

ブラインド音源分離における 2×2 分離行列の自由度について*

☆下川凜, 松本和樹, 矢田部浩平 (農工大)

1 はじめに

優決定ブラインド音源分離では, Projection Back (PB) を後処理として用いることで分離行列の周波数特性の歪みを修正する. また, PB を制約として組み込んで分離行列を最適化する手法も提案されており, この場合は通常の音源分離アルゴリズムよりも探索空間が小さくなる [1-3]. 本来, 2×2 分離行列の自由度は実部虚部合わせて 8 であるのに対し, PB を制約として取り入れると 4 まで削減できる.

本稿では, PB 適用後の 2×2 分離行列を振幅と位相を用いた極座標形式で表し, 振幅に制約を追加することで, 3 および 2 自由度へと削減できる余地があるかを検討する. また, 全探索を用いた実験により自由度の削減がどの程度性能に影響を与えるか確認した.

2 音源分離の分離行列の自由度

本稿ではマイク数と音源数が同じ状況を仮定する. 時間周波数領域における K 個の音源信号を $\mathbf{s}[f, t] \in \mathbb{C}^K$, K 個のマイクの観測信号を $\mathbf{x}[f, t] \in \mathbb{C}^K$ と表す. 観測信号が混合行列 $\mathbf{H}[f] \in \mathbb{C}^{K \times K}$ を用いて

$$\mathbf{x}[f, t] = \mathbf{H}[f] \mathbf{s}[f, t] \quad (1)$$

で得られると仮定すると, 分離信号 $\mathbf{y}[f, t]$ は, 混合行列の逆に対応する分離行列 $\mathbf{W}[f] \in \mathbb{C}^{K \times K}$ を用いて

$$\mathbf{y}[f, t] = \mathbf{W}[f] \mathbf{x}[f, t] \in \mathbb{C}^K \quad (2)$$

と求められる.

最適化アルゴリズムの出力として得られる分離行列 $\mathbf{W}$ や分離信号 $\mathbf{y}$ は周波数特性が歪んでいる場合が多い. これを解消する後処理として PB がよく用いられる. 本稿ではマイク 1 を参照マイクロホンとして用いた PB を考える. このとき PB 行列 $\mathbf{D}[f] \in \mathbb{C}^{K \times K}$ は, 混合行列 $(h_{i,j}[f])_{i=1,j=1}^{K,K} = (\mathbf{W}[f])^{-1}$ の 1 行目を対角に並べることで

$$\mathbf{D}[f] = \begin{bmatrix} h_{1,1}[f] & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & h_{1,K}[f] \end{bmatrix} \quad (3)$$

と表せる. この PB 行列 $\mathbf{D}[f]$ を用いて, PB 適用後の分離行列 $\mathbf{W}^{(\text{PB})}[f]$ は

$$\mathbf{W}^{(\text{PB})}[f] = \mathbf{D}[f] \mathbf{W}[f] \quad (4)$$

のように求められる. PB 後の分離行列 $\mathbf{W}$ は

$$\sum_{k=1}^K \mathbf{W}_{kj}[f] = \begin{cases} 1 & (j=1) \\ 0 & (j \neq 1) \end{cases} \quad (5)$$

を満たすことが知られており [1, 2], 本稿では式 (5) を PB 制約と呼ぶ. 以降, 周波数のインデックスは省略する. この分離行列は, PB 制約を用いることで分離行列をパラメータ $a, b \in \mathbb{C}$ を用いて

$$\mathbf{W} = \begin{bmatrix} 0.5 + a & b \\ 0.5 - a & -b \end{bmatrix} \quad (6)$$

と表現できる. 通常の 2×2 の分離行列は 4 つの要素を持ち, その自由度は実部と虚部合わせて 8 である. 一方, 式 (6) では 2 つの複素数のパラメータ a, b で表現されていることから, PB 制約によって自由度は 4 まで削減できることが分かる.

3 極座標を用いた分離行列の表現

本章では, 分離行列を極座標形式のパラメータで表現し, 振幅に制約を加えることで, 自由度を 2 まで削減することを考える. まずは, 音源の混合過程を平面波モデルと仮定する場合, 各周波数における混合行列の逆行列は PB 制約を満たし, さらに自由度が 4 であることを示す.

マイク 1 を基準としたときの観測信号の混合行列は, $a_1, a_2 \in \mathbb{R}_+$ を振幅, $\phi_1, \phi_2 \in [0, 2\pi)$ を位相とすると, 極座標形式を用いて

$$\mathbf{H} = \begin{bmatrix} 1 & 1 \\ a_1 e^{i\phi_1} & a_2 e^{i\phi_2} \end{bmatrix} \quad (7)$$

と表すことができる. 分離行列は $\mathbf{H}$ の逆行列であり,

$$\mathbf{H}^{-1} = \mathbf{W} = \frac{1}{a_1 e^{i\phi_1} - a_2 e^{i\phi_2}} \begin{bmatrix} -a_2 e^{i\phi_2} & 1 \\ a_1 e^{i\phi_1} & -1 \end{bmatrix} \quad (8)$$

と書ける. この行列を列ごとに縦に足すと, 混合行列の逆行列は式 (5) の PB 制約を満たすことが確認できる. また, 分離行列 (8) に注目すると, a_1, a_2, ϕ_1, ϕ_2 の 4 変数で表現されていることから, 極座標形式における分離行列の自由度は 4 であり, これは式 (6) に示したパラメータ表現の自由度と一致している.

自由度 4 から, 更なる削減の余地がないかを考える. 式 (8) において振幅が同一 (i.e., $a_1 = a_2$) であると仮

*On the degree of freedom of 2×2 separation matrices in blind source separation. By Rin SHIMOKAWA, Kazuki MATSUMOTO and Kohei YATABE (Tokyo University of Agriculture and Technology).

表-1 各手法ごとの制約と自由度

	a_1	a_2	ϕ_1	ϕ_2	自由度
Ora	○	○	○	○	4
Ora+PB	○	○	○	○	4
Amp+Ph	共有	共有	○	○	3
Ph	1で固定	1で固定	○	○	2

定する場合は、振幅の自由度を1削減できるため、3自由度で分離行列を表現できる。さらに、 $a_1 = a_2 = 1$ と仮定すると自由度は位相のみとなり、分離行列の自由度を2に削減できる。

4 自由度削減の影響評価

自由度の削減に対する分離性能への影響を検証するために実験を行った。比較手法として、真の音源信号を基準に最小二乗法で分離行列を求めたオラクルな方法(Ora), それに対してPBを適用した行列(Ora+PB)を分離行列とする手法を利用した。ここで、真の音源信号は $\tilde{s}_k = h_{1,k}s_k$ ($n = 1, 2$) とした。これはマイク1で観測される n 番目の音源由来の成分である。自由度を削減した手法としては、共通の振幅と2つの位相の3変数で全探索をした手法(Amp+Ph)と、振幅を1とし、2つの位相の全探索をした手法(Ph)を用いた。全探索における変数の範囲は、 $a \in [0.5, 2]$, $\phi_1, \phi_2 \in [0, 2\pi)$ とし、目的関数は時間周波数領域における音源信号との誤差のL2ノルムとした。それぞれの手法の制約や自由度は表-1の通りである。観測信号は、SiSEC2011のdev1データセットより作成した2音源2マイクの混合音を用いた。混合音には残響時間が(a) 130 ms, (b) 250 msの2種類含まれており、それぞれ12組ずつ計24音源を利用した。さらに、窓長の変化による性能低下の幅を確認するため、各残響時間において窓長を(i) 512, (ii) 1024, (iii) 2048 サンプルと変化させたときの性能についても評価した。サンプリング周波数16 kHz, 分離性能はSDRiで評価した。

残響時間と窓長の変化による各手法の分離性能を図-1に示す。結果より、自由度が少ない手法ほど、分離性能が低いことが読み取れる。Amp+Phは振幅を共有しているため、マイク間の振幅差を考慮できておらず、またPhは1で固定しているため、振幅の自由度が全くない。このことから、高性能な分離のためには振幅に自由度をもたせることが必要であることが確認された。

また、同じ手法内で比較すると、残響時間が長い、または、窓長が短いほど性能が全体的に低下する傾向があることが読み取れる。この結果は、残響時間に対して窓長が短いと、式(1)のモデルが成立しなくなるためだと考えられる。今回の実験では、条件(a-iii)のとき、残響時間に対する窓長が最も長く、Ora

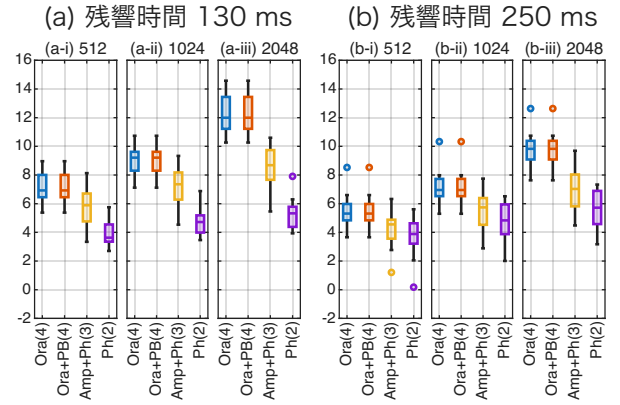


図-1 残響時間と窓長の変化によるSDRiの比較。()内の数字はその手法における分離行列の自由度を示す。(a)が残響が130 ms, (b)が250 msの場合。各図の上は窓長のサンプル数を表す。

やOra+PBは、もっとも高い性能を達成する。しかし、悪い条件でモデルが成り立たなくなるにつれ、性能は低下していく。それに対して、自由度が少ないAmp+Ph, Phも同様、条件が悪いほど性能は下がるが、その下がり幅は4自由度のOraやOra+PBよりも少ない。実際、条件(a-iii)のとき、OraとPhの中央値の性能差は7 dB程度あるが、条件(b-i)のとき、性能差は1 dB程度まで縮まっている。これより、式(1)のモデルが十分に成り立つ環境であれば、自由度が多いとより良い性能が得られる。しかし、モデルが成り立たなくなるにつれ、多くの自由度を持っても高い性能を達成するのは困難になる。結果として、そのような悪条件下では自由度による分離性能への影響は小さくなると考えられる。

5 むすび

本稿では2音源2マイクにおける分離行列の自由度についてPB制約を用いて検討した。極座標形式を用いた定式化より分離行列の自由度は4必要であるが、振幅の条件を追加することで自由度を削減した。比較実験では自由度が多いほど、高い性能が得られたが、残響時間が長い、窓長が短いといった式(1)のモデルへ影響が大きい程、自由度に対する性能差が小さくなることがわかった。今後は、今回得られた知見を音源分離のアルゴリズムの改善に役立てたい。

参考文献

- [1] Y. Kuriki, T. Nakashima and N. Ono, "Direct update of back-projected demixing matrices in blind source separation," *Proc. EUSIPCO*, pp. 922–926 (2024).
- [2] K. Matsumoto and K. Yatabe, "Determined blind source separation using metric projection and proximity operator of log-det function under projection-back constraint," *Proc. SampTA*, pp. 1–4 (2025).
- [3] K. Koiso, T. Nakashima and N. Ono, "Prior-guided source separation with direct update of back-projected demixing vectors," *Proc. APSIPA ASC*, pp. 166–171 (2025).