Supplementary Material for the CMTF Algorithm used in the paper "Tensor Analysis and Fusion of Multimodal Brain Images"

Esin Karahan, Pedro A. Rojas-Lopez, Maria L. Bringas-Vega, Pedro A. Valdes-Hernandez, Pedro A. Valdes-Sosa

This document is prepared for the explanation of the CMTF software presented in http://www.cneuro.cu/software/tensor and http://neurosignal.boun.edu.tr/software/tensor .

I. NOTATION

$\mathbf{X}^T$ and $\mathbf{X}^H$ are the transpose and conjugate (Hermitian) transpose of the matrix $\mathbf{X}$, respectively. $\mathbf{X}^\dagger$ is the Moore-Penrose pseudoinverse of $\mathbf{X}$. $\mathbf{I}_K$ is $K \times K$ dimensional identity matrix and $\mathbf{1}_K$ is $K \times 1$ dimensional column vector of 1's. *Chol* denotes Cholesky decomposition.

Definition: Mode-*n* unfolding of a tensor is the transformation of the tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ to a matrix $\mathcal{X}_{(n)} \in \mathbb{R}^{I_n \times I_1 \ldots I_{n-1} I_{n+1} \ldots I_N}$ where mode-*n* fibers are arranged to be columns of the resulting matrix. Tensor element $(i_1, i_2, \ldots, i_N)$ corresponds to matrix element (i_n, j), where $j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^{N} (i_k - 1) J_k$ with $J_k = \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m$.

Definition: Khatri-Rao product is the columnwise Krocker product . Let $\mathbf{A} \in \mathbb{R}^{I \times K}$ and $\mathbf{B} \in \mathbb{R}^{J \times K}$, Khatri-Rao product $\mathbf{A} \odot \mathbf{B}$ defined as follows:

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{A}(:,1) \otimes \mathbf{B}(:,1) \;\; \mathbf{A}(:,2) \otimes \mathbf{B}(:,2) \cdots \mathbf{A}(:,K) \otimes \mathbf{B}(:,K)]$$

Definition: PARAFAC decomposition

Given $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, PARAFAC is formulated as:

$$\mathcal{X} = [\![\lambda; \mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_N]\!] \tag{1}$$

where $\mathbf{U}_n \in \mathbb{R}^{I_n \times R}$ for $n = 1, \ldots, N$ are the factor matrices or signatures and R is the model order.

PARAFAC decomposition can be expressed in different ways (Kolda, 2006):

Expression 1: Elementwise notation

$$\mathcal{X}(i_1, \ldots, i_N) = \sum_{r=1}^{R} \mathbf{U}_1(\mathrm{i}_1, \mathrm{r}) \ldots \mathbf{U}_N(\mathrm{i}_N, \mathrm{r})$$

Expression 2: Sum of outer products

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{U}_1(:,r) \circ \ldots \circ \mathbf{U}_N(:,r)$$

Expression 3: Matricized notation

$$\mathcal{X}_{(n)} = \mathbf{U}_n (\mathbf{U}_N \odot \ldots . \odot \mathbf{U}_{n+1} \odot \mathbf{U}_{n-1} \odot \ldots \odot \mathbf{U}_1)^T$$

Factors of the PARAFAC decomposition $\mathbf{U}_1$ to $\mathbf{U}_N$ can be estimated by using Alternating Least Squares (ALS) algorithm. ALS is an iterative algorithm such that each factor is estimated sequentially by fixing the others. $\mathbf{U}_n$ is estimated by:

$$\hat{\mathbf{U}}_n = \arg\min_{\mathbf{U}_n} \left\{ \left\| \boldsymbol{\mathcal{X}} - [\![\mathbf{U}_1, \ldots, \mathbf{U}_n, \ldots, \mathbf{U}_N]\!] \right\|_2^2 \right\}$$

By using expression 3, this is equivalent to:

$$\hat{\mathbf{U}}_n = \arg\min_{\mathbf{U}_n} \left\{ \left\| \boldsymbol{\mathcal{X}}_{(\mathrm{n})} - \mathbf{U}_n (\mathbf{U}_N \odot \ldots \odot \mathbf{U}_{n+1} \odot \mathbf{U}_{n-1} \odot \ldots \odot \mathbf{U}_1)^T \right\|_2^2 \right\}$$

The closed form solution of $\hat{\mathbf{U}}_n$ is found as:

$$\hat{\mathbf{U}}_n = \boldsymbol{\mathcal{X}}_{(\mathrm{n})} (\mathbf{U}_N \odot \ldots \odot \mathbf{U}_{n+1} \odot \mathbf{U}_{n-1} \odot \ldots \odot \mathbf{U}_1)^{\dagger}$$

II. Penalized CMTF

The time–varying EEG spectrum is organized as tensor $\boldsymbol{\mathcal{S}}_T \in \mathbb{R}^{I_E \times I_{T\delta} \times I_{F\delta}}$ where I_E is the number of electrodes, $I_{T\delta}$ is the number of time samples, $I_{F\delta}$ is the number of frequency samples. fMRI data matrix $\mathbf{B} \in \mathbb{R}^{I_{Cx} \times I_{T\delta}}$ is formed as a matrix over I_{Cx} voxels on the cortical grid and $I_{T\delta}$ time samples. Lead field $\mathbf{K} \in \mathbb{R}^{I_E \times I_{Cx}}$ is included into the model to project cortical source density on the sensor space.

The EEG tensor $\boldsymbol{\mathcal{S}}_T$ is decomposed into source spatial $\mathbf{M}_{eeg}$, temporal $\mathbf{T_V}$ and spectral $\mathbf{F_V}$ signatures, and the fMRI data matrix $\mathbf{B}$ is decomposed into spatial $\mathbf{M}_{fmri}$ and temporal $\mathbf{T_B}$ signatures. EEG and fMRI data tensors are coupled in the spatial dimension and the spatial signatures are divided into common, $\mathbf{M_C}$ and discriminative sigantures, $\mathbf{M_G}$ and $\mathbf{M_B}$. Then, $\mathbf{M}_{eeg} = [\mathbf{M_C}, \mathbf{M_G}]$ and $\mathbf{M}_{fmri} = [\mathbf{M_C}, \mathbf{M_B}]$

Modality specific and common signatures are estimated by:

$$(\hat{\mathbf{M}}_\mathbf{C}, \hat{\mathbf{M}}_\mathbf{G}, \hat{\mathbf{T}}_\mathbf{V}, \hat{\mathbf{F}}_\mathbf{V}, \hat{\mathbf{M}}_\mathbf{B}, \hat{\mathbf{T}}_\mathbf{B}) = \arg\min_{\substack{\mathbf{M_C}, \mathbf{M_G}, \mathbf{T_V}, \mathbf{F_V}, \\ \mathbf{M_B}, \mathbf{T_B}}} \left\{ \begin{array}{l} \frac{1}{2} \left\| \boldsymbol{\mathcal{S}}_T - [\![\mathbf{K}[\mathbf{M_C}, \mathbf{M_G}], \mathbf{T_V}, \mathbf{F_V}]\!] \right\|_2^2 + \gamma \frac{1}{2} \left\| \mathbf{B} - [\![[\mathbf{M_C}, \mathbf{M_B}], \mathbf{T_B}]\!] \right\|_2^2 \\ + \lambda_1 \left\| \mathbf{M_C} \right\|_1 + \frac{1}{2} \lambda_2 \left\| \mathbf{L}\,\mathbf{M_C} \right\|^2 + \lambda_3 \left\| \mathbf{M_G} \right\|_1 + \frac{1}{2} \lambda_4 \left\| \mathbf{L}\,\mathbf{M_G} \right\|^2 \\ + \lambda_5 \left\| \mathbf{M_B} \right\|_1 + \frac{1}{2} \lambda_6 \left\| \mathbf{L}\,\mathbf{M_B} \right\|^2 \end{array} \right\}$$

$$\text{s.t.} [\mathbf{M_C}, \mathbf{M_G}]^T [\mathbf{M_C}, \mathbf{M_G}] = \mathbf{I}, [\mathbf{M_C}, \mathbf{M_B}]^T [\mathbf{M_C}, \mathbf{M_B}] = \mathbf{I}, \quad (2)$$
$$\mathbf{M_C} \geq 0,\ \mathbf{M_G} \geq 0, \mathbf{M_B} \geq 0, \mathbf{F_V} \geq 0$$

Estimation of Spatial Signatures:

Common spatial signature $\mathbf{M_C}$, individual spatial signature of EEG $\mathbf{M_G}$ and individual spatial signature of fMRI $\mathbf{M_B}$ are estimated by matricizing the (2) as follows:

$$(\hat{\mathbf{M}}_{\mathbf{C}},\hat{\mathbf{M}}_{\mathbf{G}},\hat{\mathbf{M}}_{\mathbf{B}}) = \underset{\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{G}},\mathbf{M}_{\mathbf{B}}}{\arg\min} \left\{ \begin{array}{l} \frac{1}{2}\left\| \boldsymbol{\mathcal{S}}_{T(1)} - \mathbf{K}[\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{G}}](\mathbf{F}_{\mathbf{V}} \odot \mathbf{T}_{\mathbf{V}})^T \right\|_2^2 + \gamma \frac{1}{2}\left\| \mathbf{B} - [\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{B}}]\mathbf{T}_{\mathbf{B}}^T \right\|_2^2 \\ +\lambda_1 \|\mathbf{M}_{\mathbf{C}}\|_1 + \frac{1}{2}\lambda_2 \|\mathbf{L}\mathbf{M}_{\mathbf{C}}\|^2 + \lambda_3 \|\mathbf{M}_{\mathbf{G}}\|_1 + \frac{1}{2}\lambda_4 \|\mathbf{L}\mathbf{M}_{\mathbf{G}}\|^2 \\ +\lambda_5 \|\mathbf{M}_{\mathbf{B}}\|_1 + \frac{1}{2}\lambda_6 \|\mathbf{L}\mathbf{M}_{\mathbf{B}}\|^2 \end{array} \right\}$$

$$\text{s.t.}\,[\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{G}}]^T[\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{G}}] = \mathbf{I},[\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{B}}]^T[\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{B}}] = \mathbf{I},$$
$$\mathbf{M}_{\mathbf{C}} \geq 0,\ \mathbf{M}_{\mathbf{G}} \geq 0, \mathbf{M}_{\mathbf{B}} \geq 0$$

Assume that the number of common components is R_C, the number of discriminative components of EEG is R_{DV} and the number of discriminative components of fMRI is R_{DB}. Then the model order of the PARAFAC for EEG tensor is $R_V = R_C + R_{DV}$ and for fMRI is $R_B = R_C + R_{DB}$.

We use Hierarchical Alternating Least Squares (HALS) algorithm to estimate the spatial factors of the PARAFAC decomposition. The HALS algorithm is a modified ALS algorithm in which at each step of ALS, only one column of a factor is estimated by fixing the rest of the columns of the factor.

HALS algorithm fits very well into the coupled factorization since the spatial signature matrix is divided into common and discriminative atoms in a columnwise manner. Call $\mathbf{P} = (\mathbf{F}_{\mathbf{V}} \odot \mathbf{T}_{\mathbf{V}})$ and similarly represent $\mathbf{P}$ in two subspaces as follows $\mathbf{P} = [\mathbf{P}_{\mathbf{C}},\mathbf{P}_{\mathbf{G}}]$. It is clear that $\mathbf{P}_{\mathbf{C}} = (\mathbf{F}_{\mathbf{V}}(:,1:R_C) \odot \mathbf{T}_{\mathbf{V}}(:,1:R_C))$ and $\mathbf{P}_{\mathbf{G}} = (\mathbf{F}_{\mathbf{V}}(:,R_C+1:R_V) \odot \mathbf{T}_{\mathbf{V}}(:,R_C+1:R_V))$. And we do the same formulation for fMRI: $\mathbf{Q} = [\mathbf{Q}_{\mathbf{C}},\mathbf{Q}_{\mathbf{B}}]$ where $\mathbf{Q}_{\mathbf{C}} = \mathbf{T}_{\mathbf{B}}(:,1:R_C)$ and $\mathbf{Q}_{\mathbf{B}} = \mathbf{T}_{\mathbf{B}}(:,R_C+1:R_B)$.

Orthogonality Constraint on the Nonnegative Spatial Signatures

Orthogonality constraint on the nonnegative spatial signatures can be imposed column-wise (Kimura, Tanaka, & Kudo, 2014). The reason for this is that for a nonnegative matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$, orthogonality condition $\mathbf{X}^T\mathbf{X} = \mathbf{I}_J$ can be replaced by $2J$ columnwise coefficients $\mathbf{X}(:,j)^T\mathbf{X}(:,j) = 1$ and $\sum_{k \neq j}^{J} \mathbf{X}(:,k)^T\mathbf{X}(:,j) = 0$ for $j = 1,2,\ldots,J$.

For our case, the orthogonality condition is expressed as follows:

$$[\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{G}}]^T[\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{G}}] = \mathbf{I}_{RV} \quad \Rightarrow \quad \begin{cases} \mathbf{M}_{\mathbf{C}}(:,j)^T\mathbf{M}_{\mathbf{C}}(:,j) = 1, & j = 1,2,\ldots,\mathrm{R}_C \quad \wedge \\ \mathbf{M}_{\mathbf{G}}(:,j)^T\mathbf{M}_{\mathbf{G}}(:,j) = 1, & j = 1,2,\ldots,\mathrm{R}_{DV} \quad \wedge \\ \sum_{k \neq j}^{R_C} \mathbf{M}_{\mathbf{C}}(:,k)^T\mathbf{M}_{\mathbf{C}}(:,j) = 0, \ j = 1,2,\ldots,\mathrm{R}_C & \wedge \\ \sum_{k \neq j}^{R_{DV}} \mathbf{M}_{\mathbf{G}}(:,k)^T\mathbf{M}_{\mathbf{G}}(:,j) = 0, \ j = 1,2,\ldots,\mathrm{R}_{DV} & \wedge \\ \sum_{k=1}^{R_{DV}} \mathbf{M}_{\mathbf{G}}(:,k)^T\mathbf{M}_{\mathbf{C}}(:,j) = 0, j = 1,2,\ldots,\mathrm{R}_C & \wedge \\ \sum_{k=1}^{R_C} \mathbf{M}_{\mathbf{C}}(:,k)^T\mathbf{M}_{\mathbf{G}}(:,j) = 0, j = 1,2,\ldots,\mathrm{R}_{DV}. \end{cases} \tag{3}$$

$$[\mathbf{M_C},\mathbf{M_B}]^T[\mathbf{M_C},\mathbf{M_B}]=\mathbf{I}_{RB} \quad \Rightarrow \quad \begin{cases} \mathbf{M_C}(:,j)^T\mathbf{M_C}(:,j)=1, \quad j=1,2,\ldots,\mathrm{R}_C \quad \wedge \\ \mathbf{M_B}(:,j)^T\mathbf{M_B}(:,j)=1, \quad j=1,2,\ldots,\mathrm{R}_{DB} \quad \wedge \\ \sum_{k\neq j}^{R_C}\mathbf{M_C}(:,k)^T\mathbf{M_C}(:,j)=0,\ j=1,2,\ldots,\mathrm{R}_C \quad \wedge \\ \sum_{k\neq j}^{R_{DB}}\mathbf{M_B}(:,k)^T\mathbf{M_B}(:,j)=0,\ j=1,2,\ldots,\mathrm{R}_{DB} \quad \wedge \\ \sum_{k=1}^{R_{DB}}\mathbf{M_B}(:,k)^T\mathbf{M_C}(:,j)=0,\ j=1,2,\ldots,\mathrm{R}_C \quad \wedge \\ \sum_{k=1}^{R_C}\mathbf{M_C}(:,k)^T\mathbf{M_B}(:,j)=0,\ j=1,2,\ldots,\mathrm{R}_{DB}. \end{cases} \tag{4}$$

(3) and (4) are unified for $\mathbf{M_C}$:

$$\mathbf{W}^{(j)}=\sum_{k\neq j}^{R_C}\mathbf{M_C}(:,k)+\sum_{k=1}^{R_{DV}}\mathbf{M_G}(:,k)+\sum_{k=1}^{R_{DB}}\mathbf{M_B}(:,k) \tag{5}$$

And the orthogonality constraint is written as: $\mathbf{W}^{(j)T}\mathbf{M_C}(:,j)=0, \quad j=1,2,\ldots,\mathrm{R}_C$

First, we will present the estimation of the common spatial signature $\mathbf{M_C}$. Estimation of the others will follow. The objective function for the estimation of the *j*th column of $\mathbf{M_C}$ with the orthogonality constraint can be formulated as Lagrangian as follows:

$$\mathcal{L}(\mathbf{M_C},\beta_1(j))=\begin{Bmatrix} \frac{1}{2}\left\|\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}-\mathbf{K}\,\mathbf{M_C}(:,j)\mathbf{P_C}(:,j)^T\right\|_2^2+\gamma\frac{1}{2}\left\|\tilde{\mathbf{B}}-\mathbf{M_C}(:,j)\mathbf{Q_C}(:,j)^T\right\|_2^2 \\ +\lambda_1\left\|\mathbf{M_C}(:,j)\right\|_1+\frac{1}{2}\lambda_2\left\|\mathbf{L}\,\mathbf{M_C}(:,j)\right\|^2+\beta_1(j)\mathbf{W}^{(j)}\mathbf{M_C}(:,j) \end{Bmatrix}$$

where $\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}=\boldsymbol{\mathcal{S}}_{T(1)}-\mathbf{K}\sum_{k\neq j}^{R_C}\mathbf{M_C}(:,k)\mathbf{P_C}(:,k)^T-\mathbf{K}\,\mathbf{M_G}\mathbf{P_G}^T$ and $\tilde{\mathbf{B}}=\mathbf{B}-\sum_{k\neq j}^{R_C}\mathbf{M_C}(:,k)\mathbf{Q_C}(:,k)^T-\mathbf{M_B}\mathbf{Q_B}^T$.

$\beta_1(j)$ is the weighting parameter for the orthogonality constraint on the *j*th column of $\mathbf{M_C}$.

Gradient of the objective function is found as follows:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{M_C}(:,j)}=\begin{Bmatrix} -\mathbf{K}^T\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P_C}(:,j)+\mathbf{K}^T\mathbf{K}\mathbf{M_C}(:,j)\,\mathbf{P_C}(:,j)^T\mathbf{P_C}(:,j)-\gamma\tilde{\mathbf{B}}\mathbf{Q_C}(:,j)+\gamma\mathbf{M_C}(:,j)\mathbf{Q_C}(:,j)^T\mathbf{Q_C}(:,j) \\ +\lambda_1\mathbf{M_C}(:,j)+\lambda_2\mathbf{L}^T\mathbf{L}\,\mathbf{M_C}(:,j)+\beta_1(j)\mathbf{W}^{(j)} \end{Bmatrix}$$

(6)

Since factors are normalized $\mathbf{P_C}(:,j)^T\mathbf{P_C}(:,j)=1$ and $\mathbf{Q_C}(:,j)^T\mathbf{Q_C}(:,j)=1$.

Then $\mathbf{M_C}(:,j)$ is estimated as follows:

$$\hat{\mathbf{M}}_{\mathbf{C}}(:,j)=\left[(\mathbf{K}^T\mathbf{K}+\lambda_2\mathbf{L}^T\mathbf{L}+\gamma\mathbf{I}_{I_{Cx}})^{-1}(\mathbf{K}^T\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P_C}(:,j)+\gamma\tilde{\mathbf{B}}\mathbf{Q_C}(:,j)-\lambda_1\mathbf{1}_{I_{Cx}}-\beta_1(j)\mathbf{W}^{(j)})\right]_+ \tag{7}$$

Nonnegativity is imposed on the factor by thresholding the elements below than a certain value shown by the function $[]_+$.

We set the regularization parameter for orthogonality constraint as described in (Kimura et al., 2014). Multiplication of (6) by $\mathbf{W}^{(j)T}(\mathbf{K}^T\mathbf{K}+\lambda_2\mathbf{L}^T\mathbf{L}+\gamma\mathbf{I}_{I_{Cx}})^{-1}$ from the left and noting $\mathbf{W}^{(j)T}\mathbf{M}_\mathbf{C}(:,j)=0$, the regularization parameter $\beta(j)$ is found as follows:

$$\beta_1(j)=\frac{\mathbf{W}^{(j)T}(\mathbf{K}^T\mathbf{K}+\lambda_2\mathbf{L}^T\mathbf{L}+\gamma\mathbf{I}_{I_{Cx}})^{-1}(\mathbf{K}^T\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P}_\mathbf{C}(:,j)+\gamma\tilde{\mathbf{B}}\mathbf{Q}_\mathbf{C}(:,j)-\lambda_1\mathbf{1}_{I_{Cx}})}{\mathbf{W}^{(j)T}(\mathbf{K}^T\mathbf{K}+\lambda_2\mathbf{L}^T\mathbf{L}+\gamma\mathbf{I}_{I_{Cx}})^{-1}\mathbf{W}^{(j)}} \tag{8}$$

Note that in (7), the size of the matrix to be inverted is $I_{Cx}\times I_{Cx}$, which can be very high in real problems. So we use the inversion formula in Chapter 3 of (Tarantola, 2005), for the reformulation.

Call $(\frac{\lambda_2}{\gamma}\mathbf{L}^T\mathbf{L}+\mathbf{I}_{I_{Cx}})=\mathbf{R}^T\mathbf{R}$ and $\mathbf{H}=(\tilde{\mathbf{B}}\mathbf{Q}_\mathbf{C}(:,j)-\frac{\lambda_1}{\gamma}\mathbf{1}_{I_{Cx}}-\frac{\beta_1(j)}{\gamma}\mathbf{W}^{(j)})$

$\mathbf{R}$ can be found from Cholesky decomposition. (7) will be:

$$\begin{aligned}\hat{\mathbf{M}}_\mathbf{C}(:,j)&=(\mathbf{K}^T\mathbf{K}+\gamma\mathbf{R}^T\mathbf{R})^{-1}(\mathbf{K}^T\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P}_\mathbf{C}(:,j)^T+\gamma\mathbf{H})\\&=\mathbf{R}^{-1}(\tilde{\mathbf{K}}^T\tilde{\mathbf{K}}+\gamma\mathbf{I}_{I_{Cx}})^{-1}(\tilde{\mathbf{K}}^T\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P}_\mathbf{C}(:,j)^T+\gamma\mathbf{R}^{-T}\mathbf{H})\\&=\mathbf{R}^{-1}\left\{\tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T+\gamma\mathbf{I}_{I_E})^{-1}(\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P}_\mathbf{C}(:,j)^T-\tilde{\mathbf{K}}\mathbf{R}^{-T}\mathbf{H})+\mathbf{R}^{-T}\mathbf{H}\right\}\end{aligned}$$

where $\tilde{\mathbf{K}}=\mathbf{K}\mathbf{R}^{-1}$

The same matrix manipulation can be used for the computation of orthogonality parameter in (8):

$$\beta_1(j)=\frac{\mathbf{W}^{(j)T}\mathbf{R}^{-1}\left\{\tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T+\gamma\mathbf{I})^{-1}(\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P}_\mathbf{C}(:,j)^T-\tilde{\mathbf{K}}\mathbf{R}^{-T}\mathbf{H})+\mathbf{R}^{-T}\mathbf{H}\right\}}{\frac{1}{\gamma}\mathbf{W}^{(j)T}\mathbf{R}^{-T}(\mathbf{I}-\tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T+\gamma\mathbf{I})^{-1}\tilde{\mathbf{K}})\mathbf{R}^{-1}\mathbf{W}^{(j)}}$$

where $\mathbf{H}=(\tilde{\mathbf{B}}\mathbf{Q}_\mathbf{C}(:,j)-\frac{\lambda_1}{\gamma}\mathbf{1}_{I_{Cx}})$

We skip the derivations of the discriminative signatures since formulation is very similar to common one. We present the final results.

Discriminative signature of EEG is estimated as:

$$\hat{\mathbf{M}}_\mathbf{G}(:,j)=\left[\mathbf{L}^{-1}\left\{\tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T+\mathbf{I}_{I_E})^{-1}(\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P}_\mathbf{G}(:,j)^T-\tilde{\mathbf{K}}\mathbf{L}^{-T}\mathbf{H})\right\}\right]_+$$

where $\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}=\boldsymbol{\mathcal{S}}_{T(1)}-\mathbf{K}\sum_{k\neq j}^{R_{DV}}\mathbf{M}_\mathbf{G}(:,k)\mathbf{P}_\mathbf{G}(:,k)^T-\mathbf{K}\,\mathbf{M}_\mathbf{C}\mathbf{P}_\mathbf{C}{}^T$, $\mathbf{W}^{(j)}=\sum_{k=1}^{R_C}\mathbf{M}_\mathbf{C}(:,k)+\sum_{k\neq j}^{R_{DV}}\mathbf{M}_\mathbf{V}(:,k)$,
$\mathbf{H}=(-\beta_2(j)\mathbf{W}^{(j)}-\lambda_3\mathbf{1}_{I_{Cx}})$, $\tilde{\mathbf{K}}=\mathbf{K}\mathbf{L}^{-1}$.

Regularization parameter for orthogonality constraint is found as:

$$\beta_2(j) = \frac{\mathbf{W}^{(j)T}\mathbf{L}^{-1}\left\{\tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \mathbf{I}_{I_E})^{-1}(\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P}_{\mathbf{G}}(:,j)^T - \tilde{\mathbf{K}}\mathbf{L}^{-T}\mathbf{H})\right\}}{\mathbf{W}^{(j)T}\mathbf{L}^{-1}(\mathbf{I}_{I_{Cx}} - \tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \mathbf{I}_{I_E})^{-1}\tilde{\mathbf{K}})\mathbf{L}^{-T}\mathbf{W}^{(j)}}$$

Discriminative signature of fMRI is estimated as:

$$\hat{\mathbf{M}}_{\mathbf{B}}(:,j) = \left[(\mathbf{I}_{I_{Cx}} + \lambda_6\mathbf{L}^T\mathbf{L})^{-1}(\gamma\tilde{\mathbf{B}}\mathbf{Q}_{\mathbf{B}}(:,j) - \lambda_5\mathbf{1}_{I_{Cx}} - \beta_3(j)\mathbf{W}^{(j)})\right]_+$$

where $\tilde{\mathbf{B}} = \mathbf{B} - \sum_{k\neq j}^{R_{DB}}\mathbf{M}_{\mathbf{B}}(:,k)\mathbf{Q}_{\mathbf{B}}(:,k)^T - \mathbf{M}_{\mathbf{C}}\mathbf{Q}_{\mathbf{C}}{}^T$, $\mathbf{W}^{(j)} = \sum_{k=1}^{R_C}\mathbf{M}_{\mathbf{C}}(:,k) + \sum_{k\neq j}^{R_{DB}}\mathbf{M}_{\mathbf{B}}(:,k)$.

Orthogonality regularization parameter is found as:

$$\beta_3(j) = \frac{\mathbf{W}^{(j)T}(\mathbf{I}_{I_{Cx}} + \lambda_6\mathbf{L}^T\mathbf{L})^{-1}(\gamma\tilde{\mathbf{B}}\mathbf{Q}_{\mathbf{B}}(:,j) - \lambda_1\mathbf{1}_{I_{Cx}})}{\mathbf{W}^{(j)T}(\mathbf{I}_{I_{Cx}} + \lambda_6\mathbf{L}^T\mathbf{L})^{-1}\mathbf{W}^{(j)}}.$$

Estimation of Other Signatures:

Other signatures are estimated from ALS as follows:

$$\mathbf{T}_{\mathbf{V}} = \boldsymbol{\mathcal{S}}_{T(2)}(\mathbf{F}_{\mathbf{V}} \odot [\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{V}}])^{\dagger}$$

$$\mathbf{F}_{\mathbf{V}} = \boldsymbol{\mathcal{S}}_{T(3)}(\mathbf{T}_{\mathbf{V}} \odot [\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{V}}])^{\dagger}$$

$$\mathbf{T}_{\mathbf{B}} = \mathbf{B}^T([\mathbf{M}_{\mathbf{C}},\mathbf{M}_{\mathbf{B}}])^{\dagger}.$$

CMTF Algorithm

Inputs : $\boldsymbol{\mathcal{S}}_T$, $\mathbf{B}, \mathbf{K}, \mathbf{L}, R_C, R_{DV}, R_{DB}, \gamma, \{\lambda_j\}_{j=1}^{6}$

Outputs : $\mathbf{M_C}, \mathbf{M_G}, \mathbf{T_V}, \mathbf{F_V}, \mathbf{M_B}, \mathbf{T_B}$

Initialize : $\{\mathbf{M_C}, \mathbf{M_G}, \mathbf{T_V}, \mathbf{F_V}, \mathbf{M_B}, \mathbf{T_B}\}$

repeat until convergence

Estimation of Spatial Signatures

$$\mathbf{R} = \mathrm{chol}(\frac{\lambda_2}{\gamma}\mathbf{L}^T\mathbf{L} + \mathbf{I})$$

$$\mathbf{P_C} = (\mathbf{F_V}(:,1:R_C) \odot \mathbf{T_V}(:,1:R_C))$$

$$\mathbf{P_G} = (\mathbf{F_V}(:,R_C+1:R_V) \odot \mathbf{T_V}(:,R_C+1:R_V))$$

$$\mathbf{Q_C} = \mathbf{T_B}(:,1:R_C)$$

$$\mathbf{Q_B} = \mathbf{T_B}(:,R_C+1:R_B)$$

for j = 1,2, ...,max(R_V,R_B) **do**

if $j \le R_C$

Estimation of $\mathbf{M_C}$:

$$\tilde{\boldsymbol{\mathcal{S}}}_{T(1)} = \boldsymbol{\mathcal{S}}_{T(1)} - \mathbf{K}\sum_{k\neq j}^{R_C}\mathbf{M_C}(:,k)\mathbf{P_C}(:,k)^T - \mathbf{K}\,\mathbf{M_G}\mathbf{P_G}^T$$

$$\tilde{\mathbf{B}} = \mathbf{B} - \sum_{k\neq j}^{R_C}\mathbf{M_C}(:,k)\mathbf{Q_C}(:,k)^T - \mathbf{M_B}\mathbf{Q_B}^T$$

$$\mathbf{W}^{(j)} = \sum_{k\neq j}^{R_C}\mathbf{M_C}(:,k) + \sum_{k=1}^{R_{DV}}\mathbf{M_G}(:,k) + \sum_{k=1}^{R_{DB}}\mathbf{M_B}(:,k)$$

$$\mathbf{H} = (\tilde{\mathbf{B}}\mathbf{Q_C}(:,j) - \frac{\lambda_1}{\gamma}\mathbf{1})$$

$$\beta_1(j) = \frac{\mathbf{W}^{(j)T}\mathbf{R}^{-1}\left\{\tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \gamma\mathbf{I})^{-1}(\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P_C}(:,j)^T - \tilde{\mathbf{K}}\mathbf{R}^{-T}\mathbf{H}) + \mathbf{R}^{-T}\mathbf{H}\right\}}{\frac{1}{\gamma}\mathbf{W}^{(j)T}\mathbf{R}^{-1}(\mathbf{I} - \tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \gamma\mathbf{I})^{-1}\tilde{\mathbf{K}})\mathbf{R}^{-T}\mathbf{W}^{(j)}}$$

$$\mathbf{H} = (\mathbf{H} - \frac{\beta_1(j)}{\gamma}\mathbf{W}^{(j)})$$

$$\tilde{\mathbf{K}} = \mathbf{K}\mathbf{R}^{-1}$$

$$\hat{\mathbf{M}}_{\mathbf{C}}(:,j) = \left[\mathbf{R}^{-1}\left\{\tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \gamma\mathbf{I})^{-1}(\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P_C}(:,j)^T - \tilde{\mathbf{K}}\mathbf{R}^{-T}\mathbf{H}) + \mathbf{R}^{-T}\mathbf{H}\right\}\right]_+$$

end if

Estimation of $\mathbf{M_G}$:

if $(j > R_C)\,\&\,(j < R_V)$

$$\tilde{\boldsymbol{\mathcal{S}}}_{T(1)} = \boldsymbol{\mathcal{S}}_{T(1)} - \mathbf{K}\sum_{k\neq j}^{R_{DV}}\mathbf{M_G}(:,k)\mathbf{P_G}(:,k)^T - \mathbf{K}\,\mathbf{M_C}\mathbf{P_C}^T$$

$$\mathbf{W}^{(j)} = \sum_{k=1}^{R_C}\mathbf{M_C}(:,k) + \sum_{k\neq j}^{R_{DV}}\mathbf{M_V}(:,k)$$

$$\mathbf{H} = (-\lambda_3\mathbf{1})$$

$$\beta_2(j) = \frac{\mathbf{W}^{(j)T}\mathbf{L}^{-1}\left\{\tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \mathbf{I})^{-1}(\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P_G}(:,j)^T - \tilde{\mathbf{K}}\mathbf{L}^{-T}\mathbf{H})\right\}}{\mathbf{W}^{(j)T}\mathbf{L}^{-1}(\mathbf{I} - \tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \mathbf{I})^{-1}\tilde{\mathbf{K}})\mathbf{L}^{-T}\mathbf{W}^{(j)}}$$

$$\mathbf{H} = (\mathbf{H} - \beta_2(j)\mathbf{W}^{(j)})$$

$$\tilde{\mathbf{K}} = \mathbf{K}\mathbf{L}^{-1}$$

$$\hat{\mathbf{M}}_{\mathbf{G}}(:,j) = \left[\mathbf{L}^{-1}\left\{\tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \mathbf{I})^{-1}(\tilde{\boldsymbol{\mathcal{S}}}_{T(1)}\mathbf{P_G}(:,j)^T - \tilde{\mathbf{K}}\mathbf{L}^{-T}\mathbf{H})\right\}\right]_+$$

end if

Estimation of $\mathbf{M_B}$:

if $(j > R_C)\,\&\,(j < R_B)$

$$\tilde{\mathbf{B}} = \mathbf{B} - \sum_{k \neq j}^{R_{DB}} \mathbf{M_B}(:,k)\mathbf{Q_B}(:,k)^T - \mathbf{M_C}\mathbf{Q_C}^T$$

$$\mathbf{W}^{(j)} = \sum_{k=1}^{R_C} \mathbf{M_C}(:,k) + \sum_{k \neq j}^{R_{DB}} \mathbf{M_B}(:,k)$$

$$\beta_3(j) = \frac{\mathbf{W}^{(j)T}\left\{(\mathbf{I} + \lambda_6 \mathbf{L}^T\mathbf{L})^{-1}(\gamma\tilde{\mathbf{B}}\mathbf{Q_B}(:,j) - \lambda_5 \mathbf{1})\right\}}{\mathbf{W}^{(j)T}(\mathbf{I} + \lambda_6 \mathbf{L}^T\mathbf{L})^{-1}\mathbf{W}^{(j)}}$$

$$\hat{\mathbf{M}}_\mathbf{B}(:,j) = \left[(\mathbf{I} + \lambda_6 \mathbf{L}^T\mathbf{L})^{-1}(\gamma\tilde{\mathbf{B}}\mathbf{Q_B}(:,j) - \lambda_5 \mathbf{1} - \beta_3(j)\mathbf{W}^{(j)})\right]_+$$

end if

end for

Estimation of Other Signatures

$$\mathbf{T_V} = \boldsymbol{\mathcal{S}}_{T(2)}(\mathbf{F_V} \odot [\mathbf{M_C}, \mathbf{M_V}])^{\dagger}$$

$$\mathbf{F_V} = \boldsymbol{\mathcal{S}}_{T(3)}(\mathbf{T_V} \odot [\mathbf{M_C}, \mathbf{M_V}])^{\dagger}$$

$$\mathbf{T_B} = \mathbf{B}^T([\mathbf{M_C}, \mathbf{M_B}])^{\dagger}$$

References

Kimura, K., Tanaka, Y., & Kudo, M. (2014). A Fast Hierarchical Alternating Least Squares Algorithm for Orthogonal Nonnegative Matrix Factorization. In *JMLR: Workshop and Conference Proceedings* (pp. 129–141).

Kolda, T. G. (2006). *Multilinear operators for higher-order decompositions*. *SAND2006-2081*. Albuquerque, NM, Livermore, CA. Retrieved from http://www.osti.gov/bridge/product.biblio.jsp?osti_id=923081

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. *SIAM*. SIAM. doi:http://dx.doi.org/10.1137/1.9780898717921