由于 $\boldsymbol{\Phi}$ 是一酉矩阵，所以 $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$ 具有相同的信号子空间和噪声子空间，即子阵列 1 和子阵列 2 具有相同的观测空间 (信号子空间 + 噪声子空间)。这就是等距线阵的平移不变性的物理解释。

由式 (8.8.23) 得

$$\begin{aligned}\boldsymbol{R}_{xx} &= \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}} + \sigma^2\boldsymbol{I}\\ &= [\boldsymbol{U}_s, \boldsymbol{U}_n]\begin{bmatrix}\boldsymbol{\Sigma}_s & \boldsymbol{O}\\ \boldsymbol{O} & \sigma^2\boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_s^{\mathrm{H}}\\ \boldsymbol{U}_n^{\mathrm{H}}\end{bmatrix}\\ &= [\boldsymbol{U}_s\boldsymbol{\Sigma}_s, \sigma^2\boldsymbol{U}_n]\begin{bmatrix}\boldsymbol{U}_s^{\mathrm{H}}\\ \boldsymbol{U}_n^{\mathrm{H}}\end{bmatrix}\\ &= \boldsymbol{U}_s\boldsymbol{\Sigma}_s\boldsymbol{U}_s^{\mathrm{H}} + \sigma^2\boldsymbol{U}_n\boldsymbol{U}_n^{\mathrm{H}}\end{aligned} \tag{8.8.56}$$

由于 $\boldsymbol{I} - \boldsymbol{U}_n\boldsymbol{U}_n^{\mathrm{H}} = \boldsymbol{U}_s\boldsymbol{U}_s^{\mathrm{H}}$，故由式 (8.8.56) 易知

$$\boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}} + \sigma^2\boldsymbol{U}_s\boldsymbol{U}_s^{\mathrm{H}} = \boldsymbol{U}_s\boldsymbol{\Sigma}_s\boldsymbol{U}_s^{\mathrm{H}} \tag{8.8.57}$$

用 $\boldsymbol{U}_s$ 右乘上式两边，注意到 $\boldsymbol{U}_s^{\mathrm{H}}\boldsymbol{U}_s = \boldsymbol{I}$，并加以重排，即得

$$\boldsymbol{U}_s = \boldsymbol{A}\boldsymbol{T} \tag{8.8.58}$$

式中，$\boldsymbol{T}$ 是一个非奇异矩阵，且

$$\boldsymbol{T} = \boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}_s(\boldsymbol{\Sigma}_s - \sigma^2\boldsymbol{I})^{-1} \tag{8.8.59}$$

虽然 $\boldsymbol{T}$ 是一未知矩阵，但它只是下面分析中的一个“虚拟参数”，我们只用到它的非奇异性。用 $\boldsymbol{T}$ 右乘式 (8.8.52)，则有

$$\boldsymbol{A}\boldsymbol{T} = \begin{bmatrix}\boldsymbol{A}_1\boldsymbol{T}\\ \text{最后一行}\end{bmatrix} = \begin{bmatrix}\text{第 1 行}\\ \boldsymbol{A}_2\boldsymbol{T}\end{bmatrix} \tag{8.8.60}$$

采用相同的分块形式，将 $\boldsymbol{U}_s$ 也分块成

$$\boldsymbol{U}_s = \begin{bmatrix}\boldsymbol{U}_1\\ \text{最后一行}\end{bmatrix} = \begin{bmatrix}\text{第 1 行}\\ \boldsymbol{U}_2\end{bmatrix} \tag{8.8.61}$$

由于 $\boldsymbol{A}\boldsymbol{T} = \boldsymbol{U}_s$，故比较式 (8.8.60) 与式 (8.8.61)，立即有

$$\boldsymbol{U}_1 = \boldsymbol{A}_1\boldsymbol{T} \tag{8.8.62}$$

$$\boldsymbol{U}_2 = \boldsymbol{A}_2\boldsymbol{T} \tag{8.8.63}$$

将式 (8.8.53) 代入式 (8.8.63)，即有

$$\boldsymbol{U}_2 = \boldsymbol{A}_1\boldsymbol{\Phi}\boldsymbol{T} \tag{8.8.64}$$

由式 (8.8.62) 及式 (8.8.64)，又有

$$\boldsymbol{U}_1\boldsymbol{T}^{-1}\boldsymbol{\Phi}\boldsymbol{T} = \boldsymbol{A}_1\boldsymbol{T}\boldsymbol{T}^{-1}\boldsymbol{\Phi}\boldsymbol{T} = \boldsymbol{A}_1\boldsymbol{\Phi}\boldsymbol{T} = \boldsymbol{U}_2 \tag{8.8.65}$$

定义

$$\boldsymbol{\Psi} = \boldsymbol{T}^{-1}\boldsymbol{\Phi}\boldsymbol{T} \tag{8.8.66}$$

矩阵 $\boldsymbol{\Psi}$ 称为矩阵 $\boldsymbol{\Phi}$ 的相似变换，因此它们具有相同的特征值，即 $\boldsymbol{\Psi}$ 的特征值也为 $\mathrm{e}^{\mathrm{j}\phi_m}$，$m=1,2,\cdots,M$。

将式 (8.8.66) 代入式 (8.8.65)，则得到一个重要的关系式，即

$$\boldsymbol{U}_2 = \boldsymbol{U}_1\boldsymbol{\Psi} \tag{8.8.67}$$

式 (8.8.67) 启迪了基本 ESPRIT 算法的另一种算法。

算法 8.8.6 (基本 ESPRIT 算法 2)

步骤 1 计算阵列协方差矩阵 $\hat{\boldsymbol{R}}_{xx}$ 的特征值分解 $\hat{\boldsymbol{R}}_{xx} = \hat{\boldsymbol{U}}\boldsymbol{\Sigma}\hat{\boldsymbol{U}}^{\mathrm{H}}$。

步骤 2 矩阵 $\hat{\boldsymbol{U}}$ 与 $\hat{\boldsymbol{R}}_{xx}$ 的 p 个主特征值对应的部分组成 $\hat{\boldsymbol{U}}_s$。

步骤 3 抽取 $\hat{\boldsymbol{U}}_s$ 的前面 $m-1$ 行组成矩阵 $\hat{\boldsymbol{U}}_1$，后面 $m-1$ 行组成矩阵 $\hat{\boldsymbol{U}}_2$。计算 $\boldsymbol{\Psi} = (\hat{\boldsymbol{U}}_1^{\mathrm{H}}\hat{\boldsymbol{U}}_1)^{-1}\hat{\boldsymbol{U}}_1^{\mathrm{H}}\hat{\boldsymbol{U}}_2$ 的特征值分解。矩阵 $\boldsymbol{\Psi}$ 的特征值 $\mathrm{e}^{\mathrm{j}\omega_i}(i=1,2,\cdots,p)$ 给出估计值 $\hat{\omega}_i$, $i=1,2,\cdots,p$。

ESPRIT 方法在通信信号处理尤其是在空时二维处理中有着重要的应用，感兴趣的读者可参考文献 [523]。

8.9 Rayleigh 商

在物理和信息科学技术中，常常会遇到 Hermitian 矩阵的二次型函数的商的最大化或者最小化。这种商有两种形式，它们分别是一个 Hermitian 矩阵的 Rayleigh 商 (有时也叫 Rayleigh-Ritz 比) 和两个 Hermitian 矩阵的广义 Rayleigh 商 (或广义 Rayleigh-Ritz 比)。

8.9.1 Rayleigh 商

在研究振动系统的小振荡时，为了找到合适的广义坐标，Rayleigh 于 1930 年代提出了一种特殊形式的商 [384]，被后人称为 Rayleigh 商。下面是现在被广泛采用的 Rayleigh 商定义。

定义 8.9.1 Hermitian 矩阵 $\boldsymbol{A} \in C^{n\times n}$ 的 Rayleigh 商或 Rayleigh-Ritz 比 $R(\boldsymbol{x})$ 是一个标量，定义为

$$R(\boldsymbol{x}) = R(\boldsymbol{x},\boldsymbol{A}) = \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \tag{8.9.1}$$

其中，$\boldsymbol{x}$ 是待选择的向量，其目的是使 Rayleigh 商最大化或者最小化。

Rayleigh 商的重要性质如下 [355],[356],[81],[211]。

性质 1 (齐次性) 若 α 和 β 为标量，则

$$R(\alpha\boldsymbol{x},\beta\boldsymbol{A}) = \beta R(\boldsymbol{x},\boldsymbol{A}) \tag{8.9.2}$$

证明 根据定义 8.9.1，易知

$$R(\alpha\boldsymbol{x},\beta\boldsymbol{A})=\frac{\alpha^*\boldsymbol{x}^{\mathrm{H}}\beta\boldsymbol{A}\alpha\boldsymbol{x}}{\alpha^*\boldsymbol{x}^{\mathrm{H}}\alpha\boldsymbol{x}}=\beta R(\boldsymbol{x},\boldsymbol{A})$$

这即是需要的结果。 ■

性质 2 (平移不变性)

$$R(\boldsymbol{x},\boldsymbol{A}-\alpha\boldsymbol{I})=R(\boldsymbol{x},\boldsymbol{A})-\alpha \tag{8.9.3}$$

证明 直接计算 Rayleigh 商，得

$$R(\boldsymbol{x},\boldsymbol{A}-\alpha\boldsymbol{I})=\frac{\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{A}-\alpha\boldsymbol{I})\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}-\frac{\boldsymbol{x}^{\mathrm{H}}\alpha\boldsymbol{I}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=R(\boldsymbol{x},\boldsymbol{A})-\alpha$$

这就证明了本性质。 ■

性质 3 (正交性)

$$\boldsymbol{x}\perp(\boldsymbol{A}-R(\boldsymbol{x})\boldsymbol{I})\boldsymbol{x} \tag{8.9.4}$$

证明 由定义 8.9.1 知 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=R(\boldsymbol{x})\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}$，或写作 $\boldsymbol{x}^{\mathrm{H}}[\boldsymbol{A}-R(\boldsymbol{x})\boldsymbol{I}]\boldsymbol{x}=0$，即有 $\boldsymbol{x}\perp(\boldsymbol{A}-R(\boldsymbol{x})\boldsymbol{I})\boldsymbol{x}$。 ■

性质 4 (有界性) 当向量 $\boldsymbol{x}$ 在所有非零向量的范围变化时，Rayleigh 商 $R(\boldsymbol{x})$ 落在一复平面的区域 (称为矩阵 $\boldsymbol{A}$ 的值域) 内。这一区域是闭合的、有界的和凸的。若 $\boldsymbol{A}$ 是 Hermitian 的，即满足 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$，则这一区域是一个闭区间 $[\lambda_1,\lambda_n]$。

性质 5 (最小残差) 对于所有向量 $\boldsymbol{x}\neq\boldsymbol{0}$ 和所有标量 μ，恒有

$$\|[\boldsymbol{A}-R(\boldsymbol{x})\boldsymbol{I}]\boldsymbol{x}\|\leqslant\|[\boldsymbol{A}-\mu\boldsymbol{I}]\boldsymbol{x}\| \tag{8.9.5}$$

关于有界性，可进一步参考文献 [299]。

Hermitian 矩阵的 Rayleigh 商的有界性可以用下面的定理严格叙述。

定理 8.9.1 (Rayleigh-Ritz 定理) 令 $\boldsymbol{A}\in C^{n\times n}$ 是 Hermitian 的，并令 $\boldsymbol{A}$ 的特征值按递增次序

$$\lambda_{\min}=\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_{n-1}\leqslant\lambda_n=\lambda_{\max} \tag{8.9.6}$$

排列，则

$$\boxed{\max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\max_{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}=1}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\lambda_{\max},\quad 若\ \boldsymbol{A}\boldsymbol{x}=\lambda_{\max}\boldsymbol{x}} \tag{8.9.7}$$

$$\boxed{\min_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\min_{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}=1}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\lambda_{\min},\quad 若\ \boldsymbol{A}\boldsymbol{x}=\lambda_{\min}\boldsymbol{x}} \tag{8.9.8}$$

更一般地，矩阵 $\boldsymbol{A}$ 的所有特征向量和特征值分别称为 Rayleigh 商 $R(\boldsymbol{x})$ 的临界点 (critical point) 和临界值 (critical value)。

这个定理的证明方法有多种，例如参考文献 [184]，[211]，[95]。下面采用的是文献 [184] 的证明方法。

证明　若 $\boldsymbol{x}\neq \mathbf{0}$，则

$$\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\left(\frac{\boldsymbol{x}}{\sqrt{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}}\right)^{\mathrm{H}}\boldsymbol{A}\left(\frac{\boldsymbol{x}}{\sqrt{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}}\right)$$

和

$$\left(\frac{\boldsymbol{x}}{\sqrt{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}}\right)^{\mathrm{H}}\left(\frac{\boldsymbol{x}}{\sqrt{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}}\right)=1$$

于是，Rayleigh 商具有下面的等价表示：

$$\left.\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right|_{\boldsymbol{x}\neq 0}=\left.\frac{\tilde{\boldsymbol{x}}^{\mathrm{H}}\boldsymbol{A}\tilde{\boldsymbol{x}}}{\tilde{\boldsymbol{x}}^{\mathrm{H}}\tilde{\boldsymbol{x}}}\right|_{\tilde{\boldsymbol{x}}^{\mathrm{H}}\tilde{\boldsymbol{x}}=1}\tag{8.9.9}$$

式中，$\tilde{\boldsymbol{x}}=\dfrac{\boldsymbol{x}}{\sqrt{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}}$ 是具有单位范数 $\tilde{\boldsymbol{x}}^{\mathrm{H}}\tilde{\boldsymbol{x}}=1$ 的向量。式 (8.9.9) 表明，在使 Rayleigh 商最小化或者最大化时，可以选择具有单位范数的向量 $\boldsymbol{x}$。

因为 $\boldsymbol{A}$ 是 Hermitian 矩阵，所以存在一个酉矩阵 $\boldsymbol{U}$，使得 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$，其中，$\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$。对于任一列向量 $\boldsymbol{x}\in C^n$，我们有

$$\begin{aligned}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}&=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{x}=(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{x})^{\mathrm{H}}\boldsymbol{\Lambda}(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{x})\\&=\sum_{i=1}^{n}\lambda_i\left|(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{x})_i\right|^2\end{aligned}$$

由于每一项 $\left|(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{x})_i\right|$ 都是非负的，故有

$$\lambda_{\min}\sum_{i=1}^{n}|(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{x})_i|^2\leqslant \boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=\sum_{i=1}^{n}\lambda_i|(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{x})_i|^2\leqslant\lambda_{\max}\sum_{i=1}^{n}|(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{x})_i|^2$$

注意到 $\boldsymbol{U}$ 是酉矩阵，又有

$$\sum_{i=1}^{n}\left|(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{x})_i\right|^2=\sum_{i=1}^{n}|x_i|^2=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}$$

这就证明了

$$\lambda_{\min}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}\leqslant\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}\leqslant\lambda_{\max}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}\tag{8.9.10}$$

若选择向量 $\boldsymbol{x}$ 是与 Hermitian 矩阵 $\boldsymbol{A}$ 的最小特征值 $\lambda_{\min}$ 相对应的特征向量，则

$$\boldsymbol{A}\boldsymbol{x}=\lambda_{\min}\boldsymbol{x}\Longrightarrow\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=\lambda_{\min}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}$$

故式 (8.9.10) 第一个不等式取等号的条件是 $\boldsymbol{A}\boldsymbol{x}=\lambda_{\min}\boldsymbol{x}$。类似地，若选择向量 $\boldsymbol{x}$ 是与 Hermitian 矩阵 $\boldsymbol{A}$ 的最大特征值 $\lambda_{\min}$ 相对应的特征向量，则

$$\boldsymbol{A}\boldsymbol{x}=\lambda_{\max}\boldsymbol{x}\Longrightarrow\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=\lambda_{\max}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}$$

即式 (8.9.10) 第二个不等式取等号的条件是 $\boldsymbol{A}\boldsymbol{x}=\lambda_{\max}\boldsymbol{x}$。

由于内积 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ 为标量，故式 (8.9.10) 可等价写作

$$\lambda_{\min} \leqslant \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \leqslant \lambda_{\max}$$

第一个不等式取等号的条件是 $\boldsymbol{A}\boldsymbol{x} = \lambda_{\min}\boldsymbol{x}$；第二个不等式取等号的条件是 $\boldsymbol{A}\boldsymbol{x} = \lambda_{\max}\boldsymbol{x}$。这就完成了定理的证明。 ■

下面考虑 Rayleigh 商的梯度与 Hessian 矩阵 (即梯度的梯度)[211],[95]。为简便计，将 Rayleigh 商 $R(\boldsymbol{x})$ 简记作 R。

Rayleigh 商的梯度为

$$\nabla_{\boldsymbol{x}} = \frac{\partial R}{\partial \boldsymbol{x}} = \frac{2}{\|\boldsymbol{x}\|_2^2}(\boldsymbol{A} - R\boldsymbol{I})\boldsymbol{x} \tag{8.9.11}$$

而 Rayleigh 商的 Hessian 矩阵为

$$\boldsymbol{H}_R = \frac{\partial^2 R}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}} = \frac{2}{\|\boldsymbol{x}\|_2^2}\left[\boldsymbol{A} - \nabla_{\boldsymbol{x}}(R)\boldsymbol{x}^{\mathrm{T}} - \boldsymbol{x}\nabla_{\boldsymbol{x}}^{\mathrm{T}}(R)\boldsymbol{x} - R\boldsymbol{I}\right] \tag{8.9.12}$$

令 $\boldsymbol{u}_i, \lambda_i, i = 1, 2, \cdots, n$ 分别是矩阵 $\boldsymbol{A}$ 的特征向量和特征值，即它们分别是 Rayleigh 商的临界点和临界值，即有

$$R(\boldsymbol{u}_i) = \lambda_i, \quad i = 1, 2, \cdots, n \tag{8.9.13}$$

计算 Hessian 矩阵在临界点 $\boldsymbol{u}_i$ 的值，易得

$$\boldsymbol{H}_R(\boldsymbol{u}_i) = \boldsymbol{A} - \lambda_i \boldsymbol{I}, \quad i = 1, 2, \cdots, n \tag{8.9.14}$$

由式 (8.9.14) 可得两个重要结果：

(1) Hessian 矩阵的行列式

$$\det[\boldsymbol{H}_R(\boldsymbol{u}_i)] = \det(\boldsymbol{A} - \lambda_i \boldsymbol{I}) = 0 \tag{8.9.15}$$

因为 $\boldsymbol{A} - z\boldsymbol{I}$ 是矩阵 $\boldsymbol{A}$ 的特征多项式。上式意味着 Hessian 矩阵 $\boldsymbol{H}_R(\boldsymbol{u}_i)$ 对于所有临界点 $\boldsymbol{u}_i$ 都是奇异矩阵。

(2) 用向量 $\boldsymbol{u}_j$ 右乘式 (8.9.14)，立即有

$$\boldsymbol{H}_R(\boldsymbol{u}_i)\boldsymbol{u}_j = (\boldsymbol{A} - \lambda_i \boldsymbol{I})\boldsymbol{u}_j = \boldsymbol{A}\boldsymbol{u}_j - \lambda_i \boldsymbol{u}_j = \lambda_j \boldsymbol{u}_j - \lambda_i \boldsymbol{u}_j$$

因为 $\boldsymbol{A}\boldsymbol{u}_j = \lambda_j \boldsymbol{u}_j$。上式即是

$$\boldsymbol{H}_R(\boldsymbol{u}_i)\boldsymbol{u}_j = \begin{cases} 0, & j = i \\ (\lambda_j - \lambda_i)\boldsymbol{u}_j, & j \neq i \end{cases} \tag{8.9.16}$$

这说明，由 Rayleigh 商的临界点计算得到的 Hessian 矩阵 $\boldsymbol{H}_R(\boldsymbol{u}_i)$ 与矩阵 $\boldsymbol{A}$ 具有相同的特征向量，但特征值不同。此外，由于 $\lambda_j - \lambda_{\min} \geqslant 0$，故只有在临界点 $\boldsymbol{u}_{\min}$ 的 Hessian 矩阵是半正定的，满足 $\boldsymbol{H}_R(\boldsymbol{u}_{\min}) \geqslant 0$。

关于 Rayleigh 商，还有下面的重要定理 [270]。

定理 8.9.2 (Courant-Fischer 定理) 令 $n\times n$ 矩阵 $\boldsymbol{A}$ 为 Hermitian 矩阵，其特征值为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$，则 λ_k $(1\leqslant k\leqslant n)$ 服从关系式

$$\lambda_k=\min_{S,\dim(S)=n-k+1}\ \max_{\boldsymbol{x}\in S,\boldsymbol{x}\neq\boldsymbol{0}}\left\{\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right\}\tag{8.9.17}$$

证明 令 $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_n$ 分别是矩阵 $\boldsymbol{A}$ 与特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 对应的特征向量，并且它们组成子空间 S 的一组标准正交基，即 $S=\mathrm{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_n\}$。又令 $\mu(S)=\max\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}/(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x})$ 表示 Rayleigh 商在子空间 S 的所有非零向量范围内的极大值。若 S_k 是由前 k 个特征向量张成的子空间，即 $S_k=\mathrm{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_k\}$，则 k 维向量空间 S_k 是 n 维向量空间 S 的子空间，这意味着在 S 和 S_k 的交集中存在非零向量 $\boldsymbol{x}$。因此，向量 $\boldsymbol{x}$ 是 S_k 的基向量的线性组合，即 $\boldsymbol{x}=\sum\limits_{i=1}^{k}\alpha_i\boldsymbol{x}_i$。于是，易得

$$\mu(S)=\max_{\boldsymbol{x}\in S,\boldsymbol{x}\neq\boldsymbol{0}}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\frac{\sum\limits_{i=1}^{k}\lambda_i|\alpha_i|^2}{\sum\limits_{i=1}^{k}|\alpha_i|^2}\geqslant\lambda_k$$

即 $\max\limits_{\boldsymbol{x}\in S,\boldsymbol{x}\neq\boldsymbol{0}}\dfrac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\geqslant\lambda_k$。

类似地，由于 $n-k+1$ 维子空间 $S_0=\mathrm{Span}\{\boldsymbol{u}_{k+1},\boldsymbol{u}_{k+2},\cdots,\boldsymbol{u}_n\}$ 也是 n 维向量空间 S 的子空间，所以交集 $S_0\cap S$ 中的向量 $\boldsymbol{x}$ 可以表示为线性组合 $\boldsymbol{x}=\sum\limits_{i=k}^{n}\alpha_i\boldsymbol{u}_i$，从而有

$$\mu(S_0)=\min_{\boldsymbol{x}\in S_0,\dim(S_0)=n-k+1}\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\frac{\sum\limits_{i=k}^{n}\lambda_i|\alpha_i|^2}{\sum\limits_{i=k}^{n}|\alpha_i|^2}\leqslant\lambda_k$$

综合 $\mu(S)\geqslant\lambda_k$ 和 $\mu(S_0)\leqslant\lambda_k$，即得到式 (8.9.17)。 ■

Courant-Fisher 定理也称 Courant-Fisher 极小 – 极大定理或极小 – 极大原理 (min-max principle)。事实上，除了式 (8.9.17) 外，Courant-Fisher 定理还有其他三种表示形式，例如极大 – 极小原理为

$$\lambda_k=\max_{S,\dim S=k}\ \min_{\boldsymbol{x}\in S,\boldsymbol{x}\neq\boldsymbol{0}}\left\{\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right\}\tag{8.9.18}$$

另外，如果特征值按照递增次序排列为 $\lambda_1\leqslant\lambda_2\cdots\leqslant\lambda_n$，则 Rayleigh 商的极小 – 极大原理和极大 – 极小原理分别与式 (8.9.17) 和式 (8.9.18) 有所不同。这一点希望加以注意。

下面介绍 Rayleigh 商的几种典型应用。

8.9.2 应用举例 1：特征滤波器

有限冲激响应 (FIR) 滤波器的最小二乘设计方法可分为两类：矩阵求逆方法 [451],[69] 以及特征方法 [455],[361],[362],[363],[333]。矩阵求逆方法由 Tufts 与 Francis 于 1970 年提

出，它利用矩阵求逆，求解线性方程组，得到滤波器系数。特征方法由 Vaidynathan 与 Nguyen[455] 于 1987 年提出，它将滤波器系数向量的求解转换为一个实对称的正定矩阵的特征向量的计算。

设一个因果的 N 阶 FIR 滤波器的传递函数为

$$H(z)=\sum_{n=0}^{N}h(n)z^{-n} \tag{8.9.19}$$

滤波器称为具有线性相位，若其相位响应是频率的线性函数。这里考虑滤波器的幅值响应的设计。

无论滤波器的长度 N 取偶或奇数，也无论滤波器冲激响应 $h(n)$ 是否对称，滤波器的幅值响应 $A(\omega)$ 都可以表示为

$$A(\omega)=\sum_{n=1}^{M}a_n p(\omega,n) \tag{8.9.20}$$

式中，$p(\omega,n)$ 是一个合适选择的三角函数，系数 a_n 与滤波器的冲激响应 $h(n)$ 有关。

定义

$$\boldsymbol{a}=[a_1,a_2,\cdots,a_M]^{\mathrm{T}} \tag{8.9.21}$$

$$\boldsymbol{p}(\omega)=[p(\omega,1),p(\omega,2),\cdots,p(\omega,M)]^{\mathrm{T}} \tag{8.9.22}$$

则式 (8.9.20) 可以写作

$$A(\omega)=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{p}(\omega)=\boldsymbol{p}^{\mathrm{T}}(\omega)\boldsymbol{a} \tag{8.9.23}$$

滤波器设计问题的提法是：给定期望幅值响应 $D(\omega)$，求系数向量 $\boldsymbol{a}$，使得由式 (8.9.23) 设计的幅值响应 $A(\omega)$ 尽可能逼近 $D(\omega)$。

Tufts 与 Francis 提出使用下列目标函数 [451]

$$\begin{aligned}J_{\mathrm{TF}}(\boldsymbol{a})&=\int_{\omega\in R}[D(\omega)-A(\omega)]^2\mathrm{d}\omega\\&=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{Q}_{\mathrm{TF}}\boldsymbol{a}-2\boldsymbol{q}^{\mathrm{T}}\boldsymbol{a}+d\end{aligned} \tag{8.9.24}$$

式中，$R=[0,2\pi]$ 表示一角频率闭区间，并且

$$\boldsymbol{Q}_t=\int_{\omega\in R}\boldsymbol{p}(\omega)\boldsymbol{p}^{\mathrm{T}}(\omega)\mathrm{d}\omega \tag{8.9.25}$$

$$\boldsymbol{q}=\int_{\omega\in R}D(\omega)\boldsymbol{p}(\omega)\mathrm{d}\omega \tag{8.9.26}$$

$$d=\int_{\omega\in R}D^2(\omega)\mathrm{d}\omega \tag{8.9.27}$$

由 $\dfrac{\partial J_{\mathrm{TF}}(\boldsymbol{a})}{\partial \boldsymbol{a}}=0$，立即得

$$2\boldsymbol{Q}_{\mathrm{TF}}\boldsymbol{a}-2\boldsymbol{q}=\boldsymbol{0}$$

或者

$$a_{\mathrm{TF}} = Q_{\mathrm{TF}}^{-1} q \tag{8.9.28}$$

由于系数向量 a_{TF} 可以通过矩阵求逆和上式得到，所以 Tufts 和 Francis 的设计方法称为矩阵求逆方法。

在文献 [455] 中，Vaidynathan 与 Nguyen 提出使用目标函数

$$J_{\mathrm{VN}}(a) = \int_{\omega\in R} \left| \frac{D(\omega)}{D(\omega_0)} A(\omega_0) - A(\omega) \right|^2 \mathrm{d}\omega \tag{8.9.29}$$

式中，ω_0 为满足 $D(\omega_0) \neq 0$ 的参考频率点。利用式 (8.9.23)，上述目标函数又可重新写作

$$\begin{aligned} J_{\mathrm{VN}}(a) &= \int_{\omega\in R} \left| \frac{D(\omega)}{D(\omega_0)} a^{\mathrm{T}} p(\omega_0) - a^{\mathrm{T}} p(\omega) \right|^2 \mathrm{d}\omega \\ &= a^{\mathrm{T}} Q_{\mathrm{VN}} a \end{aligned} \tag{8.9.30}$$

式中

$$Q_{\mathrm{VN}} = \int_{\omega\in R} \left(\frac{D(\omega)}{D(\omega_0)} p(\omega_0) - p(\omega) \right) \left(\frac{D(\omega)}{D(\omega_0)} p(\omega_0) - p(\omega) \right)^{\mathrm{T}} \mathrm{d}\omega \tag{8.9.31}$$

由定义知，矩阵 Q_{VN} 是一个实对称的正定矩阵。

令 $\dfrac{J_{\mathrm{DN}}(a)}{\partial a} = 0$，得 $Q_{\mathrm{VN}} a = 0$，只有平凡解 $a = 0$。为了避免这种平凡解，最简单和最有效的做法是将无约束的最优化问题 $\min J_{\mathrm{VN}}(a) = \min a^{\mathrm{T}} Q_{\mathrm{VN}} a$ 变成下列有约束的最优化问题

$$\min a^{\mathrm{T}} Q_{\mathrm{VN}} a \tag{8.9.32}$$

约束条件为

$$a^{\mathrm{T}} a = 1$$

很显然，这一个带约束的最优化问题又可等价写作

$$\min R(a) = \min \left\{ \frac{a^{\mathrm{T}} Q_{\mathrm{VN}} a}{a^{\mathrm{T}} a} \right\} \tag{8.9.33}$$

这是一个典型的 Rayleigh 商的最小化问题。因此，系数向量 a 等于实对称的正定矩阵 Q_{VN} 与最小特征值对应的特征向量。由于系数向量的确定转换为实对称正定矩阵的特征向量计算，故 Vaidynathan 与 Nguyen 的滤波器设计方法称为特征滤波器方法，而由此方法设计的滤波器称为特征滤波器 (eigenfilter)。

一个实对称矩阵的最小特征向量可以利用前面介绍的共轭梯度法或者 Rayleigh 商迭代法计算。

2001 年，Pei 与 Tseng[363] 提出了一种新的特征滤波器，它基于总体最小二乘误差准