Supplementary File for Learning Dynamic Coupling of Neural Oscillations for Motor Imagery Classification

Shoulin Huang, *Member, IEEE,* Ruifeng Zhang, Xiaohao Wen, *Member, IEEE,* and Mengchu Zhou, *Fellow, IEEE*

This supplementary file consists of the following parts:

1) **Notations and Operations**
2) **Optimization of CFC-based Tensor Decomposition**

I. Notations and Operations

Tensors are denoted by boldface Euler script letters. For example, $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times ... \times I_D}$ is a D-order (or mode) tensor, where $I_d(d \in \{1, ..., D\})$ is the size of the mode-d. The element (i_1, i_2,..., i_d) of $\boldsymbol{\mathcal{X}}$ is represented by $x_{i_1,i_2,...,i_d}$. Matrices are denoted by boldface capital letters, e.g., $\mathbf{X} \in \mathbb{R}^{K \times L}$. A vector is denoted by boldface lowercase letters, e.g., $\boldsymbol{x} \in \mathbb{R}^{I_1}$. The element (i_1) of $\boldsymbol{x}$ is denoted by x_{i_1}

The mode-d product of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times ... \times I_D}$ with a vector $\mathbf{v} \in \mathbb{R}^{I_d}$ is denoted by $\boldsymbol{\mathcal{X}} \times_d \mathbf{v}$ and is of size $I_1 \times ... \times I_{d-1} \times I_{d+1} \times ... \times I_D$. The element $(i_1, ..., i_{d-1}, i_{d+1}, ..., i_D)$ of $\boldsymbol{\mathcal{X}} \times_d \mathbf{v}$ is given by

$$(\boldsymbol{\mathcal{X}} \times_d \mathbf{v})_{i_1,...,i_{d-1},i_{d+1},...,i_D} = \sum_{i_d=1}^{I_d} x_{i_1,i_2,...,i_D} v_{i_d} \tag{1}$$

II. Optimization of CFC-based Tensor Decomposition (CTD)

This section describes the optimized detail for the proposed CFC-based Tensor Decomposition (CTD) used in CFCNet. To ensure that the extractive features are discriminative, we employ the CSP technique to find the parameters of $\boldsymbol{s}$ and $\mathbf{B}$. To be more specific, the optimization objective of CTD is to maximize the following power ratios for class 1 and class 2:

$$\max_{\boldsymbol{s},\mathbf{B}} \frac{E_{\boldsymbol{\mathcal{X}}\in C_1}[\alpha(\boldsymbol{\mathcal{X}}; \boldsymbol{s}, \mathbf{B})]}{E_{\boldsymbol{\mathcal{X}}\in C_2}[\alpha(\boldsymbol{\mathcal{X}}; \boldsymbol{s}, \mathbf{B})] + E_{\boldsymbol{\mathcal{X}}\in C_1}[\alpha(\boldsymbol{\mathcal{X}}; \boldsymbol{s}, \mathbf{B})]} \tag{2}$$

Here, C_y is the set in training data, which contains the test belonging to class $y \in \{1, 2\}$; $\alpha(\boldsymbol{\mathcal{X}}; \boldsymbol{s}, \mathbf{B}) = \boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}$ denotes the power calculated from the projection signal $\boldsymbol{z}$, i.e., $\alpha(\boldsymbol{\mathcal{X}}; \boldsymbol{s}, \mathbf{B}) = \boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}$; and $E_{\boldsymbol{\mathcal{X}}\in C_y}[\,\cdot\,]$ denotes the expected covariance matrix over C_y. Since it is difficult to determine $\boldsymbol{s}$ and $\mathbf{B}$ simultaneously, a standard alternating optimization strategy is adopted in the optimization process.

While $\mathbf{B}$ is fixed, $\alpha(\boldsymbol{\mathcal{X}}; \boldsymbol{s}, \mathbf{B})$ can be denoted by $\boldsymbol{s}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\boldsymbol{s}$, where $\mathbf{G} = \boldsymbol{\mathcal{X}} \times_2 [\![\mathbf{B}^{\mathrm{T}}]\!] \in \mathbb{R}^{K \times L}$. Let $cov(\mathbf{A}, y)$ be the function that computes the average covariance matrix for class y. Then it follows that $E_{\boldsymbol{\mathcal{X}}\in C_1}[\alpha(\boldsymbol{\mathcal{X}}; \boldsymbol{s}, \mathbf{B})] = \boldsymbol{s}^{\mathrm{T}}cov(\mathbf{G}, 1)\boldsymbol{s}$ and $E_{\boldsymbol{\mathcal{X}}\in C_2}[\alpha(\boldsymbol{\mathcal{X}}; \boldsymbol{s}, \mathbf{B})] = \boldsymbol{s}^{\mathrm{T}}cov(\mathbf{G}, 2)\boldsymbol{s}$. Thus, (2) can be re-expressed as a Rayleigh quotient maximization problem.

$$\max_{\boldsymbol{s}_m} J(\boldsymbol{s}_m|\mathbf{B}_m) = \frac{\boldsymbol{s}_m^{\mathrm{T}} cov(\mathbf{G}_m, d)\boldsymbol{s}_m}{\boldsymbol{s}_m^{\mathrm{T}}[cov(\mathbf{G}_m, 1) + cov(\mathbf{G}_m, 2)]\boldsymbol{s}_m} \tag{3}$$

Problem (3) can be solved by generalized eigenvalue decomposition:

$$cov(\mathbf{G}_m, 1)\boldsymbol{s}_m = \lambda[cov(\mathbf{G}_m, 1) + cov(\mathbf{G}_m, 2)]\boldsymbol{s}_m \tag{4}$$

The spatial factor $\boldsymbol{s}$ is given by the eigenvector corresponding to the eigenvalue λ. The largest or smallest eigenvalues mean discriminability of the eigenvectors. We set M_s as an even number $2r$. And then half of the factors, $\{\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_r\}$, are chosen as the eigenvectors corresponding to the largest r eigenvalues for class 1, and the other factors, $\{\boldsymbol{s}_{r+1}, \boldsymbol{s}_{r+2}, ..., \boldsymbol{s}_{2r}\}$ are the eigenvectors corresponding to the smallest r eigenvalues for class 2.

Next, we find $\{\mathbf{B}_m,\ m = 1, 2, ..., M\}$ assuming that $\{\boldsymbol{s}_m, m = 1, 2, ..., M\}$ is given. Let $\boldsymbol{\beta}_m = vec(\mathbf{B}_m) \in \mathbb{R}^{KN}$, where $vec(\mathbf{B}_m)$ denotes the vectorization operation that sequentially concatenates the column vectors of $\mathbf{B}_m$ to form a new vector. Then it follows that $\alpha(\boldsymbol{\mathcal{X}}; \boldsymbol{s}_m, \mathbf{B}_m) = \boldsymbol{\beta}_m^{\mathrm{T}}\mathbf{Q}_m\mathbf{Q}_m^{\mathrm{T}}\boldsymbol{\beta}_m$, where the matrix $\mathbf{Q}_m = [s_{m,1}\mathbf{X}_1^{\mathrm{T}}\ \ s_{m,2}\mathbf{X}_2^{\mathrm{T}}\ \ ...\ \ s_{m,K}\mathbf{X}_K^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{(KN)\times L}$, with $s_{m,k}$ being the k-th element of $\boldsymbol{s}_m$. Similarly, the optimization problem can be written as

$$\max_{\mathbf{B}_m} J(\mathbf{B}_m|\boldsymbol{s}_m) = \frac{\boldsymbol{\beta}_m^{\mathrm{T}} cov(\mathbf{Q}_m, 1)\boldsymbol{\beta}_m}{\boldsymbol{\beta}_m^{\mathrm{T}}[cov(\mathbf{Q}_m, 1) + cov(\mathbf{Q}_m, 2)]\boldsymbol{\beta}_m} \tag{5}$$

Algorithm 1 Optimization details for CTD

Input: The training set $\{\mathbf{X}_{i\text{-th}} \in \mathbb{R}^{K \times L}, i = 1, 2, ..., I\}$ with their labels $\mathbf{y} \in \mathbb{R}^{I}$, the maximum number of iterations T and a small value ε for testing convergence, and the number of spatial-spectral factors $2r$.
Output: $\{(\boldsymbol{s}_m, \mathbf{B}_m), m = 1, 2, ..., M\}$ (note that $M = 2r$).
calculate $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{K \times N \times L}$ for $1 \leq i \leq I$;
calculate average covariance: $cov(\mathbf{R}, 1)$ and $cov(\mathbf{R}, 2)$;
calculate $\{\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_{2r}\}$ according to (8);
initialize $\{\mathbf{B}_m, m = 1, 2, ..., 2r\}$ according to (9);
for $1 \leq t \leq T$ **do**
  calculate $cov(\mathbf{G}_m, 1)$ and $cov(\mathbf{G}_m, 2)$;
  obtain $\{\boldsymbol{s}_m^{(t)}, m = 1, 2, ..., 2r\}$ by solving (3);
  calculate $cov(\mathbf{Q}_m, 1)$ and $cov(\mathbf{Q}_m, 2)$;
  obtain $\{\boldsymbol{b}_{m,k}^{(t)}, k = 1, 2, ..., K,\ m = 1, 2, ..., 2r\}$ by solving (5) and (6);
  calculate $\{\mathbf{B}_m^{(t)}, m = 1, 2, ..., 2r\}$ according to (9);
  if $t = T$ or $J(\boldsymbol{s}_m^{(t)} | \mathbf{B}_m^{(t-1)}) - J(\mathbf{B}_m^{(t)} | \boldsymbol{s}_m^{(t)}) \leq \varepsilon$ for all m
    break;
end for
return $\{(\boldsymbol{s}_m^{(t)}, \mathbf{B}_m^{(t)}), m = 1, 2, ..., M\}$

In addition, problem (5) is solved by generalized eigenvalue decomposition. $\boldsymbol{\beta}_m$ is chosen as the eigenvector corresponding to the largest eigenvalue for m within the range of 1 to r, and the eigenvector corresponding to the smallest eigenvalue for m within the range of $r + 1$ to $2r$. The k-th column of $\mathbf{B}_m$, $\boldsymbol{b}_{m,k}$, can be restored from $\boldsymbol{\beta}_m$ by

$$\begin{aligned}\boldsymbol{b}_{m,k} = [\beta_{m,p}\ \beta_{m,p+1}\ ...\ \beta_{m,p+N-1}]^{\mathrm{T}}, \\ k = 1, 2, ..., K,\ m = 1, 2, ..., 2r\end{aligned} \tag{6}$$

where $p = (k-1)N + 1$, and $\beta_{m,p}$ is the p-th element of $\boldsymbol{\beta}_m$.

To facilitate fast convergence, we initialize $\{\mathbf{B}_m, m = 1, 2, ..., M\}$ with effective initial values. Let us define a coupling factor $\boldsymbol{w} \in \mathbb{R}^{KN}$ as

$$\boldsymbol{w} = \boldsymbol{s} \odot vec(\mathbf{B}) = [s_1 \boldsymbol{b}_1^{\mathrm{T}}\ s_2 \boldsymbol{b}_2^{\mathrm{T}}\ ...\ s_K \boldsymbol{b}_K^{\mathrm{T}}]^{\mathrm{T}} \tag{7}$$

where $\odot$ denotes the Khatri–Rao product of two vectors. It is easy to find $\alpha(\boldsymbol{\mathcal{X}}; \boldsymbol{s}, \mathbf{B}) = \boldsymbol{w}^{\mathrm{T}} \mathbf{R} \mathbf{R}^{\mathrm{T}} \boldsymbol{w}$, where $\mathbf{R} = [\mathbf{X}_1;\ \mathbf{X}_2; ...;\ \mathbf{X}_K] \in \mathbb{R}^{(KN) \times L}$. Thus, (2) can be re-expressed as

$$\max_{\boldsymbol{w}} J(\boldsymbol{w}) = \frac{\boldsymbol{w}^{\mathrm{T}} cov(\mathbf{R}, 1) \boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}} [cov(\mathbf{R}, 1) + cov(\mathbf{R}, 2)] \boldsymbol{w}} \tag{8}$$

Problem (8) also can be solved by generalized eigenvalue decomposition: $\gamma[cov(\mathbf{R}, 1) + cov(\mathbf{R}, 2)]\boldsymbol{w}$. Half of the factors, $\{\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_r\}$, are chosen as eigenvectors corresponding to the largest r eigenvalues for class 1, and the others are eigenvectors corresponding to the smallest r eigenvalues for class 2. The initial values of $\mathbf{B}_m$ can be estimated from $\boldsymbol{w}_m$, as follows:

$$\begin{aligned}\boldsymbol{b}_{m,k} = [w_{m,p}\ \ w_{m,p+1}\ ...\ w_{m,p+N-1}]^{\mathrm{T}} / h_p, \\ k = 1, 2, ..., K,\ m = 1, 2, ..., 2r\end{aligned} \tag{9}$$

where $p = (k-1)N + 1$; $w_{m,p}$ is the p-th element of $\boldsymbol{w}_m$; and $h_p = \sqrt{(w_{m,p})^2 + (w_{m,p+1})^2 + ... + (w_{m,p+N-1})^2}$. In this way, $\boldsymbol{b}_{m,k}$ is assumed to be unit l_2-norm. The pseudo-code of the CTD finding $\{(\boldsymbol{s}_m, \mathbf{B}_m), m = 1, 2, ..., M\}$ is given in **Algorithm 1**. Similarly to CSP, CTD can easily be extended to multi-class applications, by adopting strategies, such as Pair-Wise and One-Versus-Rest.

The proposed CTD method finds spatial-spectral factors alternatively by solving a single Rayleigh quotient maximization objective function, where we never decrease the value of cost function between two successive iterations. The convergence of the alternating optimization process is proved in $Theorem$ 1.

$Theorem$ 1: The alternating optimization process for CTD guarantees monotonic convergence:

$$\begin{aligned}\sigma \leq J(\boldsymbol{s}_m^{(1)} | \mathbf{B}_m^{(0)}) \leq J(\mathbf{B}_m^{(1)} | \boldsymbol{s}_m^{(1)}) \leq J(\boldsymbol{s}_m^{(2)} | \mathbf{B}_m^{(1)}) \leq ... \\ \leq J(\boldsymbol{s}_m^{(t)} | \mathbf{B}_m^{(t-1)}) \leq J(\mathbf{B}_m^{(t)} | \boldsymbol{s}_m^{(t)}) < 1\end{aligned} \tag{10}$$

where σ is a limiting value in the $\mathbb{R}^{+}$ space $(0 < \sigma < 1)$, and $J(\mathbf{B}_m^{(t)} | \boldsymbol{s}_m^{(t)})$ is calculated with the given $\boldsymbol{s}_m^{(t)}$ in the t-th $(t \geq 1)$

iteration. $\mathbf{B}_m^{(0)}$ is initialized according to (9).

$Proof$: Let $\sigma_m = J(\boldsymbol{w}_m)$ for $m = 1, 2, ..., 2r$, and it is easy to know that $0 < \sigma_m < 1$. Since $\boldsymbol{s}_m^{(1)}$ is calculated by arg maximizing $J(\boldsymbol{s}_m|\mathbf{B}_m)$ with the given $\mathbf{B}_m^{(0)}$, we have the following inequality: $\sigma \leq J(\boldsymbol{s}_m^{(1)}|\mathbf{B}_m^{(0)})$, where $\sigma = \min(\sigma_1, \sigma_2, ..., \sigma_{2r})$.

According to (5), we have $J(\mathbf{B}_m^{(t)}|\boldsymbol{s}_m^{(t)}) < 1$. In addition, because $\boldsymbol{s}_m^{(t)}$ maximizes $J(\boldsymbol{s}_m^{(t)}|\mathbf{B}_m^{(t-1)})$ with fixed $\mathbf{B}_m^{(t-1)}$ and then $\mathbf{B}_m^{(t)}$ maximizes $J(\mathbf{B}_m^{(t)}|\boldsymbol{s}_m^{(t)})$ with calculated $\boldsymbol{s}_m^{(t)}$, the relationship $J(\mathbf{B}_m^{(t-1)}|\boldsymbol{s}_m^{(t-1)}) \leq J(\boldsymbol{s}_m^{(t)}|\mathbf{B}_m^{(t-1)}) \leq J(\mathbf{B}_m^{(t)}|\boldsymbol{s}_m^{(t)})$ is given. Hence, we finally obtain the following relationship: $\sigma \leq J(\boldsymbol{s}_m^{(1)}|\mathbf{B}_m^{(0)}) \leq J(\mathbf{B}_m^{(1)}|\boldsymbol{s}_m^{(1)}) \leq ... \leq J(\boldsymbol{s}_m^{(t)}|\mathbf{B}_m^{(t-1)}) \leq J(\mathbf{B}_m^{(t)}|\boldsymbol{s}_m^{(t)}) < 1$.