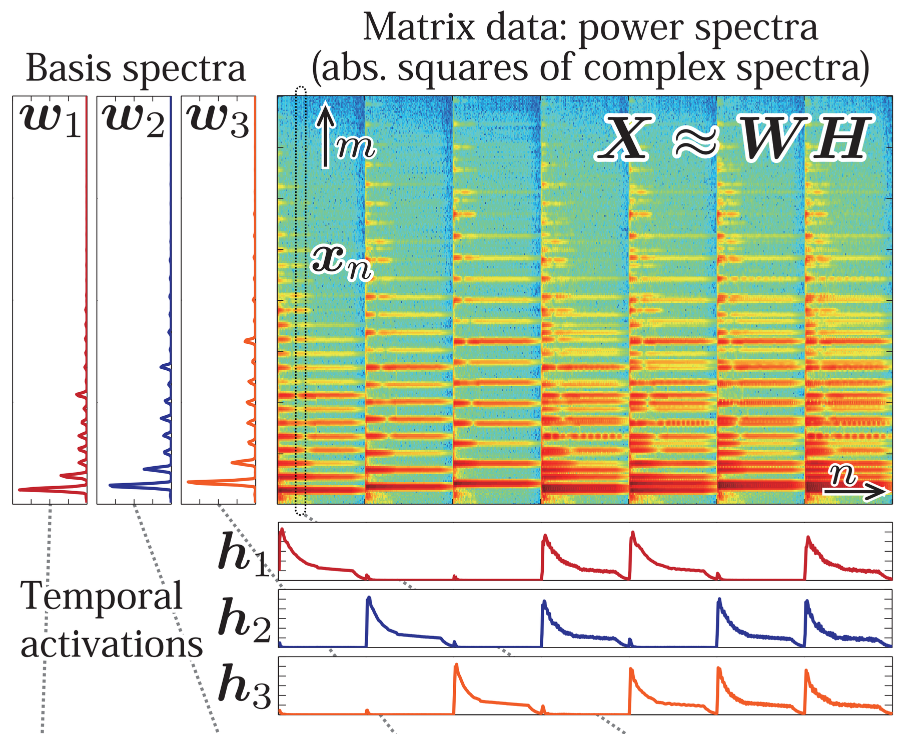
Infinite Positive Semidefinite Tensor Factorization for Source Separation of Mixture Signals

Kazuyoshi Yoshii[1] Ryota Tomioka[2] Daichi Mochihashi[3] Masataka Goto[1]

[1] National Institute of Advanced Industrial Science and Technology (AIST)
[2] The University of Tokyo [3] The Institute of Statistical Mathematics (ISM)

Conventional: Nonnegative Matrix Factorization (NMF)

Each nonnegative vector is approximated by a convex combination of nonnegative vectors



The diagonal elements of PSD matrices are always nonnegative

$$x_n \approx \sum_{k=1}^{K} w_k h_{kn} \overset{\text{def}}{=} y_n$$

Vector-wise factorization

Observed matrix (nonnegative vectors)

$X = [x_1, \cdots, x_N] \in \mathbb{R}^{M\times N}$

Basis matrix (nonnegative vectors)

$W = [w_1, \cdots, w_K] \in \mathbb{R}^{M\times K}$

Activation matrix (nonnegative vectors)

$H = [h_1, \cdots, h_K]^T \in \mathbb{R}^{K\times N}$

Reconstruction error (Bregman divergence) $\phi(x)$: Strictly convex function

$$\mathcal{D}_\phi(x_n|y_n) = \phi(x_n) - \phi(y_n) - \phi'(y_n)^T(x_n - y_n)$$

Kullback-Leibler (KL) divergence $\phi(x) = \sum_m (x_m \log x_m - x_m)$

$$\mathcal{D}_{\text{KL}}(x_n|y_n) = \sum_m \left(x_{mn} \log x_{mn} y_{mn}^{-1} - x_{mn} + y_{mn}\right)$$

Itakura-Saito (IS) divergence $\phi(x) = -\sum_m \log x_m$

$$\mathcal{D}_{\text{IS}}(x_n|y_n) = \sum_m \left(-\log x_{mn} y_{mn}^{-1} + x_{mn} y_{mn}^{-1} - 1\right)$$

IS-NMF is theoretically justified for factorizing the power spectrogram

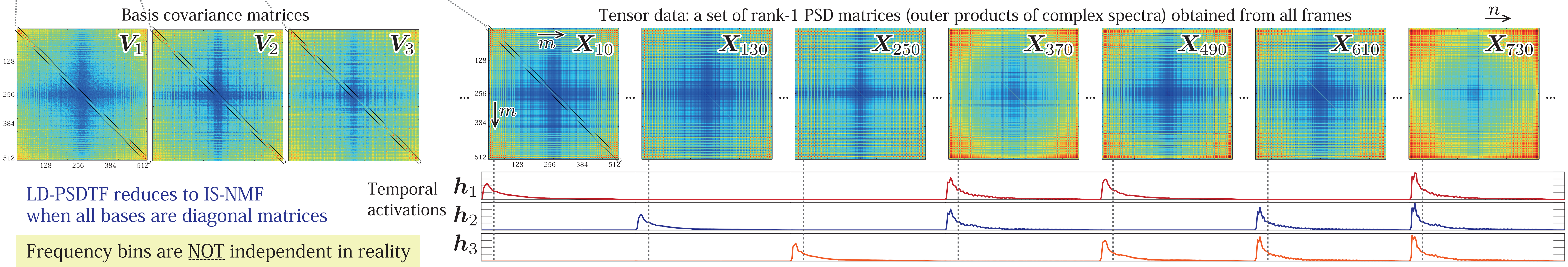
Given X, NMF tries to estimate W and H such that $\mathcal{C}(X|Y) = \sum_n \mathcal{D}_\phi(x_n|y_n)$ is minimized

Maximum-likelihood estimation

Multiplicative update rules can be derived for efficient learning

Proposed: Positive Semidefinite Tensor Factorization (PSDTF)

Each positive semidefinite matrix is approximated by a convex combination of positive semidefinite matrices



LD-PSDTF reduces to IS-NMF when all bases are diagonal matrices

Frequency bins are NOT independent in reality

$$X_n \approx \sum_{k=1}^{K} V_k h_{kn} \overset{\text{def}}{=} Y_n$$

Matrix-wise factorization

Observed tensor (PSD matrices)

$X = [X_1, \cdots, X_N] \in \mathbb{R}^{M\times M\times N}$

Basis tensor (PSD matrices)

$V = [V_1, \cdots, V_K] \in \mathbb{R}^{M\times M\times K}$

Activation matrix (nonnegative vectors)

$H = [h_1, \cdots, h_K]^T \in \mathbb{R}^{K\times N}$

Reconstruction error (Bregman matrix divergence)

$$\mathcal{D}_\phi(X_n|Y_n) = \phi(X_n) - \phi(Y_n) - \text{tr}(\nabla\phi(Y_n)^T(X_n - Y_n))$$

von Neumann (vN) divergence $\phi(X) = \text{tr}(X \log X - X)$

$$\mathcal{D}_{\text{vN}}(X_n|Y_n) = \text{tr}\left(X_n \log X_n Y_n^{-1} - X_n + Y_n\right)$$

Log-determinant (LD) divergence $\phi(X) = -\log|X|$

$$\mathcal{D}_{\text{LD}}(X_n|Y_n) = -\log\left|X_n Y_n^{-1}\right| + \text{tr}\left(X_n Y_n^{-1}\right) - M$$

Given X, PSDTF tries to estimate V and H such that $\mathcal{C}(X|Y) = \sum_n \mathcal{D}_\phi(X_n|Y_n)$ is minimized

Maximum-likelihood estimation

Nonparametric Bayesian modeling

The observed tensor is allowed to consist of an unbounded number of bases

$$X_n \approx \sum_{k=1}^{K\to\infty} \theta_k V_k h_{kn} \overset{\text{def}}{=} Y_n$$

Global weight vector $\theta = [\theta_1, \cdots, \theta_\infty]^T \in \mathbb{R}^\infty$

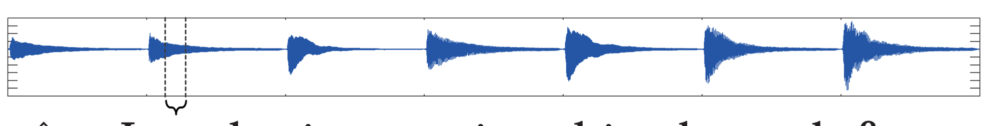
To learn an infinite-dimensional sparse vector, we place a gamma process (GaP) prior on θ

$$\theta \sim \text{GaP}(\alpha, \text{Uniform})$$

Application to Single-Channel Audio Signal Separation

Factorize frame-level covariances within a single-channel signal

Time-domain decomposition of mixture signals based on LD-PSDTF and Wiener filtering



$\hat{x}_n$: Local mixture signal in the n-th frame

Suppose $\hat{x}_n$ can be decomposed as follows:

$$\hat{x}_n = \sum_{k=1}^{K} \underset{\text{Latent}}{\hat{x}_{nk}} = \sum_{k=1}^{K} \hat{w}_{kn} \hat{h}_{kn}$$

$\hat{w}_{kn}$: Local basis signal in the n-th frame

$\hat{h}_{kn}$: Coefficient of the k-th basis signal

We assume that each basis signal is "stationary"

$$\hat{w}_{kn} \sim \mathcal{N}(\mathbf{0}, \hat{V}_k)$$

$\hat{V}_k$ defines signal characteristics (e.g., periodicity and/or whiteness)

The reproducing property of the Gaussian gives

$$\hat{x}_n | \hat{V}, \hat{H} \sim \mathcal{N}\left(\mathbf{0}, \sum_{k=1}^{K} \hat{V}_k \hat{h}_{kn}^2\right)$$

$X_n = \hat{x}_n \hat{x}_n^T \geq 0$

$V_k = \hat{V}_k \geq 0$

$h_{kn} = \hat{h}_{kn}^2 \geq 0$

$$\log p(X_n|Y_n) \overset{c}{=} -\frac{1}{2}\log|Y_n| - \frac{1}{2}\text{tr}(X_n Y_n^{-1})$$

The maximization of the Gaussian log-likelihood is equivalent to the minimization of the LD divergence

Finally, we perform Wiener filtering as follows:

$$\mathbb{E}[\hat{x}_{nk}|\hat{x}_n, V, H] = Y_{nk} Y_n^{-1} \hat{x}_n \qquad Y_{nk} = V_k h_{kn}$$

$$\mathbb{V}[\hat{x}_{nk}|\hat{x}_n, V, H] = Y_{nk} - Y_{nk} Y_n^{-1} Y_{nk} \qquad Y_n = \sum_k Y_{nk}$$

Source separation results on piano data:

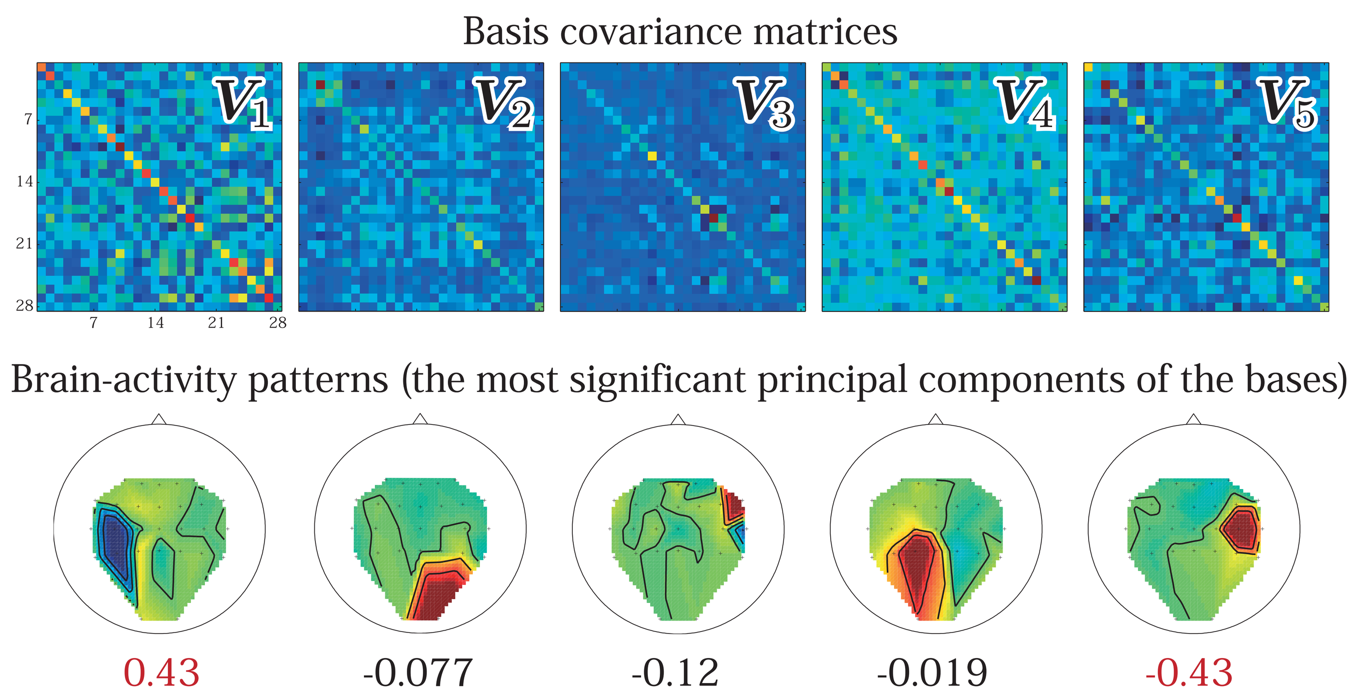
IS-NMF: SDR 18.9dB, SIR 24.1dB, SAR 20.5dB

LD-PSDTF: SDR 26.7dB, SIR 33.2dB, SAR 27.8dB

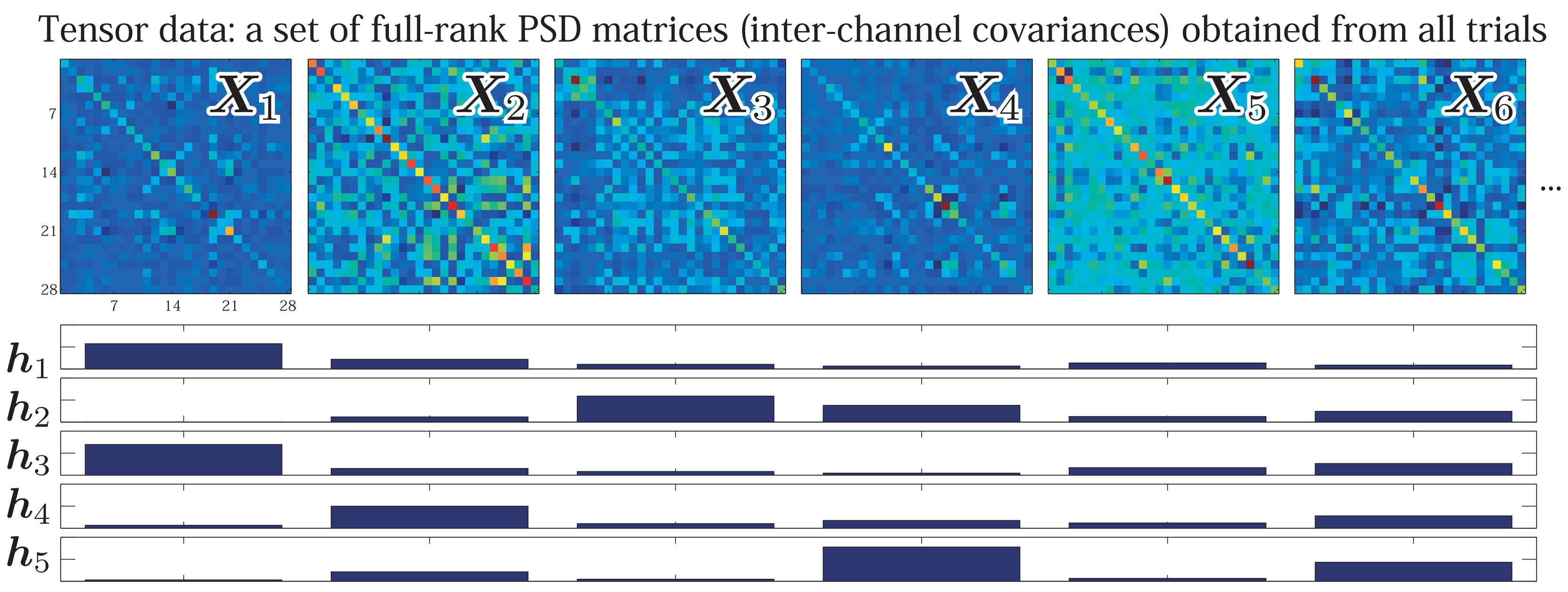
Application to Multi-Channel EEG Signal Analysis

Factorize trial-level covariances between multi-channel signals

Unsupervised learning of characteristic brain-activity patterns based on LD-PSDTF



Each value indicates the correlation between activations and ground-truth labels



Task: predict a left or right hand movement (-1 or 1) for a given EEG

There were 416 trials (N=416), in which 100 were used for evaluation.

For each trial, an EEG was recorded at 28 channels (M=28) for 500 ms and an inter-channel covariance was calculated from the EEG.

Well-known physiological processes called event-related desynchronization (ERD) were successfully discovered.

The accuracies were 73% (K=5) and 79% (K=10) when Fisher's LDA was used for discriminating an K-dim. feature vector of each trial.

Future Work

Reduce the heavy computational cost

Investigate vN-PSDTF (extension of KL-NMF)