

因此 $\exists \boldsymbol{x} \in V\cap W\setminus\{\boldsymbol{0}\}$. 由 $\boldsymbol{x}\in W$ 得

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

我们将上式放大到

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

因此，

$$\min_{\dim V = k} \max_{\boldsymbol{x}\in V\setminus\{\boldsymbol{0}\}} R(\boldsymbol{A},\boldsymbol{x}) \ge \lambda_k. \tag{8}$$

因此由 (7) 和 (8) 有

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

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

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

□

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)**.** $\boldsymbol{A}^\star=\boldsymbol{A}\in\mathbb{C}^{n\times n}$，$\boldsymbol{B}=\boldsymbol{A}(1:n-1,1:n-1)$ 是 $\boldsymbol{A}$ 的第一个 $n-1$ 阶主子式。我们有

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

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

由 Courant–Fischer min-max 定理

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

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

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 $\boldsymbol{C}$ 是 $\boldsymbol{A}$ 的 $n-k$ 阶主子阵，则

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

证明. 对 $\boldsymbol{C}$ 的阶数进行向下归纳。我们已知当 $\boldsymbol{C}$ 为 $\boldsymbol{A}$ 的 $n-1$ 阶主子阵时 $\boldsymbol{C}$ 的特征值满足题中不等式。现假设对 $m = n-1, \ldots, n-k+1$ 阶主子阵 $\boldsymbol{C}$ 都满足该不等式。现在我们来证 $\boldsymbol{C}$ 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 $\boldsymbol{B}$，使得 $\boldsymbol{C}$ 是 $\boldsymbol{B}$ 的主子阵。由 Cauchy 交错定理，

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

由归纳假设我们有

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

于是

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

□

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

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

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

$$\boldsymbol{Q}^\star\boldsymbol{A}\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{X}^\star\boldsymbol{A}\boldsymbol{X} & \boldsymbol{X}^\star\boldsymbol{A}\boldsymbol{X}_\perp \\ \boldsymbol{X}_\perp^\star\boldsymbol{A}\boldsymbol{X} & \boldsymbol{X}_\perp^\star\boldsymbol{A}\boldsymbol{X}_\perp \end{bmatrix}$$

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

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

□

这个结论也有一个推论。

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

证明.

$$\operatorname{tr}(\boldsymbol{X}^\star\boldsymbol{A}\boldsymbol{X})=\sum_{i=1}^{k}\boldsymbol{x}_i^\star\boldsymbol{A}\boldsymbol{x}_i=\sum_{i=1}^{k}R(\boldsymbol{A},\boldsymbol{x}_i).$$

由推论 3.7 知 $\lambda_i(\boldsymbol{X}^\star\boldsymbol{A}\boldsymbol{X})\geq\lambda_i(\boldsymbol{A})$，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 $\boldsymbol{A}\to\boldsymbol{A}+\boldsymbol{E}$，其中 $\boldsymbol{E}$ 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

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

Weyl 不等式很好地回答了这个问题。

定理 3.9 (Weyl 不等式)**.**

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

第一个等号成立当且仅当

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

第二个不等号成立当且仅当

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

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

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

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

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

由 $\boldsymbol{x} \in S_3$ 有

$$\lambda_i(\boldsymbol{A}+\boldsymbol{B}) \le \boldsymbol{x}^\star(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x} = \boldsymbol{x}^\star\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^\star\boldsymbol{B}\boldsymbol{x}.$$

由 $x \in S_1$ 知 $\boldsymbol{x}^\star\boldsymbol{A}\boldsymbol{x} \le \lambda_{i+j}(\boldsymbol{A})$；由 $\boldsymbol{x} \in S_2$ 知 $\boldsymbol{x}^\star\boldsymbol{B}\boldsymbol{x} \le \lambda_{n-j}(\boldsymbol{B})$. 因此

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

等号成立当且仅当

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

对于另一侧的不等式——想必你也能猜到——我们只需对 $-\boldsymbol{A}, -\boldsymbol{B}$ 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

$$\lambda_i(\boldsymbol{A}) + \lambda_1(\boldsymbol{B}) \le \lambda_i(\boldsymbol{A}+\boldsymbol{B}) \le \lambda_i(\boldsymbol{A}) + \lambda_n(\boldsymbol{B}).$$

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

$$\lambda_1(\boldsymbol{B}) \le \lambda_i(\boldsymbol{A}+\boldsymbol{B}) - \lambda_i(\boldsymbol{A}) \le \lambda_n(\boldsymbol{B}).$$

这意味着，第 i 个特征值的变化量被扰动矩阵 $\boldsymbol{B}$ 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）$\|\boldsymbol{B}\|_2 = \max(|\lambda_1(\boldsymbol{B})|, |\lambda_n(\boldsymbol{B})|)$，我们可以得到一个更简洁和著名的结果：

$$|\lambda_i(\boldsymbol{A}+\boldsymbol{B}) - \lambda_i(\boldsymbol{A})| \le \|\boldsymbol{B}\|_2.$$

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 $\boldsymbol{B}$ 的谱范数。因此，只要扰动 $\boldsymbol{B}$ 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)**.**

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

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 $(\lambda_1(\boldsymbol{A}),\dots,\lambda_n(\boldsymbol{A}))$ 及 $(\lambda_1(\boldsymbol{A}+\Delta\boldsymbol{A}),\dots,\lambda_n(\boldsymbol{A}+\delta\boldsymbol{A}))$，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 $\boldsymbol{A}(t) = \boldsymbol{A} + t\Delta\boldsymbol{A}$.

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 $\boldsymbol{A}(t) = (\boldsymbol{A}(t))^\star$，它存在解析形式的谱分解，即存在 $\boldsymbol{Q}(t), \boldsymbol{\Lambda}(t)$ 使得

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

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

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

$$\boldsymbol{Q}'(t)^\star\boldsymbol{Q}(t) + \boldsymbol{Q}(t)^\star\boldsymbol{Q}'(t) = \boldsymbol{0}.$$

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

对 $\boldsymbol{\Lambda}(t) = \boldsymbol{Q}(t)^\star\boldsymbol{A}(t)\boldsymbol{Q}(t)$ 求导得

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

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

$$\begin{aligned}
\boldsymbol{\Lambda}'(t) &= \operatorname{diag}(\boldsymbol{\Lambda}'(t)) \\
&= \operatorname{diag}(\boldsymbol{K}(t)\boldsymbol{\Lambda}(t) + \boldsymbol{Q}(t)^\star \Delta \boldsymbol{A}\boldsymbol{Q}(t) + \boldsymbol{\Lambda}(t)\boldsymbol{K}(t)^\star) \\
&= \operatorname{diag}(\boldsymbol{K}(t))\boldsymbol{\Lambda}(t) + \operatorname{diag}(\boldsymbol{Q}(t)^\star \Delta \boldsymbol{A}\boldsymbol{Q}(t)) + \boldsymbol{\Lambda}(t)\operatorname{diag}(\boldsymbol{K}(t)^\star) \\
&= \mathrm{i}\boldsymbol{D}(t)\boldsymbol{\Lambda}(t) + \operatorname{diag}(\boldsymbol{Q}(t)^\star \Delta \boldsymbol{A}\boldsymbol{Q}(t)) - \boldsymbol{\Lambda}(t)\mathrm{i}\boldsymbol{D}(t) \\
&= \operatorname{diag}(\boldsymbol{Q}(t)^\star \Delta \boldsymbol{A}\boldsymbol{Q}(t)).
\end{aligned}$$

于是

$$\begin{aligned}
\sum_{i=1}^{n} |\lambda_i(\boldsymbol{A} + \Delta\boldsymbol{A}) - \lambda_i(\boldsymbol{A})|^2 = \|\Delta\boldsymbol{\Lambda}\|_{\mathrm{F}}^2 \\
&= \|\boldsymbol{\Lambda}(1) - \boldsymbol{\Lambda}(0)\|_{\mathrm{F}}^2 \\
&= \left\| \int_0^1 \boldsymbol{\Lambda}'(t)\,\mathrm{d}t \right\|_{\mathrm{F}}^2 \\
&\le \left(\int_0^1 \|\boldsymbol{\Lambda}'(t)\|_{\mathrm{F}}\,\mathrm{d}t \right)^2 \\
&= \left(\int_0^1 \|\operatorname{diag}(\boldsymbol{Q}(t)^\star \Delta \boldsymbol{A}\boldsymbol{Q}(t))\|_{\mathrm{F}}\,\mathrm{d}t \right)^2 \\
&\le \left(\int_0^1 \|\boldsymbol{Q}(t)^\star \Delta \boldsymbol{A}\boldsymbol{Q}(t)\|_{\mathrm{F}}\,\mathrm{d}t \right)^2 \\
&= \left(\int_0^1 \|\Delta\boldsymbol{A}\|_{\mathrm{F}}\,\mathrm{d}t \right)^2 \\
&= \|\Delta\boldsymbol{A}\|_{\mathrm{F}}^2.
\end{aligned}$$

□