第32回信号処理シンポジウム
2017年11月8日～10日（盛岡）

空間事前情報を用いた独立低ランク行列分析

ILRMA incorporating with a priori spatial information

三井祥幹 † 高宗典玄 † 北村大地 † 猿渡洋 † 高橋祐 ‡ 近藤多伸 ‡
† 東京大学 ‡ ヤマハ株式会社

Yoshiki Mitsui† Norihiro Takamune† Daichi Kitamura†
Hiroshi Saruwatari† Yu Takahashi‡ Kazunobu Kondo‡
†The University of Tokyo ‡Yamaha Corporation

アブストラクト 本稿では，近年提案されたブラインド音源分離手法の独立低ランク行列分析（ILRMA）に対して空間的な事前情報を与える応用を提案する．従来の ILRMA による音源分離では，特に音声信号の分離において，音源モデルの悪い局所解や，周波数ビン間のブロックパーミュテーション不整合に起因し，十分な音源分離性能を得られないことがある．一方で，音源の到来方位が既知である場合は，特定方位の信号を強調するビームフォーマを ILRMA の正則化として新たに利用することで，前述の問題を防止できる．本手法の最適化では，従来のアルゴリズムを拡張した反復更新式を新たに導出している．実験的な評価により，本手法の有効性を示す．

1 はじめに

音源分離とは，複数の音源が混合された観測信号から，混合以前の信号を推定し復元する技術を指す．音源の空間的な配置等が未知の場合，音源間の統計的独立性を利用する多チャネル優決定音源分離手法として，周波数ビンごとに分離系（線形分離行列）を推定する周波数領域独立成分分析（frequency-domain independent component analysis: FDICA） [1] に基づく手法が多数検討されている．FDICA は周波数ビンごとの分離信号の順序が定まらない問題（パーミュテーション問題）を抱えており，種々の解決法 [2], [3] が検討されている．近年では，周波数ビン同士の間に高次の相関が存在する音源モデルを利用した独立ベクトル分析（independent vector analysis: IVA） [4] やその改良手法 [5]，非負値行列因子分解（nonnegative matrix factorization: NMF） [6] に基づく音源モデルを利用した独立低ランク行列分析（independent low-rank matrix analysis: ILRMA） [7], [8] など，音源モデルを改善することで自動的にパーミュテーション問題を解決する研究が進められており，特に ILRMA は FDICA や IVA と比較し高い音源分離性能を発揮することが報告されている．さらに，ベイズ型ポストフィルタを用いて周波数ビンごとの適応的なスパース正則化を与えるスパース ILRMA [9] も提案されている．このような音源モデルの改善により，特に隣接する周波数ビン間でのパーミュテーション問題は解決されつつあるが，まとまった帯域ブロックごとにパーミュテーション不整合が生じるブロックパーミュテーション問題 [10] や，NMF による音源モデルが適切に学習されないことにより悪い局所解に陥る問題などがこれまで報告されている．

一方で，マイクロホンアレイや音源の空間的な配置が既知の場合，人工的に指向性を制御し特定の方位より到来する音波を強調するビームフォーマが広く利用されている．音響信号処理において，空間モデルのみから構成される固定ビームフォーマをそのまま利用することは室内残響等の観点から好ましくないが，FDICA と固定ビームフォーマを融合させ，パーミュテーションの解決とより良い解への高速な誘導を同時に達成する手法 [3], [11]–[13] はこれまで多数検討されており，音源到来方位の事前情報が高精度な音源分離に有効であることが示されている．この他，IVA と空間事前情報に基づくバイナリマスキングを組み合わせた手法 [14] などが提案されている．

FDICA や IVA において，従来は勾配法に基づく反復最適化法が広く利用されていたが，近年では高速かつハイパーパラメタの不要な反復最適化法である iterative projection（IP） [15] が提案されており，ILRMA の分離行列最適化においても IP が利用されている [7]．しかし，ビームフォーマによる最適化を用いた FDICA では勾配法による反復最適化を用いており，各反復で目的関数の偏微分係数のみが必要であったのに対し，IP の場合は偏微分後の代数方程式を解く必要がある．そのため，前述の従来手法を IP に基づく ILRMA に導入することはこれまで不可能であった．したがって，分離行列最適化のための新たなアルゴリズムを導出する必要がある．本稿では，IP にインスパイアされた最適化法を提案法の目的関数に適用する．本最適化では，分離行列の各行ベクトルごとに更

新することで効率的に目的関数を下げることができる．

2 従来手法

2.1 記号の定義

音源数およびマイクロホン数をそれぞれ N, M と定義する．また，混合前の音源，観測信号，推定分離信号を短時間フーリエ変換したものをそれぞれ

$$\boldsymbol{s}_{ij} \stackrel{\mathrm{def}}{=} (s_{ij,1}, s_{ij,2}, \ldots, s_{ij,N})^{\mathsf{T}} \tag{1}$$

$$\boldsymbol{x}_{ij} \stackrel{\mathrm{def}}{=} (x_{ij,1}, x_{ij,2}, \ldots, x_{ij,M})^{\mathsf{T}} \tag{2}$$

$$\boldsymbol{y}_{ij} \stackrel{\mathrm{def}}{=} (y_{ij,1}, y_{ij,2}, \ldots, y_{ij,N})^{\mathsf{T}} \tag{3}$$

と定義する．ここで，$i = 1, \ldots, I$; $j = 1, \ldots, J$; $n = 1, \ldots, N$; および $m = 1, \ldots, M$ はそれぞれ周波数ビン，時間フレーム，音源および観測チャネルのインデクスをそれぞれ表し，$^{\mathsf{T}}$ は転置を表す．混合系が時不変かつ短時間フーリエ変換の窓長が各音源から各マイクまでのインパルス応答長より十分長い場合，ランク 1 空間モデル [16] が成立し，混合行列 $\boldsymbol{A}_i$ を用いて観測信号を

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij} \tag{4}$$

と表せる．$M = N$ かつ $\boldsymbol{A}_i$ が正則の場合は，分離行列 $\boldsymbol{W}_i = (\boldsymbol{w}_{i,1}, \ldots, \boldsymbol{w}_{i,N})^{\mathsf{H}} = \boldsymbol{A}_i^{-1}$ が存在し，分離信号を

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij} \tag{5}$$

と表せる．ここで，$^{\mathsf{H}}$ はエルミート転置を表す．

2.2 ILRMA

ILRMA は IVA と NMF を融合した音源分離手法であり，空間モデルと音源モデルを同時に推定する最適化問題として定式化される．図 1 は ILRMA による音源分離の原理を示している．分離行列及び NMF 音源モデル最適化の過程では，分離信号のパワースペクトログラムを低ランク行列としてモデル化しながら，その時間周波数構造を共変関係として加味した分離行列 $\boldsymbol{W}_i$ を推定する．混合前の各音源のパワースペクトログラムが低ランクであれば，混合信号のパワースペクトログラムのランクは基本的に増加することから，ILRMA は分離信号を低ランクに誘導することでパーミュテーション不整合を避けつつ，互いに独立となる分離信号を推定している．また，ILRMA は multichannel NMF（MNMF） [17] に対しランク 1 空間モデル (4) を仮定したものと等価であり，MNMF と比較してより効率的な反復最適化アルゴリズムを与える．

ILRMA の目的関数は，次式で定義される．

$$\mathcal{J} \stackrel{\mathrm{def}}{=} \sum_{i,j,n} \left[\frac{|\boldsymbol{w}_{i,n}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2}{\sum_l t_{il,n} v_{lj,n}} + \log \sum_l t_{il,n} v_{lj,n} \right]$$

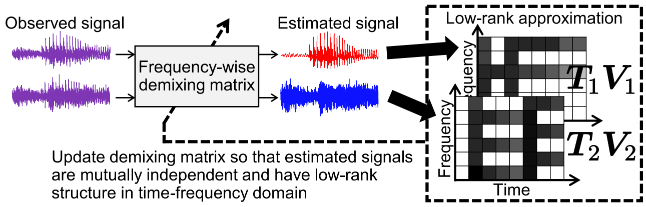

図 1: Principle of source separation based on ILRMA.

$$- 2J \sum_i \log |\det \boldsymbol{W}_i| \tag{6}$$

ここで，$t_{il,n}$ および $v_{lj,n}$ は n 番目の音源に関する基底行列 $\boldsymbol{T}_n$ およびアクティベーション行列 $\boldsymbol{V}_n$ の非負要素であり，$l = 1, \ldots, L$ は基底インデクスである．ランク L の行列 $\boldsymbol{T}_n \boldsymbol{V}_n$ は，n 番目の推定分離信号のパワースペクトログラムに対応する音源モデルである．式 (6) の第 1 項および第 3 項は IVA の目的関数に，第 1 項および第 2 項は NMF の目的関数にそれぞれ対応している．この目的関数の最小化には，補助関数法を適用した効率的な反復最適化アルゴリズムが提案されている [7]．

分離行列 $\boldsymbol{W}_i$ に着目して目的関数 $\mathcal{J}$ を書き換えると，

$$\mathcal{J} = J \sum_i \left[\sum_n \boldsymbol{w}_{i,n}^{\mathsf{H}} \boldsymbol{D}_{i,n} \boldsymbol{w}_{i,n} - \log |\det \boldsymbol{W}_i|^2 \right] + \mathcal{C} \tag{7}$$

$$\boldsymbol{D}_{i,n} \stackrel{\mathrm{def}}{=} \frac{1}{J} \sum_j \frac{\boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}}}{\sum_l t_{il,n} v_{lj,n}} \tag{8}$$

となる．ここで，$\mathcal{C}$ は $\boldsymbol{w}_{i,n}$ と無関係の項を示す．$\boldsymbol{D}_{i,n}$ は，定義より半正定値エルミート行列であり，音源分離においては正定値であることを仮定する．式 (6) は，第 1 項が二次形式，第 2 項が分離行列 $\boldsymbol{W}_i$ の log det 項となっており，上述の目的関数は iterative projection（IP） [7], [15] を利用した最適化が可能である．IP を利用した空間分離フィルタ $\boldsymbol{w}_{i,n}$ の更新式は以下で与えられる．

$$\boldsymbol{u}_{i,n} = \boldsymbol{D}_{i,n}^{-1} \boldsymbol{W}_i^{-1} \boldsymbol{e}_n \tag{9}$$

$$\boldsymbol{w}_{i,n} \leftarrow \frac{\boldsymbol{u}_{i,n}}{\sqrt{\boldsymbol{u}_{i,n}^{\mathsf{H}} \boldsymbol{D}_{i,n} \boldsymbol{u}_{i,n}}} \tag{10}$$

ここで，$\boldsymbol{e}_n$ は n 番目の要素のみが 1 であり，他の要素が全て 0 であるベクトルを示す．

2.3 ビームフォーマ

マイクロホンアレイを利用した信号処理において，指向性を制御する信号処理手法をビームフォーマと呼ぶ．ある音源の到来方位 θ が既知の場合，その方位から到来する音波を抑圧するビームフォーマが構成でき，これはヌルビームフォーマ（null beamformer: NBF）と呼ばれる．i 番目の周波数ビンにおけるビームフォーマ係数を $\boldsymbol{g}_i(\theta)$,

角度 θ より到来する音源のステアリングベクトルを $\boldsymbol{a}_i(\theta)$ と定義すると，$\boldsymbol{g}_i(\theta)$ は以下の条件を満たす．

$$\boldsymbol{g}_i(\theta)^\mathsf{T}\boldsymbol{a}_i(\theta) = 0 \tag{11}$$

マイクロホンが 2 素子の場合，角度 θ より到来する音を抑圧する NBF 係数の一例は次式で与えられる．

$$\boldsymbol{g}_i(\theta) = \rho_i(\theta)\begin{bmatrix} \exp\left(-\mathrm{j}\dfrac{\pi(i-1)f_\mathrm{s}d\sin\theta}{c_\mathrm{S}N_\mathrm{F}}\right) \\ -\exp\left(\mathrm{j}\dfrac{\pi(i-1)f_\mathrm{s}d\sin\theta}{c_\mathrm{S}N_\mathrm{F}}\right)\end{bmatrix} \tag{12}$$

ここで，j は虚数単位，c_S は音速，f_s はサンプリング周波数，N_F は短時間フーリエ変換の FFT 長，d はマイクロホン間隔である．また，$\rho_i(\theta)$ は周波数ビンごとのスケールを調節するパラメタであり，通常はある $\theta'(\neq\theta)$ に対し

$$\boldsymbol{g}_i(\theta)^\mathsf{T}\boldsymbol{a}_i(\theta') = 1 \tag{13}$$

となるように $\rho_i(\theta)$ を定めることで歪みを抑制する．

3 方位事前情報を導入した ILRMA

3.1 動機

音源の低ランク構造が明確である音楽信号の分離タスクにおいて，ILRMA は高い分離性能を実現しているが，低ランクでない音声信号等の分離タスクにおいては，特に NMF の基底数を多くした場合に，悪い局所解に陥り分離に失敗してしまう問題がある [7]．また音楽信号においても，ILRMA は IVA と同様にブロックパーミュテーション不整合を起こす場合 [18] がある．一方で，FDICA においては，音源の空間的な配置を正則化として与えることにより，パーミュテーション不整合を解消しつつより良い解に高速に誘導する手法が提案されている [3], [11], [12]．本稿では，音源の到来方位が既知の状況において，ビームフォーマより得られる空間分離フィルタとの誤差を ILRMA の目的関数に正則化として導入することで，ブロックパーミュテーション不整合を抑制しつつ，高精度な分離に誘導するアルゴリズムを提案する．

3.2 空間事前情報罰則付き ILRMA の目的関数

空間事前情報より得られる分離行列の教師を $\widehat{\boldsymbol{W}}_i = (\widehat{\boldsymbol{w}}_{i,1},\ldots,\widehat{\boldsymbol{w}}_{i,N})^\mathsf{H}$ と定義する．たとえば，2 種類の音源を 2 本のマイクロホンで録音した状況において，双方の音源到来方位 θ_1, θ_2 が既知の場合，NBF を利用して

$$\widehat{\boldsymbol{W}}_i = (\widehat{\boldsymbol{w}}_{i,1}, \widehat{\boldsymbol{w}}_{i,2})^\mathsf{H} = (\boldsymbol{g}_i(\theta_2), \boldsymbol{g}_i(\theta_1))^\mathsf{T} \tag{14}$$

と定めることができる．このとき，空間事前情報から得られる分離フィルタ $\widehat{\boldsymbol{w}}_{i,n}$ と ILRMA で最適化を行う分離フィルタ $\boldsymbol{w}_{i,n}$ の 2 乗誤差を ILRMA の目的関数に対し正則化項として追加することで，新たな目的関数

$$\begin{aligned}\mathcal{J}_\mathrm{P} &\stackrel{\mathrm{def}}{=} \mathcal{J} + J\sum_{i,n}\lambda_n\|\boldsymbol{w}_{i,n}-\widehat{\boldsymbol{w}}_{i,n}\|^2 \\ &= J\sum_i\left[\sum_n\left(\boldsymbol{w}_{i,n}^\mathsf{H}\widehat{\boldsymbol{D}}_{i,n}\boldsymbol{w}_{i,n} - \lambda_n\widehat{\boldsymbol{w}}_{i,n}^\mathsf{H}\boldsymbol{w}_{i,n}\right.\right. \\ &\qquad \left.\left. - \lambda_n\boldsymbol{w}_{i,n}^\mathsf{H}\widehat{\boldsymbol{w}}_{i,n}\right) - \log|\det\boldsymbol{W}_i|^2\right] + \mathcal{C}'\end{aligned} \tag{15}$$

を得る．ここで，$\mathcal{C}'$ は $\boldsymbol{w}_{i,n}$ と無関係の項，λ_n は罰則の重みを示すハイパーパラメタである．また，$\boldsymbol{I}_N$ を大きさ $N\times N$ の単位行列として，$\widehat{\boldsymbol{D}}_{i,n} \stackrel{\mathrm{def}}{=} \boldsymbol{D}_{i,n} + \lambda_n\boldsymbol{I}_N$ と定義する．従来の目的関数 $\mathcal{J}$ と罰則付き目的関数 $\mathcal{J}_\mathrm{P}$ を比較すると，罰則付き目的関数には $\boldsymbol{w}_{i,n}$ の内積項が新たに加わっている．目的関数が二次形式と log det 項の和として表現できる $\mathcal{J}$ に対しては IP [15] を利用した最適化が可能であるが，$\mathcal{J}_\mathrm{P}$ の場合は不可能であるため，新たな最適化手法を考案する必要がある．以下では，IP と同様に分離行列 $\boldsymbol{W}_i$ に関する罰則付き目的関数 $\mathcal{J}_\mathrm{P}$ を行ベクトル $\boldsymbol{w}_{i,n}$ ごとに最適化する手法を提案する．

3.3 余因子展開の導入

式 (15) には $\log|\det\boldsymbol{W}_i|^2$ の項が含まれているが，ここでは分離フィルタ $\boldsymbol{w}_{i,n}$ による偏微分を考えるため，余因子展開を導入する．分離行列 $\boldsymbol{W}_i$ の余因子行列を

$$\boldsymbol{B}_i = (\boldsymbol{b}_{i,1},\ldots,\boldsymbol{b}_{i,N}) \stackrel{\mathrm{def}}{=} (\det\boldsymbol{W}_i)\boldsymbol{W}_i^{-1} \tag{16}$$

と定義する．このとき，定義より

$$\boldsymbol{b}_{i,n} = (\det\boldsymbol{W}_i)\boldsymbol{W}_i^{-1}\boldsymbol{e}_n \tag{17}$$

$$\boldsymbol{w}_{i,p}^\mathsf{H}\boldsymbol{b}_{i,q} = \delta_{pq}\det\boldsymbol{W}_i \tag{18}$$

が成り立つ．ここで，δ_{pq} はクロネッカーのデルタである．この結果，目的関数 $\mathcal{J}_\mathrm{P}$ は次式のように変形できる．

$$\begin{aligned}\mathcal{J}_\mathrm{P} &= J\sum_i\left[\sum_n\left(\boldsymbol{w}_{i,n}^\mathsf{H}\widehat{\boldsymbol{D}}_{i,n}\boldsymbol{w}_{i,n} - \lambda_n\widehat{\boldsymbol{w}}_{i,n}^\mathsf{H}\boldsymbol{w}_{i,n}\right.\right. \\ &\qquad \left.\left. - \lambda_n\boldsymbol{w}_{i,n}^\mathsf{H}\widehat{\boldsymbol{w}}_{i,n}\right) - \log|\boldsymbol{w}_{i,n}^\mathsf{H}\boldsymbol{b}_{i,n}|^2\right] + \mathcal{C}'\end{aligned} \tag{19}$$

ベクトル $\boldsymbol{b}_{i,n}$ は $\boldsymbol{w}_{i,n}$ に依存しないとみなすことが可能であるから，各分離フィルタ $\boldsymbol{w}_{i,n}$ ごとに目的関数 (19) を小さくするよう更新するブロック座標降下法が適用できる．

3.4 反復更新アルゴリズムの導出

罰則付き目的関数 $\mathcal{J}_\mathrm{P}$ の停留点へ分離フィルタ $\boldsymbol{w}_{i,n}$ を更新する．このとき，$\mathcal{L} \stackrel{\mathrm{def}}{=} \mathcal{J}_\mathrm{P}/J$ と定義し，$\mathcal{L}$ を $\boldsymbol{w}_{i,n}^*$ で偏微分すると

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{w}_{i,n}^*} = \widehat{\boldsymbol{D}}_{i,n}\boldsymbol{w}_{i,n} - \lambda_n\widehat{\boldsymbol{w}}_{i,n} - \frac{\boldsymbol{b}_{i,n}}{\boldsymbol{w}_{i,n}^\mathsf{H}\boldsymbol{b}_{i,n}} \tag{20}$$

となる．$\beta_{i,n} \overset{\mathrm{def}}{=} 1/\boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{b}_{i,n}$ とおくと，$\mathcal{J}_{\mathrm{P}}$ の停留点において

$$\widehat{\boldsymbol{D}}_{i,n}\boldsymbol{w}_{i,n} - \lambda_n \widehat{\boldsymbol{w}}_{i,n} - \beta_{i,n}\boldsymbol{b}_{i,n} = \boldsymbol{0} \tag{21}$$

$$\Longleftrightarrow \boldsymbol{w}_{i,n} = \widehat{\boldsymbol{D}}_{i,n}^{-1}(\beta_{i,n}\boldsymbol{b}_{i,n} + \lambda_n \widehat{\boldsymbol{w}}_{i,n}) \tag{22}$$

が成立する．$\beta_{i,n}$ の定義より $\beta_{i,n}\boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{b}_{i,n} - 1 = 0$ であるから，式 (22) をこれに代入すると，次の方程式

$$\boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n}|\beta_{i,n}|^2 + \lambda_n \widehat{\boldsymbol{w}}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n}\beta_{i,n} - 1 = 0 \tag{23}$$

を得る．ここで，式 (23) の左辺第 1 項と第 3 項および右辺は実数であるから，

$$\mathrm{Im}\left[\lambda_n \widehat{\boldsymbol{w}}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n}\beta_{i,n}\right] = 0 \tag{24}$$

を満たす必要がある．すなわち，$\beta_{i,n} \neq 0$ であることに注意すると，$\gamma_{i,n} \in \mathbb{R}$ として，

$$\beta_{i,n} = \gamma_{i,n}\left(\lambda_n \widehat{\boldsymbol{w}}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n}\right)^* = \gamma_{i,n}\lambda_n \boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\widehat{\boldsymbol{w}}_{i,n} \tag{25}$$

または

$$\lambda_n \widehat{\boldsymbol{w}}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n} = 0 \tag{26}$$

を満たす．条件 (25) が成立する場合，方程式 (23) は

$$\begin{aligned}&\lambda_n^2 \boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n}|\boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\widehat{\boldsymbol{w}}_{i,n}|^2\gamma_{i,n}^2 \\ &\quad + \lambda_n^2|\boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\widehat{\boldsymbol{w}}_{i,n}|^2\gamma_{i,n} - 1 = 0\end{aligned} \tag{27}$$

と変形でき，この方程式の実数解は

$$\gamma_{i,n} = \frac{1}{2\boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n}}\left[-1 \pm \sqrt{1 + \frac{4\boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n}}{\lambda_n^2|\boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\widehat{\boldsymbol{w}}_{i,n}|^2}}\right] \tag{28}$$

となるから，$\beta_{i,n}$ は

$$\beta_{i,n} = \frac{\lambda_n \boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\widehat{\boldsymbol{w}}_{i,n}}{2\boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n}}\left[-1 \pm \sqrt{1 + \frac{4\boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n}}{\lambda_n^2|\boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\widehat{\boldsymbol{w}}_{i,n}|^2}}\right] \tag{29}$$

となる．根拠は 3.5 節で述べるが，$\pm$ の部分における任意性は $+$ の解を利用することで解消する．一方，条件 (26) が成立する場合，方程式 (23) は

$$\boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n}|\beta_{i,n}|^2 - 1 = 0 \tag{30}$$

$$\Longleftrightarrow \beta_{i,n} = \frac{e^{\mathrm{j}\phi_{i,n}}}{\sqrt{\boldsymbol{b}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{b}_{i,n}}} \tag{31}$$

となる．ここで，$\phi_{i,n}$ は任意の位相であり，$e^{\mathrm{j}\phi_{i,n}} = (\det \boldsymbol{W}_i)^*/|\det \boldsymbol{W}_i|$ となるように $\phi_{i.n}$ を定め，任意性を解消する．式 (22)，式 (29)，式 (31) をまとめると，分離フィルタ $\boldsymbol{w}_{i,n}$ の更新式は

$$\boldsymbol{u}_{i,n} = \widehat{\boldsymbol{D}}_{i,n}^{-1}\boldsymbol{W}_i^{-1}\boldsymbol{e}_n \tag{32}$$

$$\widehat{\boldsymbol{u}}_{i,n} = \lambda_n \widehat{\boldsymbol{D}}_{i,n}^{-1}\widehat{\boldsymbol{w}}_{i,n} \tag{33}$$

$$h_{i,n} = \boldsymbol{u}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}\boldsymbol{u}_{i,n} \tag{34}$$

$$\widehat{h}_{i,n} = \boldsymbol{u}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}\widehat{\boldsymbol{u}}_{i,n} \tag{35}$$

$$\boldsymbol{w}_{i,n} \leftarrow \begin{cases} \dfrac{\boldsymbol{u}_{i,n}}{\sqrt{h_{i,n}}} + \widehat{\boldsymbol{u}}_{i,n} & (\text{if } \widehat{h}_{i,n} = 0) \\ \dfrac{\widehat{h}_{i,n}}{2h_{i,n}}\left[-1 + \sqrt{1 + \dfrac{4h_{i,n}}{|\widehat{h}_{i,n}|^2}}\right]\boldsymbol{u}_{i,n} + \widehat{\boldsymbol{u}}_{i,n} & \\ & (\text{otherwise}) \end{cases} \tag{36}$$

となる．

3.5 解の任意性の解消

前節の各条件においては，解の任意性を解消する必要がある．式 (31) に関しては，位相 $\phi_{i,n}$ の部分に任意性が存在するが，この位相はどのように選んでも目的関数 $\mathcal{J}_{\mathrm{P}}$ の値が変化しないため，適当に定めてよく，今回は $e^{\mathrm{j}\phi_{i,n}} = (\det \boldsymbol{W}_i)^*/|\det \boldsymbol{W}_i|$ を満たすように定める．一方，式 (28) に関しては，$\pm$ の正負に関して任意性が存在する．ここで，目的関数 $\mathcal{L}$ のうち $\boldsymbol{w}_{i,n}$ に関係する項のみを取り出すと，

$$\begin{aligned}&\boldsymbol{w}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{D}}_{i,n}\boldsymbol{w}_{i,n} - \lambda_n \widehat{\boldsymbol{w}}_{i,n}^{\mathsf{H}}\boldsymbol{w}_{i,n} - \lambda_n \boldsymbol{w}_{i,n}^{\mathsf{H}}\widehat{\boldsymbol{w}}_{i,n} \\ &\qquad - \log|\boldsymbol{w}_{i,n}^{\mathsf{H}}\boldsymbol{b}_{i,n}|^2\end{aligned} \tag{37}$$

となり，$\boldsymbol{w}_{i,n}$ の更新後においては，各条件式を代入することで

$$2\log|\beta_{i,n}| - \lambda_n \widehat{\boldsymbol{w}}_{i,n}^{\mathsf{H}}\boldsymbol{w}_{i,n} + 1 \tag{38}$$

と変形できる．式 (38) のうち第 1 項に着目すると，$\beta_{i,n}^{(+)}$，$\beta_{i,n}^{(-)}$ をそれぞれ $+$，$-$ の各符号を選択した場合における解とすれば，

$$\left|-1 - \sqrt{1 + \frac{4h_{i,n}}{|\widehat{h}_{i,n}|^2}}\right| > \left|-1 + \sqrt{1 + \frac{4h_{i,n}}{|\widehat{h}_{i,n}|^2}}\right| \tag{39}$$

$$\Longleftrightarrow \log\left|\beta_{i,n}^{(-)}\right| > \log\left|\beta_{i,n}^{(+)}\right| \tag{40}$$

より，正符号の場合のほうが小さな値となる．また，第 2 項に着目すると，

$$-\lambda_n \widehat{\boldsymbol{w}}_{i,n}^{\mathsf{H}}\boldsymbol{w}_{i,n} = -\frac{|\widehat{h}_{i,n}|^2}{2h_{i,n}}\left(-1 \pm \sqrt{1 + \frac{4h_{i,n}}{|\widehat{h}_{i,n}|^2}}\right) \tag{41}$$

より，こちらも正符号の場合のほうが小さな値となる．以上より，符号の任意性の部分では正符号を選んだほうが $\mathcal{J}_{\mathrm{P}}$ の値を小さくできる．また，目的関数 $\mathcal{J}_{\mathrm{P}}$ は最小値を持つ関数であるから，目的関数が小さくなる方の停留点において必ず最小となっている．

3.6 罰則重み係数の調節

ビームフォーマとの誤差を FDICA の目的関数に加える従来の研究 [3], [12] において，最適化の前半ではビームフォーマが高速かつ高い分離性能を発揮する一方で，最適化の後半ではビームフォーマの影響を小さくし，FDICA による音源分離へ委ねるのが好ましいことが報告されている．本研究においてもこれに倣い，罰則項の重みを表すハイパーパラメタ λ_n を分離が進むにつれてだんだん小さな値へと変化させる手法を導入する．本稿では，現在の ILRMA の反復回数を c，総反復回数を C とするとき，

$$\lambda_n(c) = \alpha_n \max\left[\frac{1}{2} - \frac{c}{C}, 0\right] \tag{42}$$

として，c 回目の反復における罰則重み $\lambda_n(c)$ 定める．

4 実験

4.1 実験条件

提案法の有効性を確認するため，2 音源 2 マイクロホンによる擬似的な観測信号を利用して，以下の 3 種類の音源分離性能を比較した．

- 分離行列 $\boldsymbol{W}_i$ を単位行列で初期化し，従来の ILRMA を用いて分離（Method 1）
- 分離行列 $\boldsymbol{W}_i$ を空間事前情報より得られる分離行列 $\widehat{\boldsymbol{W}}_i$ で初期化し，従来の ILRMA を用いて分離（Method 2）
- 分離行列 $\boldsymbol{W}_i$ を空間事前情報より得られる分離行列 $\widehat{\boldsymbol{W}}_i$ で初期化し，空間事前情報より得られる分離フィルタとの誤差を罰則として加えた ILRMA を用いて分離（Method 3）

なお，Method 2 は提案手法において $\alpha_n = 0$ と設定した場合に対応し，Method 3 は $\alpha_n = \{0.1, 0.3, 1, 3, 10\}$ の 5 通りに関して検証した．分離対象とする擬似的な観測信号は，表 1 に示す音源分離コンペティション SiSEC2010 [19] で利用された音声信号に対し，RWCP データベース [20] に収録されている E2A インパルス応答（$\mathrm{RT}_{60} = 300$ ms）を畳み込むことで作成した．音源の到来方向は，マイクロホンアレイの正面を 0°，時計回りを正の方向として，$(-40°, +40°)$ と $(-40°, +20°)$ の 2 通りについて検証した．空間事前情報より得られる分離行列の教師 $\widehat{\boldsymbol{W}}_i$ は，式

表 1: Speech sources

ID	Speech name	Track name
1	`dev1_female4`	`src_1/src_2`
2	`dev1_female4`	`src_3/src_4`
3	`dev1_male4`	`src_1/src_2`
4	`dev1_male4`	`src_3/src_4`

表 2: Average SDR improvements [dB], where $(\theta_1, \theta_2) = (-40°, +40°)$

Method	α_n	Number of bases L								
		1	2	3	5	10	15	20	25	30
1	—	6.07	7.34	6.22	6.26	5.60	5.47	5.36	4.96	4.98
2	0.0	8.02	9.43	10.39	11.06	11.47	11.73	11.45	11.39	11.24
3	0.1	8.05	9.63	10.80	11.71	12.05	12.37	12.17	**12.25**	11.97
	0.3	8.06	9.84	11.03	12.06	12.24	12.48	**12.29**	12.17	11.93
	1.0	8.06	10.15	11.33	**12.16**	**12.32**	**12.56**	12.17	12.21	**12.01**
	3.0	8.06	10.31	**11.38**	12.08	12.13	12.36	12.00	11.99	11.75
	10.0	**8.06**	**10.56**	11.32	11.74	11.77	11.76	11.49	11.37	11.22

(14) により定めた．信号のサンプリング周波数は 16000 Hz であり，短時間フーリエ変換における FFT 長は 256 ms（4096 サンプル），シフト長は 128 ms（2048 サンプル）に設定した．ILRMA の総反復回数 C は 100 回に設定し，音源あたりの基底数 L は $\{1, 2, 3, 5, 10, 15, 20, 25, 30\}$ の 9 通りを検証した．音源分離の評価指標として，総合的な分離性能を示す signal-to-distortion ratio (SDR) [21] の改善量を利用した．基底行列 $\boldsymbol{T}_n$ とアクティベーション行列 $\boldsymbol{V}_n$ の初期値は一様乱数から生成し，各実験条件に対してこれらの初期値を 20 通り変化させた場合の平均 SDR 改善量を比較した．

4.2 結果

音源到来方位が $(-40°, +40°)$ と $(-40°, +20°)$ の各場合における音源分離性能は，それぞれ表 2 および表 3 に示すとおりである．なお，8 種類すべての音源の平均 SDR 改善量を示している．従来の ILRMA では，特に音声信号の音源分離において，基底数を大きくした場合に高精度な分離ができない欠点を抱えていたが，空間事前情報に由来する初期値を利用することで分離性能が大幅に改善し，空間事前情報との罰則を目的関数に加えることで更に性能が改善したことを表より読み取ることができる．NMF 音源モデルを利用してパーミュテーション解決を行う従来の ILRMA においては，音声信号の NMF によるモデリングが困難であることに起因して分離性能が低下していたが，方位事前情報を ILRMA の最適化の中で補助的に利用することで，NMF による正確な音源モデル推定が可能となり，複雑な時間周波数構造を持つ音声信号に対してもより多くの基底を用いて表現できるようになったと考えられる．

表 3: Average SDR improvement [dB], where $(\theta_1, \theta_2) = (-40°, +20°)$

Method	α_n	Number of bases L								
		1	2	3	5	10	15	20	25	30
1	—	3.14	3.51	3.14	3.25	2.86	2.58	2.46	2.17	2.20
2	0.0	**7.37**	8.39	8.86	9.57	9.74	9.84	9.81	9.81	9.81
3	0.1	7.32	8.61	9.30	10.04	10.36	10.61	10.39	10.52	10.37
	0.3	7.32	8.71	9.63	10.43	10.77	10.84	10.70	10.88	10.55
	1.0	7.32	9.05	10.05	10.70	**11.01**	**11.10**	10.79	**10.98**	**10.88**
	3.0	7.32	9.24	**10.22**	**10.70**	10.98	11.09	**10.88**	10.88	10.72
	10.0	7.33	**9.28**	10.06	10.61	10.77	10.93	10.64	10.66	10.49

5 おわりに

本稿では，ILRMA に対し音源の空間的な配置を事前情報として利用できる場合に関する検討を行った．従来 ILRMA の目的関数に対し，ビームフォーマとの誤差を罰則として加えることで得られる新たな目的関数に対し，分離行列の各行ベクトルごとに最適化を進めるアルゴリズムを導出した．新たに導出した反復最適化アルゴリズムは，ハイパーパラメタが不要であり，目的関数の単調非増加性を保証している．音源分離実験を通じ，空間事前情報を補助的に利用することで ILRMA の性能を更に向上可能であることが示された．今後の展望として，音源到来方位が未知の場合に，音源方位推定を同時に行うことでブロックパーミュテーション不整合の解消を行う手法への発展や，真の音源到来方位と事前情報に誤差が生じている場合の考察などが考えられる．

謝辞

本研究は，総合科学技術・イノベーション会議による革新的研究開発プログラム（ImPACT），セコム科学技術振興財団，および JSPS 科研費 17H06572 の助成を受けた．

参考文献

[1] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, vol. 22, no. 1, pp. 21–34, 1998.

[2] H. Sawada, R. Mukai, S. Araki, and S. Makino, “A robust and precise method for solving the permutation problem of frequency-domain blind source separation,” *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 530–538, 2004.

[3] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, “Blind source separation based on a fast-convergence algorithm combining ICA and beamforming,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 666–678, 2006.

[4] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, “Blind source separation exploiting higher-order frequency dependencies,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 1, pp. 70–79, 2007.

[5] R. Ikeshita, Y. Kawaguchi, M. Togami, Y. Fujita, and K. Nagamatsu, “Independent vector analysis with frequency range division and prior switching,” in *Proc. Eur. Signal Process. Conf.*, 2017, pp. 2393–2397.

[6] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[7] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, 2016.

[8] D. Kitamura. (2016) Algorithms for independent low-rank matrix analysis. [Online]. Available: http://d-kitamura.net/pdf/misc/AlgorithmsForIndependentLowRankMatrixAnalysis.pdf

[9] Y. Mitsui, D. Kitamura, S. Takamichi, N. Ono, and H. Saruwatari, “Blind source separation based on independent low-rank matrix analysis with sparse regularization for time-series activity,” in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2017, pp. 21–25.

[10] Y. Liang, S. Naqvi, and J. Chambers, “Overcoming block permutation problem in frequency domain blind source separation when using AuxIVA algorithm,” *Electron. Lett.*, vol. 48, no. 8, pp. 460–462, 2012.

[11] L. C. Parra and C. V. Alvino, “Geometric source separation: Merging convolutive source separation with geometric beamforming,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 10, no. 6, pp. 352–362, 2002.

[12] K. Osako, Y. Mori, Y. Takahashi, H. Saruwatari, and K. Shikano, “Fast convergence blind source separation using frequency subband interpolation by null beamforming,” *IEICE Trans. Fundam.*, vol. E91-A, no. 6, pp. 1357–1361, 2008.

[13] Y. Zheng, K. Reindl, and W. Kellermann, “BSS for improved interference estimation for blind speech signal extraction with two microphones,” in *Proc. Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, 2009, pp. 253–256.

[14] Y. Tachioka, T. Narita, and J. Ishii, “Semi-blind source separation using binary masking and independent vector analysis,” *IEEJ Trans. Electr. Electron. Eng.*, vol. 10, no. 1, pp. 114–115, 2015.

[15] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2011, pp. 189–192.

[16] N. Q. K. Duong, E. Vincent, and R. Gribonval, “Under-determined reverberant audio source separation using a full-rank spatial covariance model,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1830–1840, 2010.

[17] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, “Multichannel extensions of non-negative matrix factorization with complex-valued data,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 5, pp. 971–982, 2013.

[18] Y. Mitsui, D. Kitamura, N. Takamune, H. Saruwatari, Y. Takahashi, and K. Kondo, “Independent low-rank matrix analysis based on parametric majorization-equalization algorithm,” in *Proc. Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, 2017, (in press).

[19] S. Araki, A. Ozerov, V. Gowreesunker, H. Sawada, F. Theis, G. Nolte, D. Lutter, and N. Q. K. Duong, “The 2010 signal separation evaluation campaign (SiSEC2010): Audio source separation,” in *Proc. Int. Conf. Latent Variable Anal. Signal Separation*, 2010, pp. 114–122.

[20] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, “Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition,” in *Proc. Int. Conf. Lang. Resources Evaluation*, 2000, pp. 965–968.

[21] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.