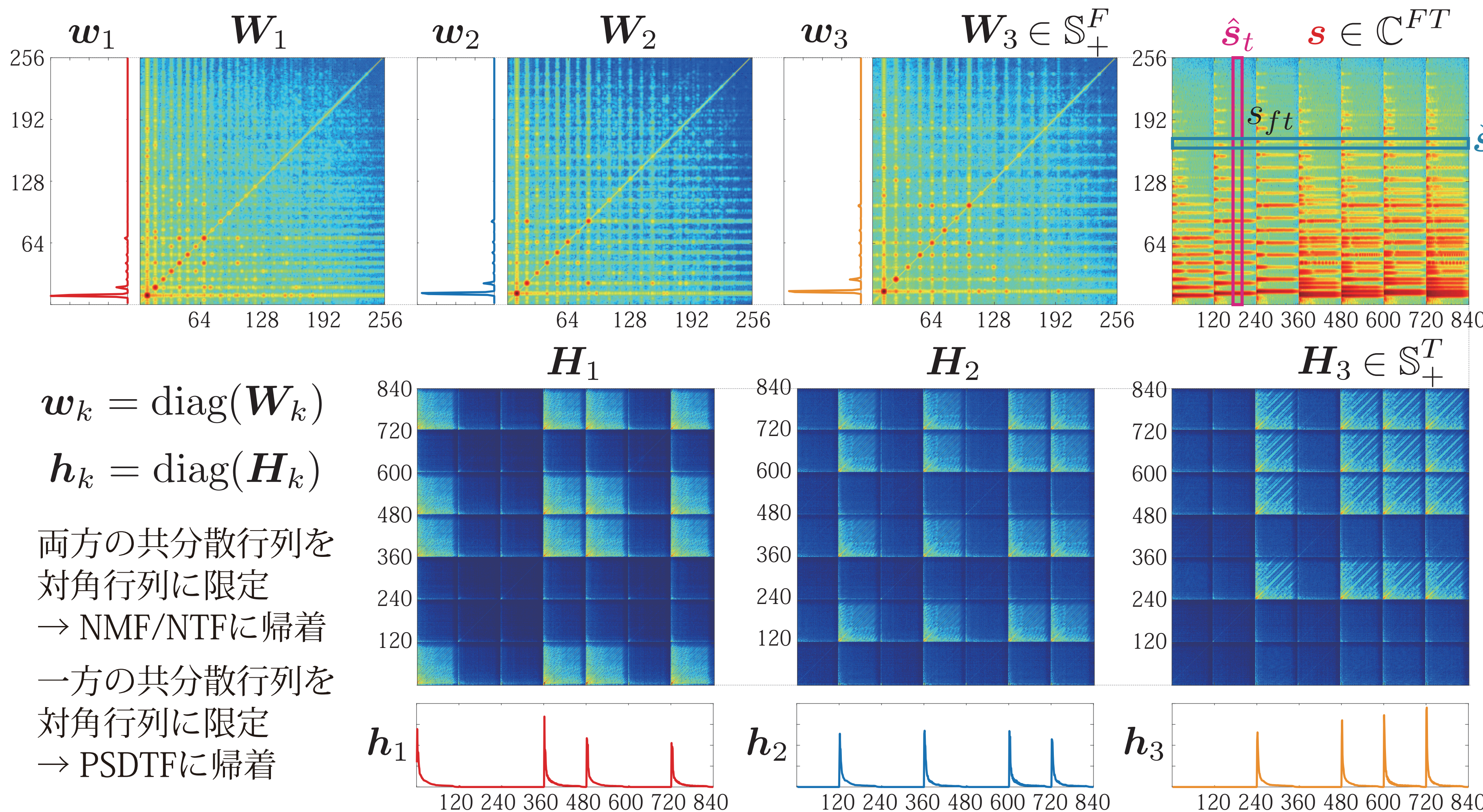


音楽音響信号解析のための ガンマ過程に基づく無限相関テンソル分解

吉井 和佳 (京都大学 / 理研 AIP)

提案法：相関テンソル分解 (Correlated Tensor Factorization: CTF) CP分解の拡張とも解釈可能

非負値行列分解 (NMF)・半正定値テンソル分解 (PSDTF)・非負値テンソル分解 (NTF) を特殊形に含む究極のテンソル分解法



$$\begin{aligned} x_{ft} &= s_{ft} s_{ft}^* = |s_{ft}|^2 \\ x_{ft} &\approx y_{ft} = \sum_{k=1}^K w_{kf} h_{kt} \\ \hat{\mathbf{x}}_t &= \text{diag}(\hat{\mathbf{X}}_t) \in \mathbb{R}_+^F \\ \hat{\mathbf{x}}_t &\approx \hat{\mathbf{y}}_t = \sum_{k=1}^K h_{kt} \mathbf{w}_k \\ \tilde{\mathbf{x}}_f &= \text{diag}(\tilde{\mathbf{X}}_f) \in \mathbb{R}_+^T \\ \tilde{\mathbf{x}}_f &\approx \tilde{\mathbf{y}}_f = \sum_{k=1}^K w_{kf} \mathbf{h}_k \\ \mathbf{x} &= \text{diag}(\mathbf{X}) \in \mathbb{R}_+^{FT} \\ \mathbf{x} &\approx \mathbf{y} = \sum_{k=1}^K \mathbf{w}_k \otimes \mathbf{h}_k \end{aligned}$$

IS or LDダイバージェンス最小化
→ (多変量)複素ガウス分布の
分散 (行列) の最尤推定

半正定値テンソル分解 (PSDTF)

$$\begin{aligned} \hat{\mathbf{X}}_t &= \hat{\mathbf{s}}_t \hat{\mathbf{s}}_t^H \in \mathbb{S}_+^F \\ \hat{\mathbf{X}}_t &\approx \hat{\mathbf{Y}}_t = \sum_{k=1}^K h_{kt} \mathbf{W}_k \\ \tilde{\mathbf{X}}_f &= \tilde{\mathbf{s}}_f \tilde{\mathbf{s}}_f^H \in \mathbb{S}_+^T \\ \tilde{\mathbf{X}}_f &\approx \tilde{\mathbf{Y}}_f = \sum_{k=1}^K w_{kf} \mathbf{H}_k \\ \mathbf{X} &= \mathbf{s} \mathbf{s}^H \in \mathbb{S}_+^{FT} \\ \mathbf{X} &\approx \mathbf{Y} = \sum_{k=1}^K \mathbf{W}_k \otimes \mathbf{H}_k \end{aligned}$$

相関テンソル分解 (CTF)

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

最小化すべき
コスト関数

$$\mathcal{C}_{\text{NMF}}(\mathbf{X}|\mathbf{Y}) = \sum_{t=1}^T \sum_{f=1}^F \mathcal{D}(x_{ft}|y_{ft})$$

すべての時間・周波数間は独立

Itakura-Saito (IS) ダイバージェンス
 $\mathcal{D}_{\text{IS}}(x|y) = -\log \frac{x}{y} + \frac{x}{y} - 1$

Kullback-Leibler (KL) ダイバージェンス
 $\mathcal{D}_{\text{KL}}(x|y) = x \log x - x \log y - x + y$

Majorization-
Minimization
アルゴリズム
に基づくパラ
メータ推定

$$\begin{aligned} p_{kf} &= \sum_{t=1}^T h_{kt} y_{ft}^{-1} \\ q_{kf} &= \sum_{t=1}^T h_{kt} x_{ft} y_{ft}^{-2} \\ w_{kf} &\leftarrow p_{kf}^{-1} \# (w_{kf} q_{kf} w_{kf}) = w_{kf} \left(\frac{q_{kf}}{p_{kf}} \right)^{\frac{1}{2}} \end{aligned}$$

$$\begin{aligned} r_{kt} &= \sum_{f=1}^F w_{kf} y_{ft}^{-1} \\ s_{kt} &= \sum_{f=1}^F w_{kf} x_{ft} y_{ft}^{-2} \\ h_{kt} &\leftarrow r_{kt}^{-1} \# (h_{kt} s_{kt} h_{kt}) = h_{kt} \left(\frac{s_{kt}}{r_{kt}} \right)^{\frac{1}{2}} \end{aligned}$$

Wiener Filtering
による音源分離
計算量

$$s_{ft}^{(k)} = y_{ft}^{(k)} y_{ft}^{-1} s_{ft}$$

$\mathcal{O}(KFT)$

時間周波数ごとに
独立に振幅を分配
(混合音の位相を
そのまま再利用)

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

$$\mathcal{C}_{\text{PSDTF}}(\mathbf{X}|\mathbf{Y}) = \sum_{t=1}^T \mathcal{D}(\hat{\mathbf{X}}_t|\hat{\mathbf{Y}}_t) \text{ or } \sum_{f=1}^F \mathcal{D}(\tilde{\mathbf{X}}_f|\tilde{\mathbf{Y}}_f)$$

周波数間は相関 (時刻間は独立) or 時刻間は相関 (周波数間は独立)

Log-Determinant (LD) ダイバージェンス

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

von Neumann (vN) ダイバージェンス

$$\mathcal{D}_{\text{vN}}(\mathbf{X}|\mathbf{Y}) = \text{tr}(\mathbf{X} \log \mathbf{X} - \mathbf{X} \log \mathbf{Y} - \mathbf{X} + \mathbf{Y})$$

$$\mathbf{P}_k = (\mathbf{I}_{F,F} \otimes \mathbf{1}_T^T) ((\mathbf{1}_{F,F} \otimes \mathbf{H}_k^T) \odot \mathbf{Y}^{-1}) (\mathbf{I}_{F,F} \otimes \mathbf{1}_T)$$

$$\mathbf{Q}_k = (\mathbf{I}_{F,F} \otimes \mathbf{1}_T^T) ((\mathbf{1}_{F,F} \otimes \mathbf{H}_k^T) \odot \mathbf{Y}^{-1} \mathbf{X} \mathbf{Y}^{-1}) (\mathbf{I}_{F,F} \otimes \mathbf{1}_T)$$

$$\mathbf{W}_k \leftarrow \mathbf{P}_k^{-1} \# (\mathbf{W}_k \mathbf{Q}_k \mathbf{W}_k) \quad \text{二つの半正定値行列の幾何平均による更新}$$

$$\mathbf{R}_k = (\mathbf{1}_F^T \otimes \mathbf{I}_{T,T}) ((\mathbf{W}_k^T \otimes \mathbf{1}_{T,T}) \odot \mathbf{Y}^{-1}) (\mathbf{1}_F \otimes \mathbf{I}_{T,T})$$

$$\mathbf{S}_k = (\mathbf{1}_F^T \otimes \mathbf{I}_{T,T}) ((\mathbf{W}_k^T \otimes \mathbf{1}_{T,T}) \odot \mathbf{Y}^{-1} \mathbf{X} \mathbf{Y}^{-1}) (\mathbf{1}_F \otimes \mathbf{I}_{T,T})$$

$$\mathbf{H}_k \leftarrow \mathbf{R}_k^{-1} \# (\mathbf{H}_k \mathbf{S}_k \mathbf{H}_k) \quad \text{二つの半正定値行列の幾何平均による更新}$$

$$\hat{\mathbf{s}}_t^{(k)} = \hat{\mathbf{Y}}_t^{(k)} \hat{\mathbf{Y}}_t^{-1} \hat{\mathbf{s}}_t \text{ or } \tilde{\mathbf{s}}_f^{(k)} = \tilde{\mathbf{Y}}_f^{(k)} \tilde{\mathbf{Y}}_f^{-1} \tilde{\mathbf{s}}_f$$

$\mathcal{O}(KF^3T)$ $\mathcal{O}(KFT^3)$

$$\mathbf{s}^{(k)} = \mathbf{Y}^{(k)} \mathbf{Y}^{-1} \mathbf{s}$$

$\mathcal{O}(KF^3T^3)$

時間周波数上の
相関を考慮しながら
音源複素スペクトロ
グラムを一挙に推定

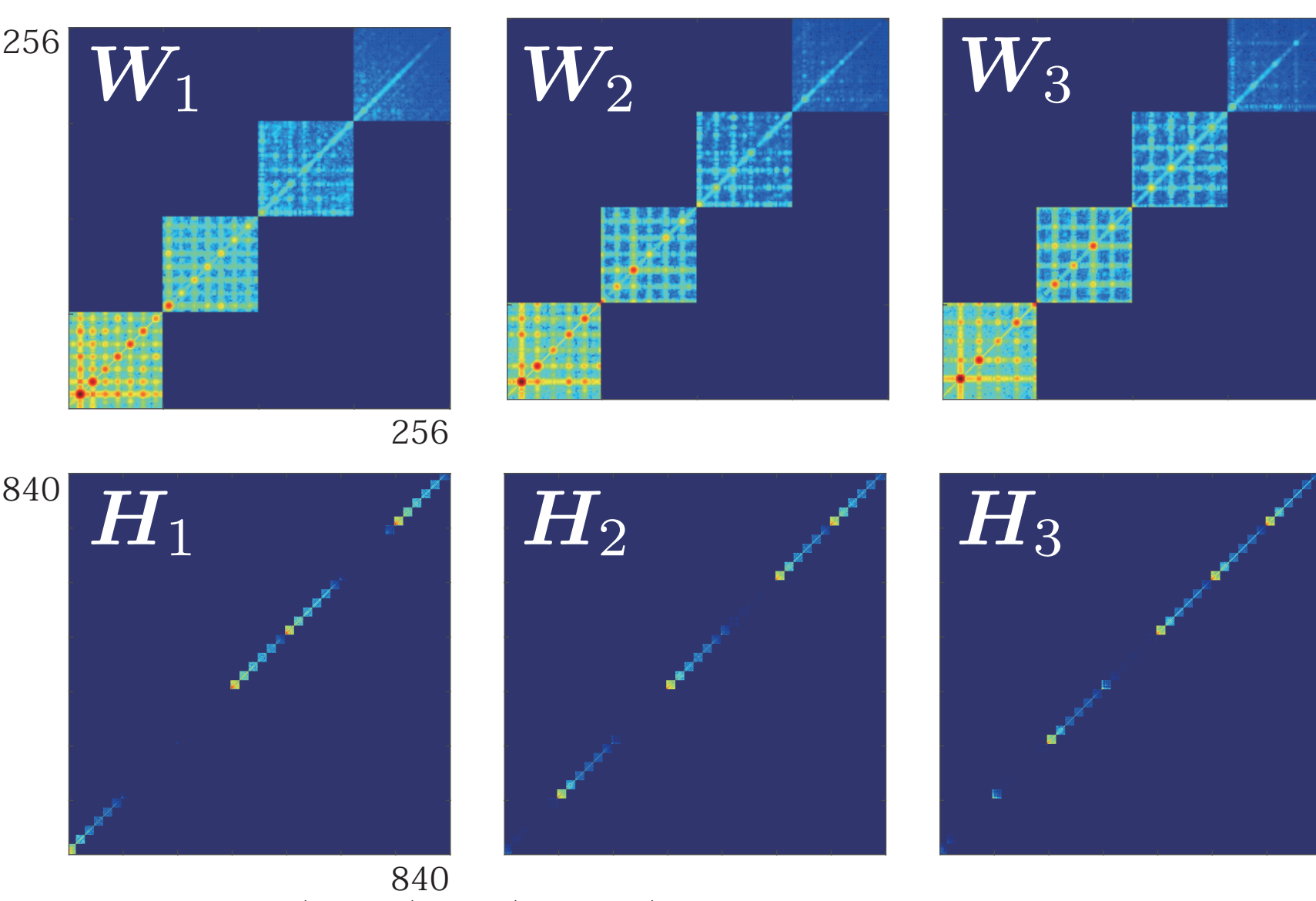
高速化のための近似手法と今後の展開

共分散行列のブロック対角化

周波数軸をPビンずつ、時間軸をQフレームずつに
分割し、各局所領域内に含まれる時間周波数ビン
間のみの相関を考慮 (異なる領域間は無相関)

本来は、周波数ビン・時間フレームともに
任意個数・サイズのクラスタリングでOK

$$\mathcal{O}(KFTP^2Q^2) \ll \mathcal{O}(KF^3T^3)$$



(P, Q) = (64, 20) の場合の推定結果

共分散行列の同時対角化

各WおよびHそれぞれについて、共分散行列を近似的に同時対角化
(Approximate Joint Diagonalization) できる空間を見つけたうえで、
(3個以上の厳密対角化は不可能)、その空間でIS-NMFを実行

もとの空間 (時間周波数平面) における確率モデル

$$\mathbf{S} \sim \mathcal{N}_c \left(\mathbf{0}, \sum_{k=1}^K \mathbf{W}_k \otimes \mathbf{H}_k \right)$$

線形写像A, Bによる空間変換

変換後の空間における確率モデル

$$\mathbf{A} \mathbf{S} \mathbf{B}^H \sim \mathcal{N}_c \left(\mathbf{0}, \sum_{k=1}^K \mathbf{A} \mathbf{W}_k \mathbf{A}^H \otimes \mathbf{B} \mathbf{H}_k \mathbf{B}^H \right)$$

統一的な確率的枠組みの下で、CTFパラメータおよび対角化パラ
メータを同時最適化 (NMFとAJDの反復実行) できる可能性が示唆

参考：二次統計量に基づくマルチチャネル音源分離, 独立ベクトル分析 (小野2011),
独立低ランク行列分析 (北村2016), 同期同時対角化 (澤田2017)

予備的評価と今後の課題

ピアノの単独音を用いて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
(16 [kHz], 1.2 [s] * 7 = 8.4 [s], F=256, T=840)

	IS-NMF	LD-PSDTF	LD-CTF (ブロック対角化版)
(P, Q)	(1, 1)	(256, 1)	(840, 1)
(128, 10)			(64, 20)
(32, 40)			
SDR	18.88	21.58	21.04
19.68		20.60	20.21
SIR	24.14	27.01	24.67
25.29		26.17	25.45
SAR	20.45	23.14	23.50
21.47		21.47	22.15

ブロック対角化近似を用いたLD-CTFはIS-NMFより
優れていたものの、LD-PSDTFには及ばなかった
→ 離れた倍音周波数間の強い相関を考慮する重要性
高速化の本命手法は同時対角化 (今後の課題)

LD-CTFは極めて初期値依存性が強いいため、IS-NMF
(これ自体局所解に陥りやすい) で初期化しておく
→ 経験的にはKL-NMFの方が初期値依存性が低い
vN-CTFの導出 (今後の課題)

他分野への応用 (例：ユーザ-アイテム-コンテキスト
などテンソルを用いた推薦システム) にも取り組む