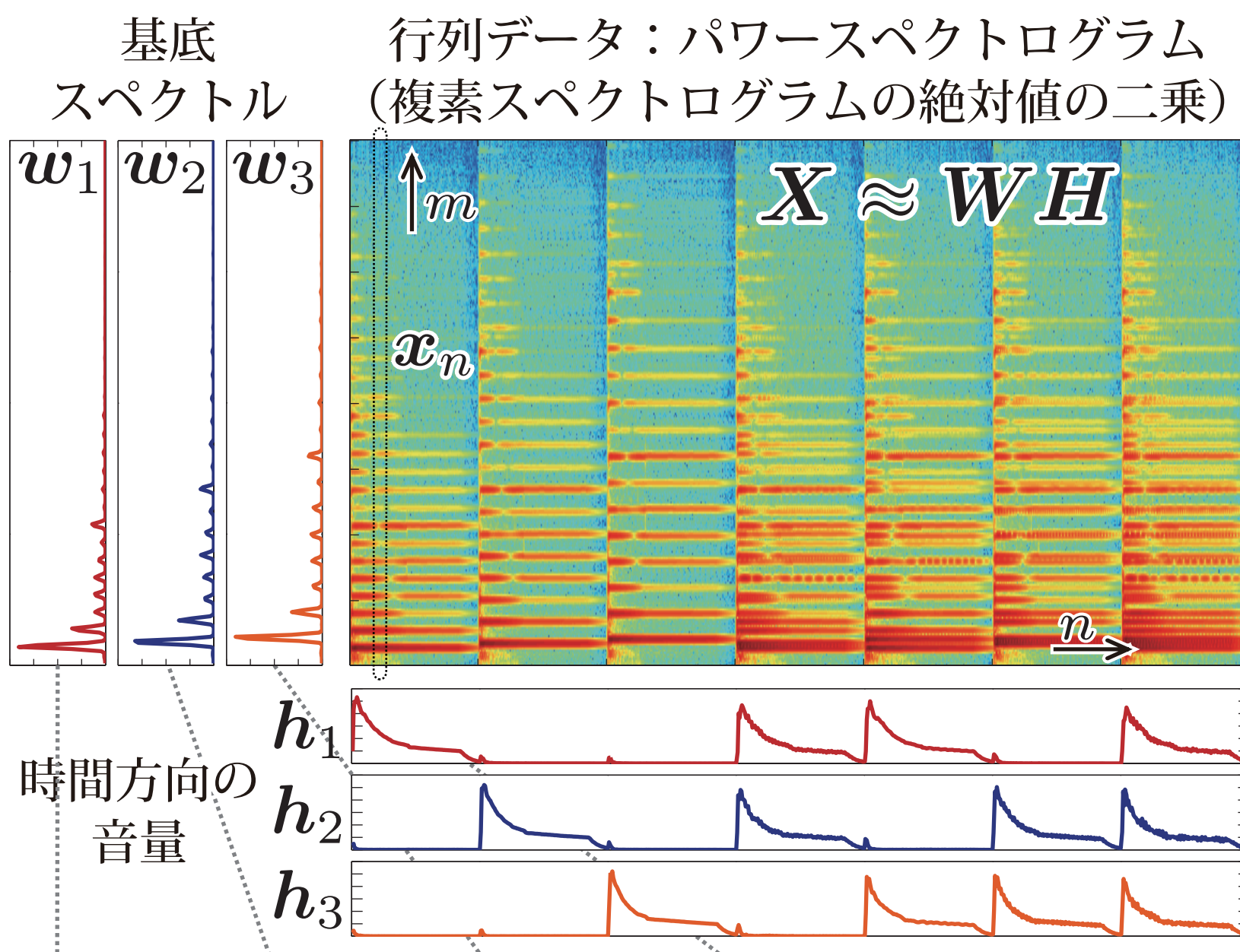


音楽音響信号解析のためのステューデント t 分布に基づく 非負値行列分解と半正定値テンソル分解

吉井 和佳 (京大) 糸山 克寿 (京大) 後藤 真孝 (産総研)

非負値行列分解 (Nonnegative Matrix Factorization: NMF) [1]

非負値ベクトル (観測ベクトル) を少数の非負値ベクトル (基底ベクトル) の線形結合で近似



観測行列 (非負値ベクトル群) $X = [x_1, \dots, x_N] \in \mathbb{R}^{M \times N}$

基底行列 (非負値ベクトル群) $W = [w_1, \dots, w_K] \in \mathbb{R}^{M \times K}$

音量行列 (非負値ベクトル群) $H = [h_1, \dots, h_K]^T \in \mathbb{R}^{K \times N}$

$x_n \approx \sum_{k=1}^K h_{kn} w_k \stackrel{\text{def}}{=} y_n$ (ベクトル単位の分解)

Itakura-Saito (IS) ダイバージェンス

$$\mathcal{D}_{\text{IS}}(x_n | y_n) = \sum_{m=1}^M (-\log x_{nm} y_{nm}^{-1} + x_{nm} y_{nm}^{-1} - 1)$$

$$X_n = s_n s_n^H$$

$$x_n = \text{diag}(X_n)$$

$$y_n = \text{diag}(Y_n)$$

$$w_k = \text{diag}(W_k)$$

単変量複素ガウス分布に基づく対数尤度

$$\log p(s_n | y_n) = \log \mathcal{N}_c(s_n | \mathbf{0}, \text{Diag}(y_n)) \stackrel{c}{=} - \sum_{m=1}^M (\log y_{nm} + x_{nm} y_{nm}^{-1})$$

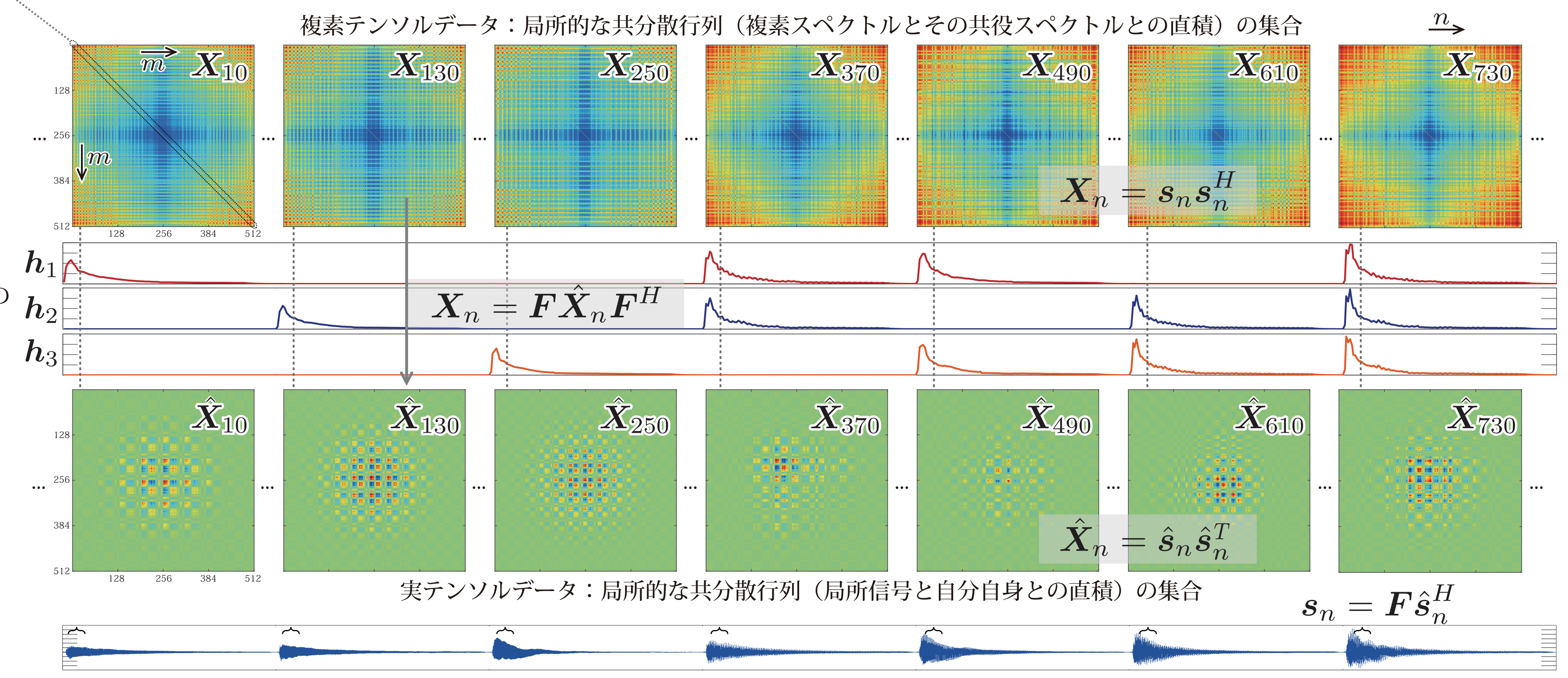
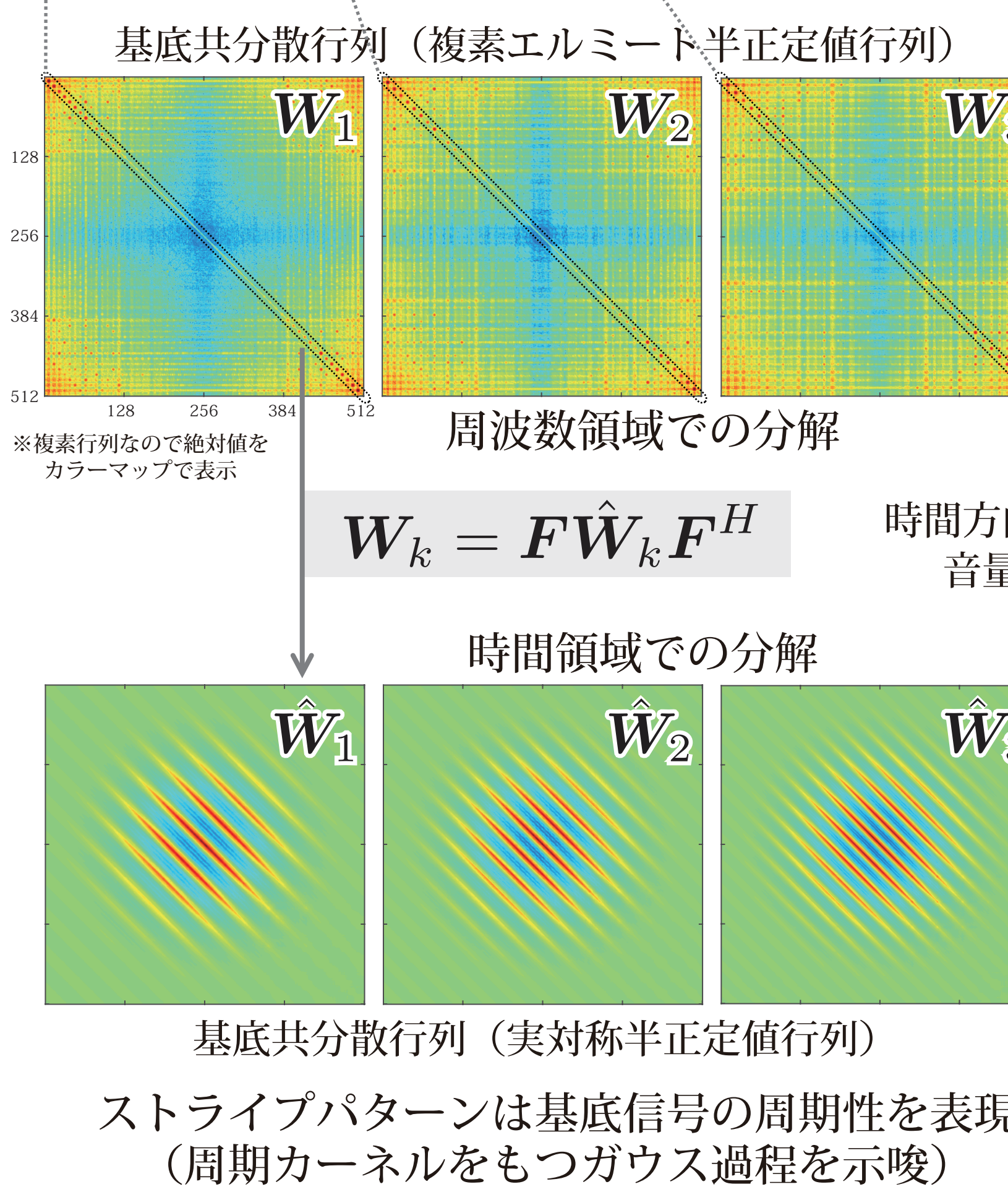
X が与えられたときに以下を満たす W, H を求める

$$\mathcal{C}(X|Y) = \sum_{n=1}^N \mathcal{D}_{\text{IS}}(x_n | y_n) \text{ を最小化} \Leftrightarrow \log p(X|Y) = \sum_{n=1}^N \log p(s_n | y_n) \text{ を最大化}$$

非負値ベクトル = 半正定値行列の対角成分 Pros: パワースペクトルの加法性 (ガウス分布の再生性) が成立 Cons: 位相が取り扱えないので分離品質に限界

半正定値テンソル分解 (Positive Semidefinite Tensor Factorization: PSDTF) [2]

半正定値行列 (観測行列) を少数の半正定値行列 (基底行列) の線形結合で近似



観測テンソル (複素エルミート半正定値行列群) $\mathcal{X} = [X_1, \dots, X_N] \in \mathbb{C}^{M \times M \times N}$

基底テンソル (複素エルミート半正定値行列群) $\mathcal{W} = [W_1, \dots, W_K] \in \mathbb{C}^{M \times M \times K}$

周波数領域表現

音量行列 (非負値ベクトル群) $H = [h_1, \dots, h_K]^T \in \mathbb{R}^{K \times N}$

$X_n \approx \sum_{k=1}^K h_{kn} W_k \stackrel{\text{def}}{=} Y_n \Leftrightarrow \hat{X}_n \approx \sum_{k=1}^K h_{kn} \hat{W}_k \stackrel{\text{def}}{=} \hat{Y}_n$ (行列単位の分解)

観測テンソル (実対称半正定値行列群) $\hat{\mathcal{X}} = [\hat{X}_1, \dots, \hat{X}_N] \in \mathbb{R}^{M \times M \times N}$

基底テンソル (実対称半正定値行列群) $\hat{\mathcal{W}} = [\hat{W}_1, \dots, \hat{W}_K] \in \mathbb{R}^{M \times M \times K}$

時間領域表現

Log-Determinant (LD) ダイバージェンス

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

多変量複素ガウス分布に基づく対数尤度

$$\log p(s_n | Y_n) = \log \mathcal{N}_c(s_n | \mathbf{0}, Y_n) \stackrel{c}{=} -\log |Y_n| - \text{tr}(X_n Y_n^{-1})$$

\mathcal{X} が与えられたときに以下を満たす \mathcal{W}, H を求める

$$\mathcal{C}(\mathcal{X} | \mathcal{Y}) = \sum_{n=1}^N \mathcal{D}_{\text{LD}}(X_n | Y_n) \text{ を最小化} \Leftrightarrow \log p(\mathcal{X} | \mathcal{Y}) = \sum_{n=1}^N \log p(s_n | Y_n) \text{ を最大化}$$

Pros: パワースペクトルの加法性が成立・位相が取り扱える Cons: 実信号は裾が重い

提案法: 複素ステューデント t 分布に基づく半正定値テンソル分解 (t-PSDTF)

IS-NMF [1], Cauchy NMF [3], LD-PSDTF [2] を含む統一的な確率モデル: 自由度の調節による初期値依存性の低減が可能

多変量複素 t 分布

$$\mathcal{T}_c^\nu(x | \mathbf{0}, \Sigma) = \frac{\Gamma(\frac{2d+\nu}{2})}{\Gamma(\frac{\nu}{2})} \frac{2^d}{(\pi\nu)^d} |\Sigma|^{-1} \left(1 + \frac{2}{\nu} x^H \Sigma^{-1} x\right)^{-\frac{2d+\nu}{2}}$$

多変量複素ガウス分布: 自由度 ∞ の t 分布

$$\mathcal{N}_c(x | \mathbf{0}, \Sigma) = \frac{1}{\pi^d} |\Sigma|^{-1} \exp(-x^H \Sigma^{-1} x)$$

単変量ガウス分布の再生性 \rightarrow パワースペクトルは加法的 \rightarrow IS-NMF [1]

多変量ガウス分布の再生性 \rightarrow パワースペクトルは加法的 \rightarrow LD-PSDTF [2]

多変量複素コーシー分布: 自由度 1 の t 分布

$$\mathcal{C}_c(x | \mathbf{0}, \Sigma) = \frac{\Gamma(\frac{2d+1}{2}) 2^d}{\pi^{\frac{2d+1}{2}}} |\Sigma|^{-1} \left(1 + 2x^H \Sigma^{-1} x\right)^{-\frac{2d+1}{2}}$$

単変量コーシー分布の再生性 \rightarrow 振幅スペクトルは加法的 \rightarrow Cauchy NMF [3]

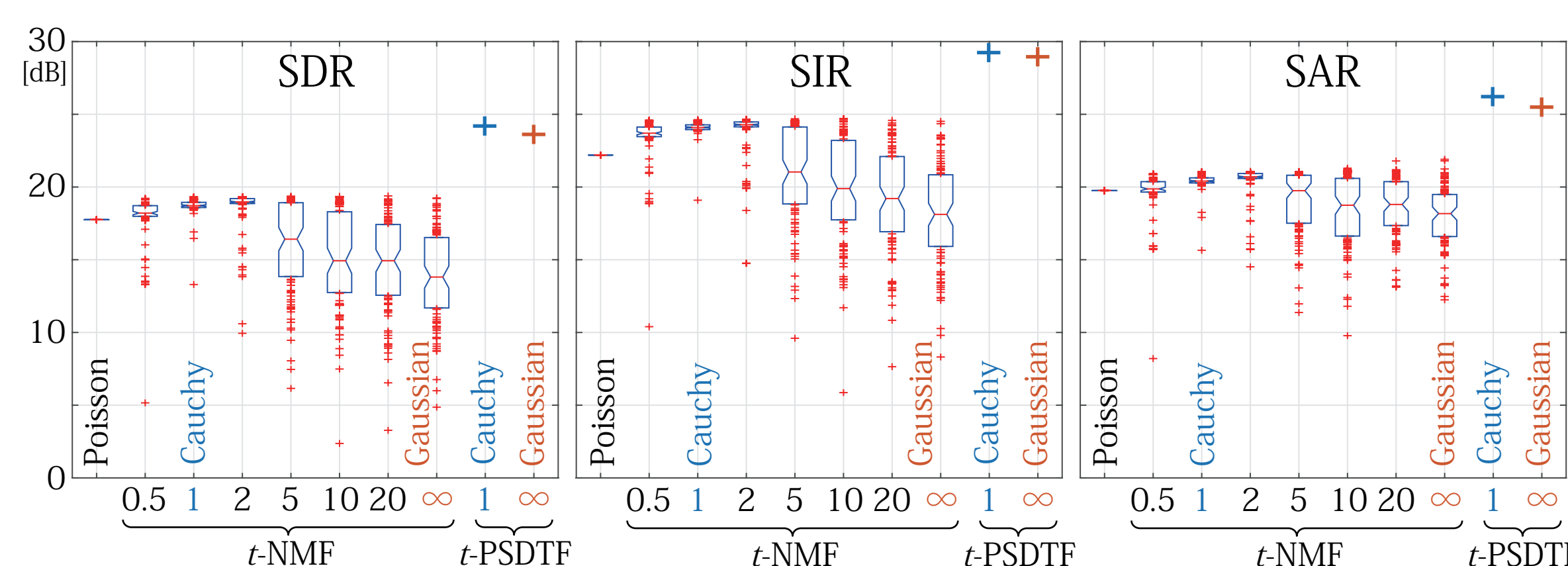
多変量コーシー分布の再生性 \rightarrow スペクトルは非加法的 \rightarrow Cauchy PSDTF [本研究]

多変量複素 t 分布に基づく対数尤度

$$\log p(s_n | Y_n) = \log \mathcal{T}_c^\nu(s_n | \mathbf{0}, Y_n) \stackrel{c}{=} -\log |Y_n| - \left(M + \frac{\nu}{2}\right) \log \left(1 + \frac{2}{\nu} \text{tr}(X_n Y_n^{-1})\right)$$

対数尤度最大化のため補助関数法を用いた効率的な乗法更新アルゴリズムを導出

(IS-NMF, LD-PSDTF の更新則を特殊形に含む・Cauchy NMF の新しい更新則を与える)



ピアノ・ギター・クラリネットの単独音を用いて C4, E4, G4, C4+E4, C4+G4, E4+G4, C4+E4+G4 を順番に配置して合成音を作成

自由度を変えた t-NMF と t-PSDTF を用いて分離精度を評価

t-NMF は IS-NMF より安定して品質の良い局所解に収束
スペクトル加法性の保証のない自由度 2 の場合も健闘

将来展望: 複素安定分布に基づく NMF/PSDTF (cf. [4])