Chapter 2
Canonical Correlation Analysis

Canonical correlation analysis (CCA), which is a multivariate analysis method, tries to quantify the amount of linear relationships between two sets of random variables, leading to different modes of maximum correlation [1]. In this chapter, we explain how CCA works from a signal processing perspective.

2.1 Preliminaries

Let $\mathbf{a}$ and $\mathbf{b}$ be two zero-mean complex-valued circular random vectors of lengths $L_{\mathbf{a}}$ and $L_{\mathbf{b}}$, respectively, where, without loss of generality, it is assumed that $L_{\mathbf{a}} \leq L_{\mathbf{b}}$. Let us define the longer random vector:

$$\mathbf{c} = \left[\mathbf{a}^T \ \mathbf{b}^T \right]^T , \tag{2.1}$$

where the superscript ^{T} denotes transpose of a vector or a matrix. The covariance matrix of $\mathbf{c}$ is then

$$\begin{aligned} \mathbf{\Phi} &= E\left(\mathbf{c}\mathbf{c}^H\right) \\ &= \begin{bmatrix} \mathbf{\Phi}_{\mathbf{a}} & \mathbf{\Phi}_{\mathbf{ab}} \\ \mathbf{\Phi}_{\mathbf{ba}} & \mathbf{\Phi}_{\mathbf{b}} \end{bmatrix}, \end{aligned} \tag{2.2}$$

where $E(\cdot)$ denotes mathematical expectation, the superscript ^{H} is the conjugate-transpose operator, $\mathbf{\Phi}_{\mathbf{a}} = E\left(\mathbf{a}\mathbf{a}^H\right)$ is the covariance matrix of $\mathbf{a}$, $\mathbf{\Phi}_{\mathbf{b}} = E\left(\mathbf{b}\mathbf{b}^H\right)$ is the covariance matrix of $\mathbf{b}$, $\mathbf{\Phi}_{\mathbf{ab}} = E\left(\mathbf{a}\mathbf{b}^H\right)$ is the cross-covariance matrix between $\mathbf{a}$ and $\mathbf{b}$, and $\mathbf{\Phi}_{\mathbf{ba}} = \mathbf{\Phi}_{\mathbf{ab}}^H$. It is assumed that $\text{rank}\left(\mathbf{\Phi}_{\mathbf{ab}}\right) = L_{\mathbf{a}}$ and, unless stated otherwise, $\mathbf{\Phi}_{\mathbf{a}}$ and $\mathbf{\Phi}_{\mathbf{b}}$ are assumed to have full rank.

J. Benesty and I. Cohen, *Canonical Correlation Analysis in Speech Enhancement*,
SpringerBriefs in Electrical and Computer Engineering,
DOI 10.1007/978-3-319-67020-1_2

The Schur complement of $\mathbf{\Phi_a}$ in $\mathbf{\Phi}$ is defined as [2]

$$\mathbf{\Phi_{b/a}} = \mathbf{\Phi_b} - \mathbf{\Phi_{ba}}\mathbf{\Phi_a^{-1}}\mathbf{\Phi_{ab}} \tag{2.3}$$

and the Schur complement of $\mathbf{\Phi_b}$ in $\mathbf{\Phi}$ is

$$\mathbf{\Phi_{a/b}} = \mathbf{\Phi_a} - \mathbf{\Phi_{ab}}\mathbf{\Phi_b^{-1}}\mathbf{\Phi_{ba}}. \tag{2.4}$$

Clearly, both matrices $\mathbf{\Phi_{b/a}}$ and $\mathbf{\Phi_{a/b}}$ are nonsingular. An equivalent and more interesting way to express the two previous expressions is

$$\begin{aligned}\mathbf{\Phi_b}^{-1/2}\mathbf{\Phi_{b/a}}\mathbf{\Phi_b}^{-1/2} &= \mathbf{I}_{L_b} - \mathbf{\Phi_b}^{-1/2}\mathbf{\Phi_{ba}}\mathbf{\Phi_a}^{-1}\mathbf{\Phi_{ab}}\mathbf{\Phi_b}^{-1/2} \\ &= \mathbf{I}_{L_b} - \left(\mathbf{\Phi_b}^{-1/2}\mathbf{\Phi_{ba}}\mathbf{\Phi_a}^{-1/2}\right)\left(\mathbf{\Phi_a}^{-1/2}\mathbf{\Phi_{ab}}\mathbf{\Phi_b}^{-1/2}\right) \\ &= \mathbf{I}_{L_b} - \mathbf{\Theta}\mathbf{\Theta}^H \\ &= \mathbf{I}_{L_b} - \mathbf{\Psi_b}\end{aligned} \tag{2.5}$$

and

$$\begin{aligned}\mathbf{\Phi_a}^{-1/2}\mathbf{\Phi_{a/b}}\mathbf{\Phi_a}^{-1/2} &= \mathbf{I}_{L_a} - \mathbf{\Phi_a}^{-1/2}\mathbf{\Phi_{ab}}\mathbf{\Phi_b}^{-1}\mathbf{\Phi_{ba}}\mathbf{\Phi_a}^{-1/2} \\ &= \mathbf{I}_{L_a} - \left(\mathbf{\Phi_a}^{-1/2}\mathbf{\Phi_{ab}}\mathbf{\Phi_b}^{-1/2}\right)\left(\mathbf{\Phi_b}^{-1/2}\mathbf{\Phi_{ba}}\mathbf{\Phi_a}^{-1/2}\right) \\ &= \mathbf{I}_{L_a} - \mathbf{\Theta}^H\mathbf{\Theta} \\ &= \mathbf{I}_{L_a} - \mathbf{\Psi_a},\end{aligned} \tag{2.6}$$

where $\mathbf{I}_{L_b}$ and $\mathbf{I}_{L_a}$ are the identity matrices of sizes $L_\mathbf{b} \times L_\mathbf{b}$ and $L_\mathbf{a} \times L_\mathbf{a}$, respectively, $\mathbf{\Theta} = \mathbf{\Phi_b}^{-1/2}\mathbf{\Phi_{ba}}\mathbf{\Phi_a}^{-1/2}$, $\mathbf{\Psi_b} = \mathbf{\Theta}\mathbf{\Theta}^H$, and $\mathbf{\Psi_a} = \mathbf{\Theta}^H\mathbf{\Theta}$. From (2.5) and (2.6), it can be observed that the eigenvalues of $\mathbf{\Psi_b}$ and $\mathbf{\Psi_a}$ are always smaller than 1 (and, of course, positive or null). Furthermore, the matrices $\mathbf{\Psi_b}$ and $\mathbf{\Psi_a}$ have the same nonzero eigenvalues [2]. Indeed, since

$$\det\left(\mathbf{I}_{L_a} - \mathbf{\Psi_a}\right) = \det\left(\mathbf{I}_{L_b} - \mathbf{\Psi_b}\right), \tag{2.7}$$

where det stands for determinant, it follows that

$$\lambda^{L_b}\det\left(\lambda\mathbf{I}_{L_a} - \mathbf{\Psi_a}\right) = \lambda^{L_a}\det\left(\lambda\mathbf{I}_{L_b} - \mathbf{\Psi_b}\right), \tag{2.8}$$

which shows that $\mathbf{\Psi_a}$ and $\mathbf{\Psi_b}$ have the same nonzero characteristic roots. Obviously, $\mathbf{\Psi_a}$ is nonsingular while $\mathbf{\Psi_b}$ is singular.

Using the well-known eigenvalue decomposition [3], the Hermitian matrix $\mathbf{\Psi_a}$ can be diagonalized as

$$\mathbf{U_a}^H\mathbf{\Psi_a}\mathbf{U_a} = \mathbf{\Lambda_a}, \tag{2.9}$$

where

$$\mathbf{U_a} = \left[\mathbf{u}_{\mathbf{a},1} \ \mathbf{u}_{\mathbf{a},2} \ \cdots \ \mathbf{u}_{\mathbf{a},L_\mathbf{a}} \right] \tag{2.10}$$

is a unitary matrix, i.e., $\mathbf{U_a}^H \mathbf{U_a} = \mathbf{U_a} \mathbf{U_a}^H = \mathbf{I}_{L_\mathbf{a}}$, and

$$\mathbf{\Lambda_a} = \text{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_{L_\mathbf{a}}\right) \tag{2.11}$$

is a diagonal matrix. The orthonormal vectors $\mathbf{u}_{\mathbf{a},1}, \mathbf{u}_{\mathbf{a},2}, \ldots, \mathbf{u}_{\mathbf{a},L_\mathbf{a}}$ are the eigenvectors corresponding, respectively, to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{L_\mathbf{a}}$ of the matrix $\mathbf{\Psi_a}$, where $1 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{L_\mathbf{a}} > 0$. In the same way, the Hermitian matrix $\mathbf{\Psi_b}$ can be diagonalized as

$$\mathbf{U_b}^H \mathbf{\Psi_b} \mathbf{U_b} = \mathbf{\Lambda_b}, \tag{2.12}$$

where

$$\mathbf{U_b} = \left[\mathbf{u}_{\mathbf{b},1} \ \mathbf{u}_{\mathbf{b},2} \ \cdots \ \mathbf{u}_{\mathbf{b},L_\mathbf{b}} \right] \tag{2.13}$$

is a unitary matrix and

$$\begin{aligned} \mathbf{\Lambda_b} &= \text{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_{L_\mathbf{a}}, 0, 0, \ldots, 0\right) \\ &= \text{diag}\left(\mathbf{\Lambda_a}, \mathbf{0}\right) \end{aligned} \tag{2.14}$$

is a diagonal matrix.

For $l = 1, 2, \ldots, L_\mathbf{a}$, we have

$$\mathbf{\Psi_b} \mathbf{u}_{\mathbf{b},l} = \mathbf{\Theta}\mathbf{\Theta}^H \mathbf{u}_{\mathbf{b},l} = \lambda_l \mathbf{u}_{\mathbf{b},l}. \tag{2.15}$$

Left multiplying both sides of the previous equation by $\mathbf{\Theta}^H / \sqrt{\lambda_l}$, we get

$$\begin{aligned} \frac{\mathbf{\Theta}^H}{\sqrt{\lambda_l}} \mathbf{\Theta}\mathbf{\Theta}^H \mathbf{u}_{\mathbf{b},l} &= \mathbf{\Theta}^H \mathbf{\Theta} \frac{\mathbf{\Theta}^H \mathbf{u}_{\mathbf{b},l}}{\sqrt{\lambda_l}} \\ &= \lambda_l \frac{\mathbf{\Theta}^H \mathbf{u}_{\mathbf{b},l}}{\sqrt{\lambda_l}}. \end{aligned} \tag{2.16}$$

We deduce that

$$\mathbf{u}_{\mathbf{a},l} = \frac{\mathbf{\Theta}^H \mathbf{u}_{\mathbf{b},l}}{\sqrt{\lambda_l}}. \tag{2.17}$$

Similarly, for $l = 1, 2, \ldots, L_\mathbf{a}$, we have

$$\mathbf{\Psi_a} \mathbf{u}_{\mathbf{a},l} = \mathbf{\Theta}^H \mathbf{\Theta} \mathbf{u}_{\mathbf{a},l} = \lambda_l \mathbf{u}_{\mathbf{a},l}. \tag{2.18}$$

Left multiplying both sides of the previous expression by $\boldsymbol{\Theta}/\sqrt{\lambda_l}$, we get

$$\begin{aligned}\frac{\boldsymbol{\Theta}}{\sqrt{\lambda_l}}\boldsymbol{\Theta}^H\boldsymbol{\Theta}\mathbf{u}_{\mathbf{a},l} &= \boldsymbol{\Theta}\boldsymbol{\Theta}^H\frac{\boldsymbol{\Theta}\mathbf{u}_{\mathbf{a},l}}{\sqrt{\lambda_l}} \\ &= \lambda_l\frac{\boldsymbol{\Theta}\mathbf{u}_{\mathbf{a},l}}{\sqrt{\lambda_l}},\end{aligned} \tag{2.19}$$

from which we find that

$$\mathbf{u}_{\mathbf{b},l} = \frac{\boldsymbol{\Theta}\mathbf{u}_{\mathbf{a},l}}{\sqrt{\lambda_l}}. \tag{2.20}$$

Relations (2.17) and (2.20) show how the first $L_{\mathbf{a}}$ eigenvectors of $\boldsymbol{\Psi}_{\mathbf{a}} = \boldsymbol{\Theta}^H\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}_{\mathbf{b}} = \boldsymbol{\Theta}\boldsymbol{\Theta}^H$ are related. From the above, we also have

$$\boldsymbol{\Theta} = \mathbf{U}'_{\mathbf{b}}\boldsymbol{\Lambda}_{\mathbf{a}}^{1/2}\mathbf{U}_{\mathbf{a}}^H, \tag{2.21}$$

where

$$\begin{aligned}\mathbf{U}'_{\mathbf{b}} &= \left[\, \mathbf{u}_{\mathbf{b},1}\ \mathbf{u}_{\mathbf{b},2}\ \cdots\ \mathbf{u}_{\mathbf{b},L_{\mathbf{a}}} \,\right] \\ &= \boldsymbol{\Theta}\mathbf{U}_{\mathbf{a}}\boldsymbol{\Lambda}_{\mathbf{a}}^{-1/2}\end{aligned} \tag{2.22}$$

is a semi-unitary matrix of size $L_{\mathbf{b}} \times L_{\mathbf{a}}$. In (2.21), we recognize the singular value decomposition (SVD) of $\boldsymbol{\Theta}$. In fact, from a practical point of view, this is all what we need for CCA.

2.2 How CCA Works

Let $\mathbf{g}$ and $\mathbf{h}$ be two complex-valued filters of lengths $L_{\mathbf{a}}$ and $L_{\mathbf{b}}$, respectively. Applying these filters to the two random vectors $\mathbf{a}$ and $\mathbf{b}$, we obtain the two random signals:

$$Z_{\mathbf{a}} = \mathbf{g}^H\mathbf{a}, \tag{2.23}$$

$$Z_{\mathbf{b}} = \mathbf{h}^H\mathbf{b}. \tag{2.24}$$

The Pearson correlation coefficient (PCC) between $Z_{\mathbf{a}}$ and $Z_{\mathbf{b}}$ is then [4]

$$\begin{aligned}\rho(\mathbf{g},\mathbf{h}) &= \frac{E\left(Z_{\mathbf{a}}Z_{\mathbf{b}}^*\right)}{\sqrt{E\left(|Z_{\mathbf{a}}|^2\right)E\left(|Z_{\mathbf{b}}|^2\right)}} \\ &= \frac{\mathbf{g}^H\boldsymbol{\Phi}_{\mathbf{ab}}\mathbf{h}}{\sqrt{\mathbf{g}^H\boldsymbol{\Phi}_{\mathbf{a}}\mathbf{g}\times\mathbf{h}^H\boldsymbol{\Phi}_{\mathbf{b}}\mathbf{h}}},\end{aligned} \tag{2.25}$$

where the superscript * is the complex conjugate. The objective of CCA [1, 5, 6] is to find the two filters $\mathbf{g}$ and $\mathbf{h}$ in such a way that $\rho(\mathbf{g}, \mathbf{h}) + \rho^*(\mathbf{g}, \mathbf{h})$, i.e., the real part of the PCC, is maximized. This is equivalent to maximizing $\mathbf{g}^H \mathbf{\Phi}_{\mathbf{ab}} \mathbf{h} + \mathbf{h}^H \mathbf{\Phi}_{\mathbf{ba}} \mathbf{g}$ subject to the constraints $\mathbf{g}^H \mathbf{\Phi}_{\mathbf{a}} \mathbf{g} = \mathbf{h}^H \mathbf{\Phi}_{\mathbf{b}} \mathbf{h} = 1$, i.e.,

$$\max_{\mathbf{g},\mathbf{h}} \left(\mathbf{g}^H \mathbf{\Phi}_{\mathbf{ab}} \mathbf{h} + \mathbf{h}^H \mathbf{\Phi}_{\mathbf{ba}} \mathbf{g} \right) \quad \text{s. t.} \quad \begin{cases} \mathbf{g}^H \mathbf{\Phi}_{\mathbf{a}} \mathbf{g} = 1 \\ \mathbf{h}^H \mathbf{\Phi}_{\mathbf{b}} \mathbf{h} = 1 \end{cases}. \tag{2.26}$$

The Lagrange function associated with the previous criterion is

$$\mathcal{L}(\mathbf{g}, \mathbf{h}) = \mathbf{g}^H \mathbf{\Phi}_{\mathbf{ab}} \mathbf{h} + \mathbf{h}^H \mathbf{\Phi}_{\mathbf{ba}} \mathbf{g} + \mu_{\mathbf{g}} \left(\mathbf{g}^H \mathbf{\Phi}_{\mathbf{a}} \mathbf{g} - 1 \right) + \mu_{\mathbf{h}} \left(\mathbf{h}^H \mathbf{\Phi}_{\mathbf{b}} \mathbf{h} - 1 \right), \tag{2.27}$$

where $\mu_{\mathbf{g}} \neq 0$ and $\mu_{\mathbf{h}} \neq 0$ are two real-valued Lagrange multipliers. Taking the gradient of $\mathcal{L}(\mathbf{g}, \mathbf{h})$ with respect to $\mathbf{g}$ and $\mathbf{h}$ and equating the results to zero, we get the two equations:

$$\mathbf{\Phi}_{\mathbf{ab}} \mathbf{h} + \mu_{\mathbf{g}} \mathbf{\Phi}_{\mathbf{a}} \mathbf{g} = \mathbf{0}, \tag{2.28}$$

$$\mathbf{\Phi}_{\mathbf{ba}} \mathbf{g} + \mu_{\mathbf{h}} \mathbf{\Phi}_{\mathbf{b}} \mathbf{h} = \mathbf{0}, \tag{2.29}$$

which can be rewritten as

$$\mathbf{g} = -\frac{1}{\mu_{\mathbf{g}}} \mathbf{\Phi}_{\mathbf{a}}^{-1} \mathbf{\Phi}_{\mathbf{ab}} \mathbf{h}, \tag{2.30}$$

$$\mathbf{h} = -\frac{1}{\mu_{\mathbf{h}}} \mathbf{\Phi}_{\mathbf{b}}^{-1} \mathbf{\Phi}_{\mathbf{ba}} \mathbf{g}. \tag{2.31}$$

Left multiplying both sides of (2.28) and (2.29) by $\mathbf{g}^H$ and $\mathbf{h}^H$, respectively, and using the constraints, we easily find that

$$\begin{aligned} \mu_{\mathbf{g}} &= -\frac{\mathbf{g}^H \mathbf{\Phi}_{\mathbf{ab}} \mathbf{h}}{\mathbf{g}^H \mathbf{\Phi}_{\mathbf{a}} \mathbf{g}} \\ &= -\mathbf{g}^H \mathbf{\Phi}_{\mathbf{ab}} \mathbf{h}, \end{aligned} \tag{2.32}$$

$$\begin{aligned} \mu_{\mathbf{h}} &= -\frac{\mathbf{h}^H \mathbf{\Phi}_{\mathbf{ba}} \mathbf{g}}{\mathbf{h}^H \mathbf{\Phi}_{\mathbf{b}} \mathbf{h}} \\ &= -\mathbf{h}^H \mathbf{\Phi}_{\mathbf{ba}} \mathbf{g}. \end{aligned} \tag{2.33}$$

As a result,

$$|\rho(\mathbf{g}, \mathbf{h})|^2 = \mu_{\mathbf{g}} \mu_{\mathbf{h}} \tag{2.34}$$

and $\rho(\mathbf{g}, \mathbf{h})$ must be a real-valued number. From all previous equations, we deduce that

$$\left[\boldsymbol{\Phi}_{\mathbf{a}}^{-1/2}\boldsymbol{\Phi}_{\mathbf{ab}}\boldsymbol{\Phi}_{\mathbf{b}}^{-1}\boldsymbol{\Phi}_{\mathbf{ba}}\boldsymbol{\Phi}_{\mathbf{a}}^{-1/2} - |\rho(\mathbf{g},\mathbf{h})|^2 \mathbf{I}_{L_{\mathrm{a}}}\right]\boldsymbol{\Phi}_{\mathbf{a}}^{1/2}\mathbf{g} = \mathbf{0}, \tag{2.35}$$

$$\left[\boldsymbol{\Phi}_{\mathbf{b}}^{-1/2}\boldsymbol{\Phi}_{\mathbf{ba}}\boldsymbol{\Phi}_{\mathbf{a}}^{-1}\boldsymbol{\Phi}_{\mathbf{ab}}\boldsymbol{\Phi}_{\mathbf{b}}^{-1/2} - |\rho(\mathbf{g},\mathbf{h})|^2 \mathbf{I}_{L_{\mathrm{b}}}\right]\boldsymbol{\Phi}_{\mathbf{b}}^{1/2}\mathbf{h} = \mathbf{0}, \tag{2.36}$$

or, using the notation from the previous section,

$$\left[\boldsymbol{\Psi}_{\mathbf{a}} - |\rho(\mathbf{g},\mathbf{h})|^2 \mathbf{I}_{L_{\mathrm{a}}}\right]\boldsymbol{\Phi}_{\mathbf{a}}^{1/2}\mathbf{g} = \mathbf{0}, \tag{2.37}$$

$$\left[\boldsymbol{\Psi}_{\mathbf{b}} - |\rho(\mathbf{g},\mathbf{h})|^2 \mathbf{I}_{L_{\mathrm{b}}}\right]\boldsymbol{\Phi}_{\mathbf{b}}^{1/2}\mathbf{h} = \mathbf{0}, \tag{2.38}$$

where we recognize the studied eigenvalue problem. From the results of Sect. 2.1, it is clear that the solution to our optimization problem in (2.26) is

$$\mathbf{g}_1 = \boldsymbol{\Phi}_{\mathbf{a}}^{-1/2}\mathbf{u}_{\mathbf{a},1}, \tag{2.39}$$

$$\mathbf{h}_1 = \boldsymbol{\Phi}_{\mathbf{b}}^{-1/2}\mathbf{u}_{\mathbf{b},1}, \tag{2.40}$$

which are the first canonical filters. They lead to the first canonical correlation:

$$\rho(\mathbf{g}_1,\mathbf{h}_1) = \sqrt{\lambda_1} \tag{2.41}$$

and to the first canonical variates:

$$\begin{aligned} Z_{\mathbf{a},1} &= \mathbf{g}_1^H\mathbf{a} \\ &= \mathbf{u}_{\mathbf{a},1}^H\boldsymbol{\Phi}_{\mathbf{a}}^{-1/2}\mathbf{a}, \end{aligned} \tag{2.42}$$

$$\begin{aligned} Z_{\mathbf{b},1} &= \mathbf{h}_1^H\mathbf{b} \\ &= \mathbf{u}_{\mathbf{b},1}^H\boldsymbol{\Phi}_{\mathbf{b}}^{-1/2}\mathbf{b}. \end{aligned} \tag{2.43}$$

The second canonical filters are obtained from the criterion:

$$\max_{\mathbf{g},\mathbf{h}} \left(\mathbf{g}^H\boldsymbol{\Phi}_{\mathbf{ab}}\mathbf{h} + \mathbf{h}^H\boldsymbol{\Phi}_{\mathbf{ba}}\mathbf{g}\right) \quad \text{s. t.} \quad \begin{cases} \mathbf{g}^H\boldsymbol{\Phi}_{\mathbf{a}}\mathbf{g} = 1 \\ \mathbf{h}^H\boldsymbol{\Phi}_{\mathbf{b}}\mathbf{h} = 1 \\ \mathbf{g}^H\boldsymbol{\Phi}_{\mathbf{a}}\mathbf{g}_1 = 0 \\ \mathbf{h}^H\boldsymbol{\Phi}_{\mathbf{b}}\mathbf{h}_1 = 0 \end{cases}. \tag{2.44}$$

It is not hard to see that the optimization of (2.44) gives the second canonical filters:

$$\mathbf{g}_2 = \boldsymbol{\Phi}_{\mathbf{a}}^{-1/2}\mathbf{u}_{\mathbf{a},2}, \tag{2.45}$$

$$\mathbf{h}_2 = \boldsymbol{\Phi}_{\mathbf{b}}^{-1/2}\mathbf{u}_{\mathbf{b},2}, \tag{2.46}$$

the second canonical correlation:

$$\rho(\mathbf{g}_2,\mathbf{h}_2) = \sqrt{\lambda_2}, \tag{2.47}$$

and the second canonical variates:

$$\begin{aligned} Z_{\mathbf{a},2} &= \mathbf{g}_2^H \mathbf{a} \\ &= \mathbf{u}_{\mathbf{a},2}^H \mathbf{\Phi}_{\mathbf{a}}^{-1/2} \mathbf{a}, \end{aligned} \tag{2.48}$$

$$\begin{aligned} Z_{\mathbf{b},2} &= \mathbf{h}_2^H \mathbf{b} \\ &= \mathbf{u}_{\mathbf{b},2}^H \mathbf{\Phi}_{\mathbf{b}}^{-1/2} \mathbf{b}. \end{aligned} \tag{2.49}$$

Obviously, following the same procedure, we can derive all the $L_{\mathbf{a}}$ canonical modes ($l = 1, 2, \ldots, L_{\mathbf{a}}$).

- The lth canonical filters:

$$\mathbf{g}_l = \mathbf{\Phi}_{\mathbf{a}}^{-1/2} \mathbf{u}_{\mathbf{a},l}, \tag{2.50}$$

$$\mathbf{h}_l = \mathbf{\Phi}_{\mathbf{b}}^{-1/2} \mathbf{u}_{\mathbf{b},l}. \tag{2.51}$$

  It is important to notice that instead of $\mathbf{g}_l$ and $\mathbf{h}_l$, we can use the filters $\varsigma_{\mathbf{g}_l} \mathbf{g}_l$ and $\varsigma_{\mathbf{h}_l} \mathbf{h}_l$, where $\varsigma_{\mathbf{g}_l} \neq 0$ and $\varsigma_{\mathbf{h}_l} \neq 0$ are arbitrary real numbers, since the canonical correlations will not be affected.
- The lth canonical correlation:

$$\rho\left(\mathbf{g}_l, \mathbf{h}_l\right) = \sqrt{\lambda_l}. \tag{2.52}$$

- The lth canonical variates:

$$\begin{aligned} Z_{\mathbf{a},l} &= \mathbf{g}_l^H \mathbf{a} \\ &= \mathbf{u}_{\mathbf{a},l}^H \mathbf{\Phi}_{\mathbf{a}}^{-1/2} \mathbf{a}, \end{aligned} \tag{2.53}$$

$$\begin{aligned} Z_{\mathbf{b},l} &= \mathbf{h}_l^H \mathbf{b} \\ &= \mathbf{u}_{\mathbf{b},l}^H \mathbf{\Phi}_{\mathbf{b}}^{-1/2} \mathbf{b}. \end{aligned} \tag{2.54}$$

Considering all the modes of the canonical filters, they can be combined in a matrix form as

$$\begin{aligned} \mathbf{G} &= \left[\mathbf{g}_1 \ \mathbf{g}_2 \ \cdots \ \mathbf{g}_{L_{\mathbf{a}}} \right] \\ &= \mathbf{\Phi}_{\mathbf{a}}^{-1/2} \mathbf{U}_{\mathbf{a}}, \end{aligned} \tag{2.55}$$

$$\begin{aligned} \mathbf{H} &= \left[\mathbf{h}_1 \ \mathbf{h}_2 \ \cdots \ \mathbf{h}_{L_{\mathbf{a}}} \right] \\ &= \mathbf{\Phi}_{\mathbf{b}}^{-1/2} \mathbf{U}_{\mathbf{b}}'. \end{aligned} \tag{2.56}$$

Then, it is easy to check that $\mathbf{G}^H \mathbf{\Phi}_{\mathbf{a}} \mathbf{G} = \mathbf{H}^H \mathbf{\Phi}_{\mathbf{b}} \mathbf{H} = \mathbf{I}_{L_{\mathbf{a}}}$ and $\mathbf{G}^H \mathbf{\Phi}_{\mathbf{ab}} \mathbf{H} = \mathbf{H}^H \mathbf{\Phi}_{\mathbf{ba}} \mathbf{G} = \mathbf{\Lambda}_{\mathbf{a}}^{1/2}$. It can also be verified that $\mathbf{H} \mathbf{\Lambda}_{\mathbf{a}}^{1/2} \mathbf{G}^H = \mathbf{\Phi}_{\mathbf{b}}^{-1} \mathbf{\Phi}_{\mathbf{ba}} \mathbf{\Phi}_{\mathbf{a}}^{-1}$. In the same way, we can combine all the modes of the canonical variates:

$$\begin{aligned}\mathbf{z}_{\mathbf{a}} &= \left[Z_{\mathbf{a},1} \; Z_{\mathbf{a},2} \; \cdots \; Z_{\mathbf{a},L_{\mathbf{a}}} \right] \\ &= \mathbf{G}^H \mathbf{a},\end{aligned} \tag{2.57}$$

$$\begin{aligned}\mathbf{z}_{\mathbf{b}} &= \left[Z_{\mathbf{b},1} \; Z_{\mathbf{b},2} \; \cdots \; Z_{\mathbf{b},L_{\mathbf{a}}} \right] \\ &= \mathbf{H}^H \mathbf{b}.\end{aligned} \tag{2.58}$$

Then, it can be verified that $E\left(\mathbf{z}_{\mathbf{a}}\mathbf{z}_{\mathbf{a}}^H\right) = E\left(\mathbf{z}_{\mathbf{b}}\mathbf{z}_{\mathbf{b}}^H\right) = \mathbf{I}_{L_{\mathbf{a}}}$ and $E\left(\mathbf{z}_{\mathbf{a}}\mathbf{z}_{\mathbf{b}}^H\right) = E\left(\mathbf{z}_{\mathbf{b}}\mathbf{z}_{\mathbf{a}}^H\right) = \mathbf{\Lambda}_{\mathbf{a}}^{1/2}$.

Let us now briefly discuss the two particular cases: $L_{\mathbf{a}} = L_{\mathbf{b}} = 1$ and $L_{\mathbf{a}} = 1, \; L_{\mathbf{b}} > 1$.

When $L_{\mathbf{a}} = L_{\mathbf{b}} = 1$, the two random vectors **a** and **b** become the two random variables A and B. In this case, CCA simplifies to the classical PCC between $Z_A = A$ and $Z_B = B$, i.e.,

$$\rho_{AB} = \frac{E\left(AB^*\right)}{\sqrt{E\left(|A|^2\right) E\left(|B|^2\right)}}. \tag{2.59}$$

In the second case ($L_{\mathbf{a}} = 1, \; L_{\mathbf{b}} > 1$), we only have one canonical mode. We find that the canonical filters are

$$G = 1 \tag{2.60}$$

$$\mathbf{h} = \mathbf{\Phi}_{\mathbf{b}}^{-1/2} \mathbf{u}_{\mathbf{b}}, \tag{2.61}$$

where

$$\mathbf{u}_{\mathbf{b}} = \mathbf{\Phi}_{\mathbf{b}}^{-1/2} \phi_{\mathbf{b}A} \tag{2.62}$$

is the (unnormalized) eigenvector corresponding to the only nonzero eigenvalue:

$$\lambda = \frac{\phi_{\mathbf{b}A}^H \mathbf{\Phi}_{\mathbf{b}}^{-1} \phi_{\mathbf{b}A}}{\phi_A}, \tag{2.63}$$

of the rank-1 matrix $\mathbf{\Theta}\mathbf{\Theta}^H = \mathbf{\Phi}_{\mathbf{b}}^{-1/2} \phi_{\mathbf{b}A} \phi_{\mathbf{b}A}^H \mathbf{\Phi}_{\mathbf{b}}^{-1/2} / \phi_A$, with $\phi_{\mathbf{b}A} = E\left(\mathbf{b}A^*\right)$ and $\phi_A = E\left(|A|^2\right)$. As a result, the canonical correlation is

$$\begin{aligned}\rho\left(G, \mathbf{h}\right) &= \frac{\phi_{\mathbf{b}A}^H \mathbf{h}}{\sqrt{\phi_A \times \mathbf{h}^H \mathbf{\Phi}_{\mathbf{b}} \mathbf{h}}} \\ &= \sqrt{\lambda}\end{aligned} \tag{2.64}$$

and the canonical variates are

$$Z_A = A, \tag{2.65}$$

$$Z_{\mathbf{b}} = \phi_{\mathbf{b}A}^H \boldsymbol{\Phi}_{\mathbf{b}}^{-1} \mathbf{b}. \tag{2.66}$$

2.3 The Singular Case

Now, assume that the covariance matrix $\boldsymbol{\Phi}_{\mathbf{a}}$ is singular but the covariance matrix $\boldsymbol{\Phi}_{\mathbf{b}}$ is still nonsingular.[1] In this case, it is clear that CCA as explained in the previous section cannot work since the existence of the inverse of $\boldsymbol{\Phi}_{\mathbf{a}}$ is required. One way to circumvent this problem is derived next.

Let us assume that rank $(\boldsymbol{\Phi}_{\mathbf{a}})$ = rank $(\boldsymbol{\Phi}_{\mathbf{ab}}) = P < L_{\mathbf{a}}$. We can diagonalize $\boldsymbol{\Phi}_{\mathbf{a}}$ as

$$\mathbf{Q}^H \boldsymbol{\Phi}_{\mathbf{a}} \mathbf{Q} = \boldsymbol{\Lambda}', \tag{2.67}$$

where

$$\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_{L_{\mathbf{a}}} \end{bmatrix} \tag{2.68}$$

is a unitary matrix and

$$\boldsymbol{\Lambda}' = \text{diag}\left(\lambda_1', \lambda_2', \ldots, \lambda_{L_{\mathbf{a}}}'\right) \tag{2.69}$$

is a diagonal matrix. The orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{L_{\mathbf{a}}}$ are the eigenvectors corresponding, respectively, to the eigenvalues $\lambda_1', \lambda_2', \ldots, \lambda_{L_{\mathbf{a}}}'$ of the matrix $\boldsymbol{\Phi}_{\mathbf{a}}$, where $\lambda_1' \geq \lambda_2' \geq \cdots \geq \lambda_P' > 0$ and $\lambda_{P+1}' = \lambda_{P+2}' = \cdots = \lambda_{L_{\mathbf{a}}} = 0$. Let

$$\mathbf{Q} = \begin{bmatrix} \mathbf{T} \ \boldsymbol{\Xi} \end{bmatrix}, \tag{2.70}$$

where the $L_{\mathbf{a}} \times P$ matrix $\mathbf{T}$ contains the eigenvectors corresponding to the nonzero eigenvalues of $\boldsymbol{\Phi}_{\mathbf{a}}$ and the $L_{\mathbf{a}} \times (L_{\mathbf{a}} - P)$ matrix $\boldsymbol{\Xi}$ contains the eigenvectors corresponding to the null eigenvalues of $\boldsymbol{\Phi}_{\mathbf{a}}$. It can be verified that

$$\mathbf{I}_{L_{\mathbf{a}}} = \mathbf{T}\mathbf{T}^H + \boldsymbol{\Xi}\boldsymbol{\Xi}^H. \tag{2.71}$$

Notice that $\mathbf{T}\mathbf{T}^H$ and $\boldsymbol{\Xi}\boldsymbol{\Xi}^H$ are two orthogonal projection matrices of rank P and $L_{\mathbf{a}} - P$, respectively. Hence, $\mathbf{T}\mathbf{T}^H$ is the orthogonal projector onto the signal subspace (where all the energy of the signal is concentrated) or the range of $\boldsymbol{\Phi}_{\mathbf{a}}$, and $\boldsymbol{\Xi}\boldsymbol{\Xi}^H$ is the orthogonal projector onto the null subspace of $\boldsymbol{\Phi}_{\mathbf{a}}$. Using (2.71), we can write the random vector $\mathbf{a}$ (of length $L_{\mathbf{a}}$) as

[1] The same approach discussed in this section can be applied when $\boldsymbol{\Phi}_{\mathbf{b}}$ is singular or when both $\boldsymbol{\Phi}_{\mathbf{a}}$ and $\boldsymbol{\Phi}_{\mathbf{b}}$ are singular.

$$\begin{aligned}\mathbf{a} &= \mathbf{Q}\mathbf{Q}^H\mathbf{a} \\ &= \mathbf{T}\mathbf{T}^H\mathbf{a} \\ &= \mathbf{T}\widetilde{\mathbf{a}},\end{aligned} \tag{2.72}$$

where

$$\widetilde{\mathbf{a}} = \mathbf{T}^H\mathbf{a} \tag{2.73}$$

is the transformed random signal vector of length P. Therefore, instead of working with the pair of random vectors $\mathbf{a}$ and $\mathbf{b}$ as we did in Sect. 2.2, we propose to handle the pair of random vectors $\widetilde{\mathbf{a}}$ and $\mathbf{b}$ since now the covariance matrix of $\widetilde{\mathbf{a}}$ denoted $\mathbf{\Phi}_{\widetilde{\mathbf{a}}} = \mathbf{T}^H\mathbf{\Phi}_{\mathbf{a}}\mathbf{T}$ has full rank. As a result, CCA with P different canonical modes can be performed by simply replacing in previous derivations $\mathbf{\Phi}_{\mathbf{a}}$ by $\mathbf{\Phi}_{\widetilde{\mathbf{a}}}$ and $\mathbf{\Phi}_{\mathbf{ab}}$ by $\mathbf{\Phi}_{\widetilde{\mathbf{a}}\mathbf{b}} = \mathbf{T}^H\mathbf{\Phi}_{\mathbf{ab}}$.

References

1. H. Hotelling, Relations between two sets of variables. Biometrika **28**, 321–377 (1936)
2. D.V. Ouellette, Schur complements and statistics. Linear Algebr. Appl. **36**, 187–295 (1981)
3. G.H. Golub, C.F. Van Loan, *Matrix Computations*, 3rd edn. (The Johns Hopkins University Press, Baltimore, Maryland, 1996)
4. J. Benesty, J. Chen, Y. Huang, On the importance of the Pearson correlation coefficient in noise reduction. IEEE Trans. Audio Speech Lang. Process. **16**, 757–765 (2008)
5. J.R. Kettenring, Canonical analysis of several sets of variables. Biometrika **58**, 433–451 (1971)
6. T.W. Anderson, Asymptotic theory for canonical correlation analysis. J. Multivar. Anal. **79**, 1–29 (1999)

http://www.springer.com/978-3-319-67019-5

Canonical Correlation Analysis in Speech Enhancement
Benesty, J.; Cohen, I.
2018, IX, 121 p. 47 illus. in color., Softcover
ISBN: 978-3-319-67019-5