MUSICを利用したマルチチャンネルNMFのパーミュテーション評価*

○太刀岡 勇気 (デンソーアイティーラボラトリ)

1 はじめに

混合音を分離する音源分離において、独立成分分析のような周波数ごとの音源分離では、音源のパーミュテーションを解決する必要がある。音源の性質に注目したモデル化 (隣接周波数間での差異に注目したモデル化 [3] やスペクトル包絡のモデル化 [4, 5]) と到来方向に基づくクラスタリング [1] が提案されている。

マルチチャンネルNMFは、音源基底によるスペクトル包絡の拘束により、前者を内包していると考えられるが、後者の到来方向については明示的な拘束を使っていない。実際、推定された空間相関行列の類似度と音源分離の性能には明確な関係がある [6]。ここでは、空間相関行列からステアリングベクトルを求めることでMUSIC法を適用するアルゴリズムに着目し [2]、MUSICスペクトルに基づいてパーミュテーションを解決する方法を提案する。提案法の有効性を確認するため、音楽データの分離実験を行った。

2 マルチチャンネルNMF

NMFでは観測行列 $\boldsymbol{X}$ を、基底行列 $\boldsymbol{T}$ とアクティベーション行列 $\boldsymbol{V}$ に因子分解する。さらに、MNMFは観測行列 $\boldsymbol{X}$ を4行列に因子分解する [7]。2行列 $\boldsymbol{H}$ と $\boldsymbol{Z}$ は、それぞれ空間相関行列とクラスター指示隠れ変数である。MNMFは K 個のスペクトル基底を L 個の音源に分解する際に、空間情報を使う。

2.1 定式化

周波数ビン i $(1 \leq i \leq I)$、時間フレーム j $(1 \leq j \leq J)$ の観測ベクトル $\boldsymbol{x}_{ij}$ を、$[x_1, \ldots, x_m, \ldots, x_M]_{ij}^{\top}$ とする。ここで、$^{\top}$ は転置、x_m は $m(1 \leq m \leq M)$ 番目のマイクで観測された短時間フーリエ変換 (STFT) の複素スペクトルである。これから、観測行列 $\boldsymbol{X} \in (\mathbb{C}^{M\times M})^{I\times J}$ の要素 i, j は以下のように表される。

$$\boldsymbol{X}_{ij} = \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}} = \begin{bmatrix} |x_1|^2 & \cdots & x_1 x_M^* \\ \vdots & \ddots & \vdots \\ x_M x_1^* & \cdots & |x_M|^2 \end{bmatrix}_{ij}, \quad (1)$$

ここで、* は複素共役、H はエルミート転置である。行列 $\boldsymbol{X}$ は階層的な行列であり、その要素 $\boldsymbol{X}_{ij}$ は $M \times M$ の複素半正定値エルミート行列となる。この行列 $\boldsymbol{X}$ は以下のように4行列 ($\boldsymbol{H}$, $\boldsymbol{Z}$, $\boldsymbol{T}$, $\boldsymbol{V}$) に分解される。

$$\boldsymbol{X} \cong \hat{\boldsymbol{X}} = [(\boldsymbol{H}\boldsymbol{Z}) \circ \boldsymbol{T}]\,\boldsymbol{V}, \quad (2)$$

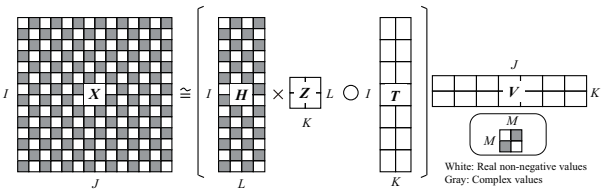

Fig. 1 An example of factorizing an observation matrix $\boldsymbol{X}$ into four matrices $\boldsymbol{H}$, $\boldsymbol{Z}$, $\boldsymbol{T}$, and $\boldsymbol{V}$ by the multi-channel NMF algorithm. ($I = J = 7$ and $K = L = M = 2$)

$\circ$ はアダマール積で、図1は、式 (2) を図示したものである。$\boldsymbol{H} \in (\mathbb{C}^{M\times M})^{I\times L}$ は、L 個の音源の空間情報を示す空間相関行列である。行列 $\boldsymbol{Z} \in \mathbb{R}^{L\times K}$ は、空間情報を各基底に関連付けるクラスター指示隠れ変数行列である。基底行列 $\boldsymbol{T} \in \mathbb{R}^{I\times K}$ は K 個の基底から構成され、$\boldsymbol{V} \in \mathbb{R}^{K\times J}$ は各基底のアクティベーションを表す。式 (2) の右辺は、以下のように表される。

$$\hat{\boldsymbol{X}}_{ij} = \sum_k \left[\sum_l \boldsymbol{H}_{il} z_{lk} \right] t_{ik} v_{kj}. \quad (3)$$

理想条件では、$\hat{\boldsymbol{X}}_{ij}$ を要素に持つ再構築された行列 $\hat{\boldsymbol{X}}$ は、元の観測行列 $\boldsymbol{X}$ に一致するが、一般的には、これらの行列は誤差により一致しない。NMFにおいて、$\boldsymbol{X}$ と $\hat{\boldsymbol{X}}$ の間の任意の距離を定義し、式 (2) の右辺の4行列は、この距離を最小化するように更新される。ここでは、板倉斎藤 (IS) 擬距離を使う。

3 MUSIC

MUSICスペクトルを空間相関行列から以下のように求める。空間相関行列 $\boldsymbol{H}_{il} \in \mathbb{C}^{M\times M}$ を固有値分解

$$\boldsymbol{H}_{il} = \boldsymbol{V}_{il}\boldsymbol{D}_{il}\boldsymbol{V}_{il}^{-1} \quad (4)$$

する。$\boldsymbol{D} \in \mathbb{R}^{M\times M}$ は ($\boldsymbol{H}$ がエルミート行列のため) 実数の固有値を対角成分に持つ対角行列で、降順にソートされている。$\boldsymbol{V} \in \mathbb{C}^{M\times M}$ は固有値に対応する固有ベクトルを列に並べたもので、互いに直交する。

到来方向を S 方向に離散化して、マイク間隔 d の直線アレイで平面波仮定すると、音源の各到来方向 $\theta_s (s \in \{1, ..., S\})$ に対応するステアリングベクトル $\boldsymbol{a}(i, \theta_s) = [a_1, ..., a_m, ..., a_M]^{\top}$ は

$$a_m = \exp\left[\jmath \left(m - \frac{M+1}{2} \right) \frac{2\pi d}{c} \varphi(i) \sin\theta_s \right] \quad (5)$$

*Evaluation of permutation problems in multi-channel NMF by using MUSIC. by TACHIOKA, Yuuki (Denso IT Laboratory)

となる。ここで φ は周波数ビンを周波数 [Hz] に変換する関数、c は音速である[1]。

$\boldsymbol{H}_{il}$ は 1 つの音源 l に対応する空間相関行列なので、最大固有値以外に対応する固有ベクトルは雑音部分空間を張る。これより、最大固有値で重みづけされた MUSIC スペクトルは

$$S_{il}(\theta_s) = \frac{\sqrt{D_{il}(1,1)}}{\boldsymbol{a}^{\mathrm{H}}(i,\theta_s)\boldsymbol{V}(:,2:M)\boldsymbol{V}^{\mathrm{H}}(:,2:M)\boldsymbol{a}(i,\theta_s)} \tag{6}$$

のように表される[2]。

音源のパワーが大きいところの MUSIC スペクトルのほうが小さいところよりも信頼性が高いと考えられるので、各音源に属するパワーで MUSIC スペクトルを除したパワー正規化された MUSIC スペクトル

$$S'_{i,l}(\theta_s) = \frac{S_{i,l}(\theta_s)}{\sum_j \sum_k z(l,k)t(i,k)v(k,j)} \tag{7}$$

を使うことも考えられる。

周波数全体での MUSIC スペクトルは

$$S_l(\theta_s) = \sum_{i=\varphi^{-1}(f_{min})}^{\varphi^{-1}(f_{max})} S_{il}(\theta_s) \tag{8}$$

で、MUSIC スペクトルの信頼性が高いと考えられる f_{min}〜f_{max}[Hz] までの周波数を用いる。

4 パーミュテーション評価

周波数全体での MUSIC スペクトルを音源ごとに比較することで、分離ができているか評価に使う。3〜7 の基準を使う際には MUSIC スペクトルを、その総和が 1 になるように正規化しておく。

$$\bar{S}_\cdot(\theta_s) = \frac{S_\cdot(\theta_s)}{\sum_s S_\cdot(\theta_s)} \tag{9}$$

4.1 最大値 (max) 基準

音源 l に対するピーク

$$\phi_l = \arg\max_{\theta_s} S_l(\theta_s) \tag{10}$$

を算出し、ピークをとるインデクスの距離で分離度を評価する。

$$\sum_l \sum_{l'>l} |\phi_l - \phi_{l'}| \tag{11}$$

4.2 内積 (IP) 基準

スペクトルの類似度を内積で測る。

$$\sum_l \sum_{l'>l} -\frac{\boldsymbol{S}_l^\top \boldsymbol{S}_{l'}}{|\boldsymbol{S}_l||\boldsymbol{S}_{l'}|} \tag{12}$$

$\boldsymbol{S}_\cdot = [S_\cdot(\theta_1), ..., S_\cdot(\theta_S)]^\top$ である。$\boldsymbol{S} > 0$ なので、内積は必ず 0 以上となる。

[1]ただし等間隔であれば、マイク間隔 d は不明でもよい。d を仮定して実際には d' であった場合でも、$a'_m(\theta_s) = a_m(\theta_s)^{d'/d}$ となるだけなので、MUSIC スペクトルの概形は変わらない。

[2]$\boldsymbol{V}$ に関するインデックス il は煩雑になるので省略した。

4.3 二乗誤差 (SE) 基準

スペクトル間の距離を 2 乗誤差

$$d_{SE}(p_1,p_2) = \sum_\theta |p_1(\theta) - p_2(\theta)|^2 \tag{13}$$

で測り、$\sum_l \sum_{l>l} d_{SE}\left(\bar{S}_l, \bar{S}_{l'}\right)$ で評価する。

4.4 重なり面積 (Overlapping area; OA) 基準

正規化されたスペクトル間の重なり面積

$$d_{CA}(p_1,p_2) = \sum_\theta -\min\left(p_1(\theta), p_2(\theta)\right) \tag{14}$$

で測り、$\sum_l \sum_{l'>l} d_{CA}\left(\bar{S}_l, \bar{S}_{l'}\right)$ で評価する。

4.5 Kullback-Leibler divergence (KLD) 基準

スペクトル間の距離を KLD

$$d_{KL}(p_1,p_2) = \sum_\theta p_1(\theta) \log\left(\frac{p_1(\theta)}{p_2(\theta)}\right) \tag{15}$$

で測り、$\frac{1}{2}\sum_l \sum_{l'\neq l} d_{KL}\left(\bar{S}_l, \bar{S}_{l'}\right)$ で評価する。

4.6 Itakura-Saito divergence (ISD) 基準

スペクトル同士の距離を測るのによいとされている ISD

$$d_{ISD}(p_1,p_2) = \sum_\theta \left[\frac{p_1(\theta)}{p_2(\theta)} - \log\left(\frac{p_1(\theta)}{p_2(\theta)}\right) - 1\right] \tag{16}$$

で測り、$\frac{1}{2}\sum_l \sum_{l'\neq l} d_{IS}\left(\bar{S}_l, \bar{S}_{l'}\right)$ で評価する。

4.7 Density power divergence (DPD) 基準

効率性とロバスト性を調整できる位相回復の分野で使われている DPD[8]

$$\begin{aligned} d_{DPD}(p_1,p_2) = \sum_\theta \Bigg[&\frac{1}{\gamma}\left(p_1(\theta)^\gamma - p_2(\theta)^\gamma\right) \\ &-\frac{1}{1+\gamma}\left(p_1(\theta)^{1+\gamma} - p_2(\theta)^{1+\gamma}\right)\Bigg] \end{aligned} \tag{17}$$

で測り、$\frac{1}{2}\sum_l \sum_{l'\neq l} d_{DPD}\left(\bar{S}_l, \bar{S}_{l'}\right)$ で評価する。ただし $\gamma = 0.2$ とした。

5 パーミュテーション解決

周波数全体での MUSIC スペクトルと各周波数ビンでの MUSIC スペクトルを以下の基準で比較することで、パーミュテーション解決を行う。以下 $\mathcal{P} = \{p_1, p_2, \ldots, p_{L!}\}$ を音源 $1, \ldots, L$ のパーミュテーションとする。例えば $L = 2$ の場合、$\mathcal{P} = \{p_1, p_2\} = \{(1,2),(2,1)\}$ である。

図 2 のように、全体のスペクトル $\sum_i S_{i,l}$ と個別の周波数ビン (ここでは $f = 10$) のスペクトル $S_{10,l}$ を比較する。$S_{10,1}$ のピーク位置は、$\sum_i S_{i,1}$ のピーク位置よりも、$\sum_i S_{i,2}$ のピーク位置に近く、ここにおいてパーミュテーションが起こっていると判断できる。

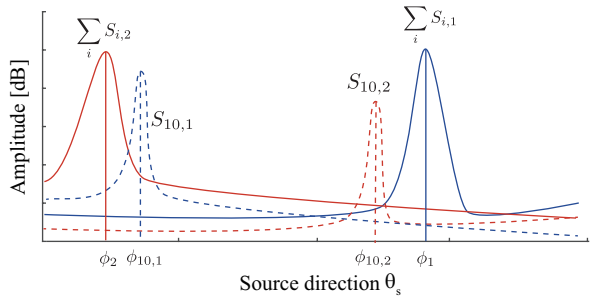

Fig. 2　Finding permutation at each frequency bin.

5.1　最大値 (max) 基準

全体でのピーク ϕ_l と各周波数ビンでのピーク

$$\phi_{il} = \arg\max_{\theta_s} S_{il}(\theta_s) \tag{18}$$

を算出し、ピーク同士がなるべく近くなるようにパーミュテーション解決を行う。

$$\min_m \sum_l |\phi_l - \phi_{ip_m(l)}| \tag{19}$$

5.2　内積 (IP) 基準

内積においても同様

$$\min_m -\frac{\boldsymbol{S}_l^\top \boldsymbol{S}_{ip_m(l)}}{|\boldsymbol{S}_l||\boldsymbol{S}_{ip_m(l)}|} \tag{20}$$

である。

5.3　その他の基準

その他の基準に関しても、$\min_m d.\left(\bar{S}_l, \bar{S}_{ip_m(l)}\right)$ とすることで同様にパーミュテーション解決できる。

6　実験

6.1　実験条件

$f_{min} = 500$、$f_{max} = \frac{c}{2d} = 4250$[Hz] とした。表 3 に示す更新回数においてパーミュテーション解決を

Table 1　Setup for music source separation.

Sampling frequency	16 kHz
Frame size and shift	1024 and 256
# bases	30
# iterations	500

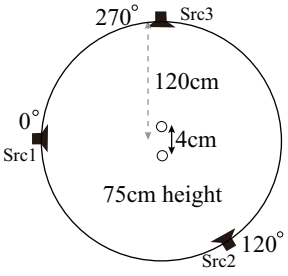

Fig. 3　Source locations.

Table 2　Source music

ID	Author/Song	Snip	Part
1	Bearlin Roads	85-99 (14 sec)	piano ambient vocals
2	Another Dreamer The Ones We Love	69-94 (25 sec)	drums vocals guitar
3	Fort Minor Remember The Name	54-78 (24 sec)	drums vocals violin+synth
4	– Ultimate Nz Tour	43-61 (18 sec)	drums guitar synth

Table 3　Schedule of permutation alignment.

sID	Schedule	sID	Schedule
1	40,45,50,55,60	3	100,150
2	50,100,150	4	50,100,150,200

行った。音源分離のパラメータは、表 1 の通り設定した。

図 3 に音源の配置を示す。表 2 に示す 3 音源 ($L = 3$) の音楽データに、それぞれ正面 (0 度)、−90 度、60 度方向からのインパルス応答 (インパルス応答長 300) を畳み込み、4cm 間隔の 2 マイク ($M = 2$) で分離した結果を信号対歪比 (SDR)[dB] で評価した。

10 の異なる乱数シードを設定し、$\boldsymbol{T}$、$\boldsymbol{V}$、$\boldsymbol{Z}$ には乱数を $\boldsymbol{H}$ は単位行列として実験した。従来研究に倣い、最初の 20 回は $\boldsymbol{H}$ を固定で乗法更新則を適用した。

6.2　MUSIC スペクトルの考察

10 の分離結果のうち、SDR 最良 (14.63[dB]) の場合 (best) と最悪 (1.63[dB]) の場合 (worst) で、最終的に得られた $\boldsymbol{H}$ を対象として、提案法により求めた MUSIC スペクトルを、図 4 に示す。明らかに worst の場合は、best の場合に比べてピークが重なっている。これから、MNMF においても、パーミュテーション解決がうまくいかない場合があることがうかがえる。例えば 4.1 の基準に従えば、左はピーク位置が音源 1,2,3 について $\{0.4, -1, -1.25\}$ なので、それぞれの絶対値の総和は $|0.4-(-1)|+|-1-(-1.25)|+|0.4-(-1.25)| = 3.3$ である。これに対して、右はそれぞれ $\{0.25, 0.25, -1\}$ なので、$|0.25-0.25|+|0.25-(-1)|+|0.25-(-1)| = 2.5$ で左の方が指標が大きい。

6.3　分離性能の予測

表 4 に SDR と 4 節での指標の相関係数を示す。曲 1 を除くと相関はあまり高くなく、SDR の直接的な予測はパーミュテーションの程度だけからは難しいことが分かった。

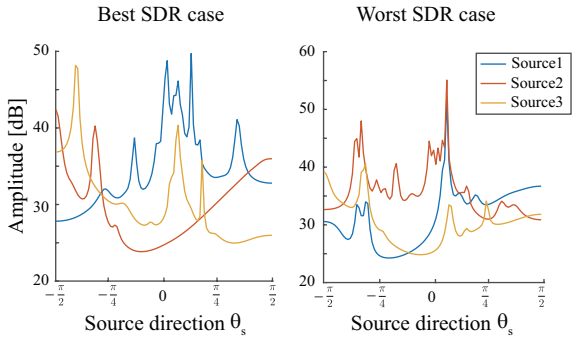

Fig. 4　MUSIC spectrum of $\boldsymbol{H}$ corresponding to the best and worst SDR among results starting from 10 random initializations.

Table 4　Correlation coefficients between SDR[dB] and criteria

ID	d_{IS}	max	IP	DPD	KL	IS	SE	OA
1	-0.56	0.89	0.89	0.80	0.77	-0.29	0.64	0.91
2	-0.15	-0.14	-0.41	-0.58	-0.57	-0.35	-0.79	-0.47
3	-0.83	0.36	0.56	0.46	0.50	-0.12	0.01	0.51
4	-0.62	0.66	0.55	0.49	0.45	-0.59	0.26	0.51

6.4　分離実験の結果

分離実験の結果を表 5 に示す。表中「proposed (Eq. (6))」は通常の MUSIC スペクトルを使った場合、「proposed (Eq. (7))」はパワーで正規化した MUSIC スペクトルを使った場合である。このように楽曲により差異があるが、平均的にみるとすべての場合で性能の改善がみられる。最良の基準は楽曲によって異なるが、max か IP が最も安定していることがわかる。

7　まとめと今後の課題

音源分離の際のパーミュテーション解決を目的として、空間相関行列の MUSIC スペクトルを用いたパーミュテーション解決法を提案した。音楽データの分離実験により、分離結果が悪い場合にはパーミュテーション解決が十分でないことを明らかにし、これを用いてパーミュテーション解決を音源分離の繰り返し最適化中に行うことで音源分離性能が向上することが示された。更新回数の最適化が今後の課題である。

Table 5　Average SDR[dB] of each song.

Criterion	sID	Music ID 1	2	3	4	Avg.
baseline	-	2.18	5.55	3.07	7.52	4.58
Based on Eq. (6)						
max	1	2.64	7.24	2.59	**9.08**	5.39
max	2	2.35	6.77	**3.08**	7.00	4.80
max	3	2.18	4.88	2.88	6.92	4.22
max	4	2.40	5.67	**3.08**	5.94	4.27
IP	1	1.60	6.63	2.42	8.17	4.70
IP	2	2.51	7.57	2.74	8.75	5.39
IP	4	**2.78**	**7.63**	2.75	8.48	**5.41**
DPD	1	2.09	6.89	2.95	8.79	5.18
DPD	2	1.28	7.25	2.92	8.48	4.98
DPD	4	1.38	7.21	2.89	7.78	4.81
KLD	1	1.81	5.76	2.61	8.72	4.72
KLD	2	1.15	6.69	2.99	8.31	4.78
KLD	4	1.18	6.63	2.94	8.40	4.79
ISD	1	1.82	5.85	2.69	8.62	4.75
ISD	2	1.11	6.62	2.82	8.26	4.70
ISD	4	1.30	6.47	2.85	8.46	4.77
SE	1	2.03	6.06	2.67	8.52	4.82
SE	2	1.24	6.58	2.81	8.35	4.75
SE	4	1.22	6.60	2.95	8.08	4.71
OA	1	1.34	6.72	2.65	8.46	4.79
OA	2	2.01	7.29	2.81	8.75	5.22
OA	4	1.85	7.23	2.82	8.86	5.19
Based on Eq. (7)						
max	1	2.85	6.31	2.53	8.13	4.95
max	2	**3.51**	7.38	**3.57**	7.88	**5.58**
IP	1	1.93	**8.11**	2.45	8.81	5.32
IP	2	2.18	5.89	2.98	9.12	5.04
DPD	1	1.47	5.44	2.47	9.37	4.69
DPD	2	1.47	6.06	2.92	9.48	4.98
KLD	1	2.00	5.57	2.35	9.42	4.84
KLD	2	1.26	6.05	2.80	9.45	4.89
ISD	1	1.36	5.41	2.35	8.59	4.43
ISD	2	1.41	5.87	2.74	**9.78**	4.95
SE	1	1.03	6.63	2.79	9.14	4.90
SE	2	1.84	6.36	2.81	9.21	5.05
OA	1	2.43	7.21	2.32	8.51	5.12
OA	2	1.57	5.78	3.30	9.09	4.93

参考文献

[1] H. Sawada, S. Araki, and S. Makino, “Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment,” IEEE Trans on ASLP, **19**, 516–527 (2011).

[2] N. Mitianoudis and M. Davies, “Permutation alignment for frequency domain ICA using subspace beamforming methods,” Proc of ICA, pp.127–132 (2004).

[3] R. Mazur and A. Mertins, “An approach for solving the permutation problem of convolutive blind source separation based on statistical signal models,” IEEE Trans on ASLP, **17**, 117–126 (2009).

[4] S. Saito, K. Oishi, and T. Furukawa, “Convolutive blind source separation using an iterative least-squares algorithm for non-orthogonal approximate joint diagonalization,” IEEE Trans on ASLP, **23**, 2434–2448 (2015).

[5] A. Sarmiento, I. Duran-Diaz, A. Cichocki, and S. Cruces, “A contrast function based on generalised divergences for solving the permutation problem in convolved speech mixtures,” IEEE Trans on ASLP, **23**, 1713–1726 (2015).

[6] 浦本昂伸，太刀岡勇気，成田知宏，三浦伊織，上ノ原進吾，古家賢一，“マルチチャネル非負値行列因子分解を用いたブラインド音源分離のためのチャネル数増加に伴う逐次的初期化法,” 信学論 D, **J101-D**, 569–577 (2018).

[7] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, “Multichannel extensions of non-negative matrix factorization with complex-valued data,” IEEE Trans on ASLP, **21**, 971–982 (2013).

[8] 塩谷浩之，郷原一寿，“位相回復 –計算アルゴリズム–,” 計測と制御, **50**, 332–337 (2011).