「PSDTF」
NMF の
最も本質的で
美しい拡張

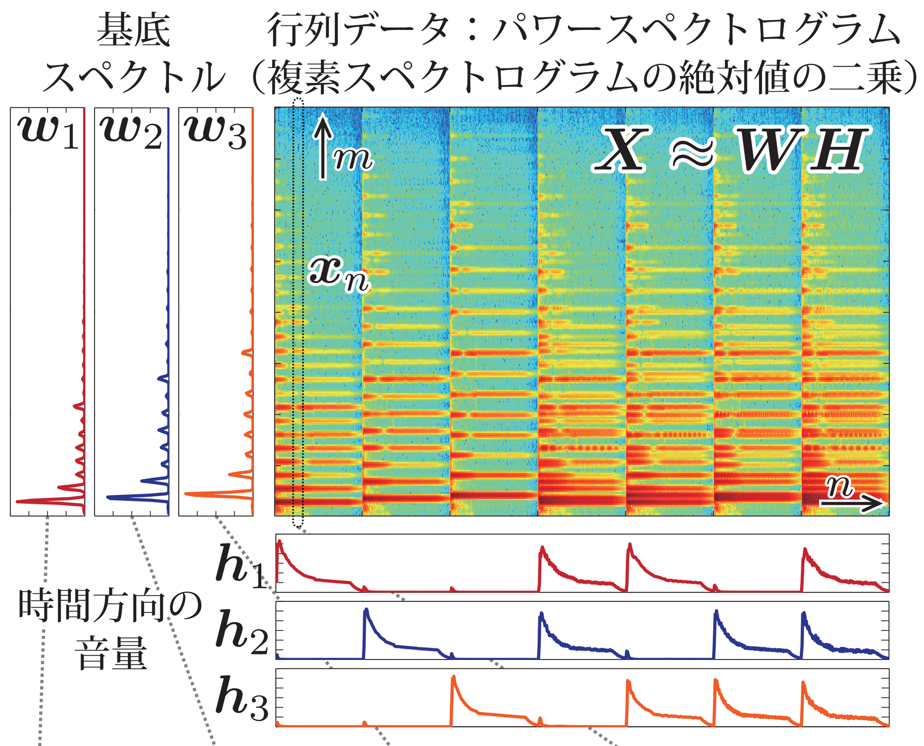
音楽音響信号解析のためのガンマ過程に基づく無限半正定値テンソル分解

吉井 和佳（産総研） 富岡 亮太（東大） 持橋 大地（統数研） 後藤 真孝（産総研）

Matlab ソースコード公開（二条項 BSD ライセンス！）http://staff.aist.go.jp/k.yoshii/psdtf/

従来法：非負値行列分解 (Nonnegative Matrix Factorization: NMF)

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）の凸結合で近似



非負値ベクトル＝半正定値行列の対角成分

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

ベクトル単位の分解

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

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

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

コスト関数（Bregmanダイバージェンス）$\phi(x)$: 厳密に凸な関数

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

Kullback-Leibler (KL) ダイバージェンス $\phi(x) = \sum_m (x_m \log x_m - x_m)$

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

Itakura-Saito (IS) ダイバージェンス $\phi(x) = -\sum_m \log x_m$

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

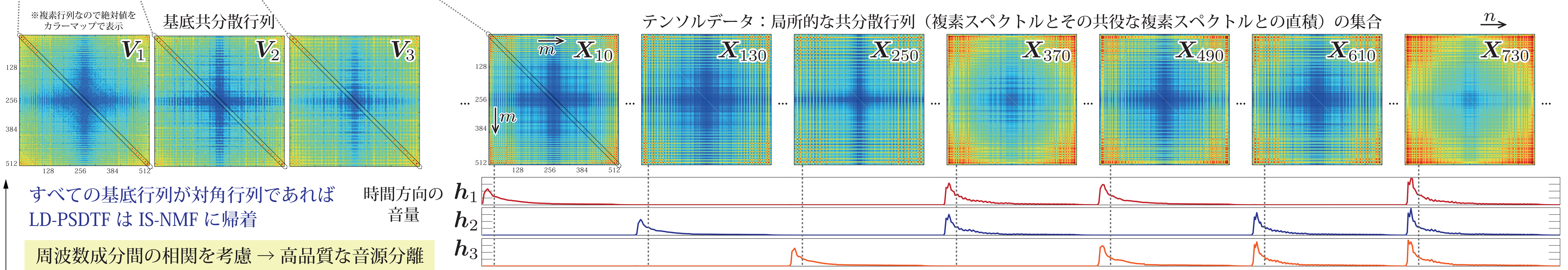
・スケール不変な尺度
・信号がガウス性雑音で
あれば理論的に妥当

X が与えられたときに以下を満たす W, H を求めたい
総コスト $\mathcal{C}(X|Y) = \sum_n \mathcal{D}_\phi(x_n|y_n)$ を最小化
ある確率モデルの最尤推定に対応

・乗法更新則による学習が可能
・ガンマ過程事前分布に基づく
基底数の自動決定も可能

提案法：半正定値テンソル分解 (Positive Semidefinite Tensor Factorization: PSDTF)

半正定値行列（観測行列）を少数の半正定値行列（基底行列）の凸結合で近似



すべての基底行列が対角行列であれば
LD-PSDTF は IS-NMF に帰着

周波数成分間の相関を考慮 → 高品質な音源分離

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

行列単位の分解

観測テンソル（半正定値行列の集合）
$X = [X_1, \cdots, X_N] \in \mathbb{R}^{M \times M \times N}$

基底テンソル（半正定値行列の集合）
$V = [V_1, \cdots, V_K] \in \mathbb{R}^{M \times M \times K}$

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

コスト関数（Bregman行列ダイバージェンス）

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

von Neumann (vN) ダイバージェンス $\phi(X) = \mathrm{tr}(X \log X - X)$

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

Log-Determinant (LD) ダイバージェンス $\phi(X) = -\log|X|$

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

X が与えられたときに以下を満たす V, H を求めたい
総コスト $\mathcal{C}(X|Y) = \sum_n \mathcal{D}_\phi(X_n|Y_n)$ を最小化
ある確率モデルの最尤推定に対応

・乗法更新則
による学習
が可能

ノンパラメトリックベイズモデル

観測行列を無限個の基底行列の凸結合で表現
（実質的には有限個の基底行列のみが活性化）

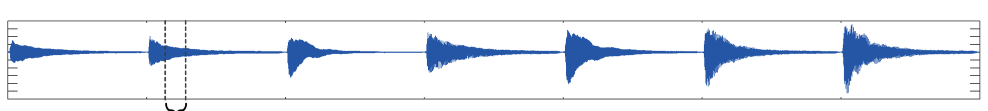
$$X_n \approx \sum_{k=1}^{K \to \infty} \theta_k V_k h_{kn} \overset{\text{def}}{=} Y_n$$

基底の重みベクトル $\theta = [\theta_1, \cdots, \theta_\infty]^T \in \mathbb{R}^\infty$
スパースかつ非負の無限次元ベクトル θ を生成するため、事前分布としてガンマ過程を用いる

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

モノラル音楽音響信号に対する音源分離への応用

LD-PSDTFによる音源信号の統計的性質の推定 ＋ ウィナーフィルタリングによる音源信号の復元



o_n：フレーム n における観測信号

観測信号に対する音源信号の瞬時混合モデル

$$\underset{\text{観測信号}}{o_n} = \sum_{k=1}^{K} \underset{\text{音源信号}}{o_{kn}} = \sum_{k=1}^{K} \pi_{kn} \phi_{kn}$$

ϕ_{kn}：フレーム n における基底信号 k
π_{kn}：上記信号のフレーム n における重み

各基底信号が定常であると仮定＝分散が一定

$$\pi_{kn} \sim \mathcal{N}(\mathbf{0}, V_k)$$

V_k は信号の性質（白色性・周期性）を表す実対称半正定値行列

基底信号の線形結合もまたガウス分布に従う

$$o_n | V, H \sim \mathcal{N}\left(\mathbf{0}, \sum_{k=1}^{K} V_k h_{kn}\right)$$

$X_n = o_n o_n^T \succeq 0$
$V_k \succeq 0$
$h_{kn} = \pi_{kn}^2 \geq 0$

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

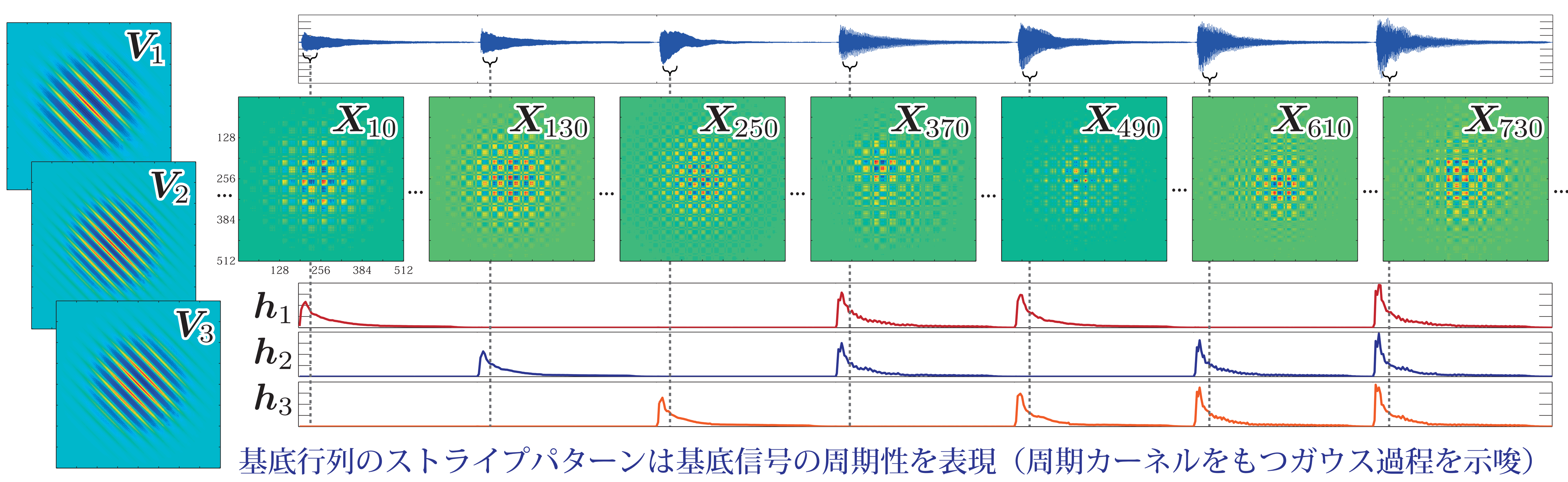
ガウス分布の対数尤度最大化 ＝ LDダイバージェンス最小化

時間領域でのウィナーフィルタリング
LD-PSDTFの結果 V, H に基づく観測信号 o_n の比例配分

$$\mathbb{E}[o_{kn}|o_n, V, H] = Y_{kn} Y_n^{-1} o_n \qquad Y_{kn} = V_k h_{kn}$$

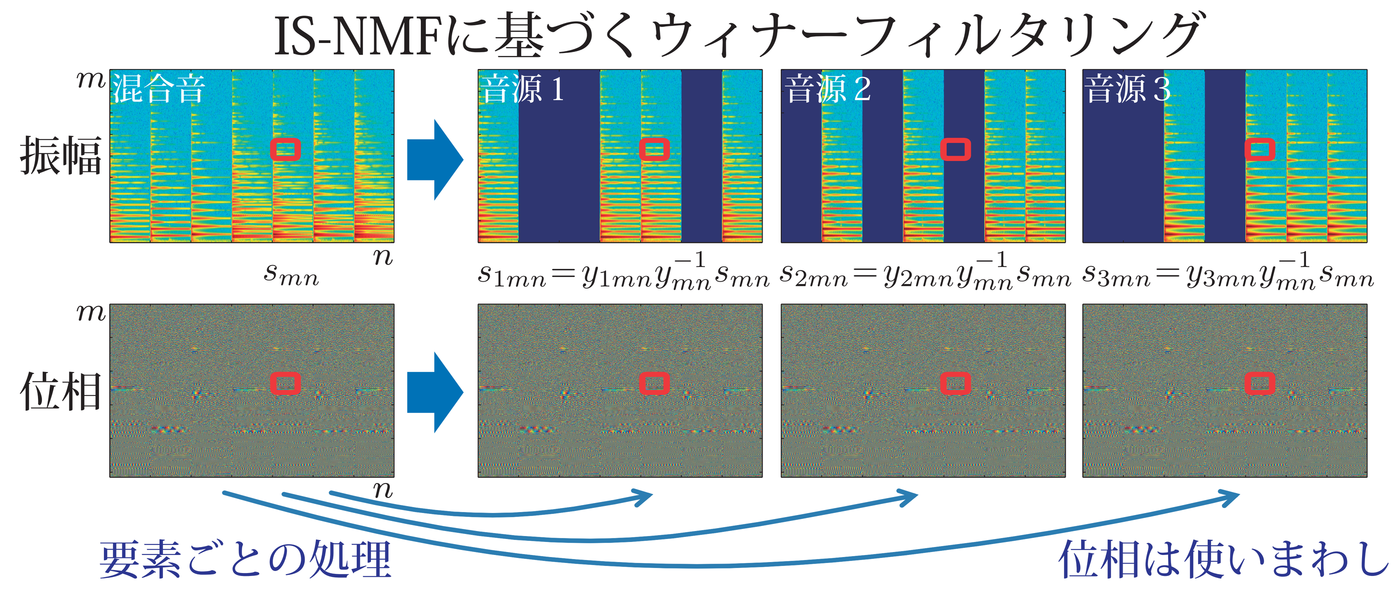
$$\mathbb{V}[o_{kn}|o_n, V, H] = Y_{kn} - Y_{kn} Y_n^{-1} Y_{kn} \qquad Y_n = \sum_k Y_{kn}$$

C4, E4, G4の音高からなる混合音の音源分離結果 (K=3)
IS-NMF: SDR 19.1dB, SIR 24.0dB, SAR 21.0dB
LD-PSDTF: SDR 23.0dB, SIR27.7dB, SAR 25.1dB



基底行列のストライプパターンは基底信号の周期性を表現（周期カーネルをもつガウス過程を示唆）

IS-NMFに基づくウィナーフィルタリング



時間領域LD-PSDTFと等価な周波数領域LD-PSDTFが存在 → 位相が適切に取り扱える！

観測スペクトルに対する音源スペクトルの瞬時混合モデル

$$s_n = \sum_{k=1}^{K} s_{kn}$$

F：離散フーリエ変換行列
$s_n = F o_n$, $s_{kn} = F o_{kn}$

周波数領域における基底行列
（複素エルミート半正定値行列）

$$s_n | V, H \sim \mathcal{N}\left(\mathbf{0}, \sum_{k=1}^{K} \left(F V_k F^H\right) h_{kn}\right)$$

V_k が巡回行列であれば対角化される → IS-NMFと等価

LD-PSDTFに基づくウィナーフィルタリング