

## [招待講演] 独立低ランク行列分析に基づくブラインド音源分離

北村 大地<sup>†</sup> 小野 順貴<sup>††</sup> 澤田 宏<sup>†††</sup> 亀岡 弘和<sup>†††</sup> 猿渡 洋<sup>†</sup>

<sup>†</sup> 東京大学 〒113-8656 東京都文京区本郷 7-3-1

<sup>††</sup> 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

<sup>†††</sup> 日本電信電話株式会社 〒619-0237 京都府相楽郡精華町光台 2-4

あらまし 本稿では、新しいブラインド音源分離手法である独立低ランク行列分析 (ILRMA) を提案する。ILRMA は、従来の有名なブラインド音源分離手法の独立ベクトル分析 (IVA) と、非負行列を対象とした低ランク近似理論の非負値行列因子分解 (NMF) の二つを融合したアルゴリズムであり、独立成分分析を起源とする統計的独立性に基づく信号分離理論の正当な拡張手法である。また、本手法は NMF を多次元観測信号用に拡張した多チャンネル NMF (MNMF) の特殊形とも解釈できることから、近年の最先端音源分離アルゴリズム (IVA 及び MNMF) の関係性を、ILRMA を通して理論的に明らかにすることができる。

キーワード ブラインド音源分離, 独立成分分析, 非負値行列因子分解

## Blind Source Separation Based on Independent Low-Rank Matrix Analysis

Daichi KITAMURA<sup>†</sup>, Nobutaka ONO<sup>††</sup>, Hiroshi SAWADA<sup>†††</sup>, Hirokazu KAMEOKA<sup>†††</sup>, and  
Hiroshi SARUWATARI<sup>†</sup>

<sup>†</sup> The University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

<sup>††</sup> National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

<sup>†††</sup> Nippon Telegraph and Telephone Corporation 2-4 Hikaridai, Seika-cho, Kyoto 619-0237, Japan

**Abstract** In this paper, we propose a new efficient algorithm for blind source separation problem (BSS) called independent low-rank matrix analysis (ILRMA). ILRMA is a unified method of independent vector analysis (IVA) and nonnegative matrix factorization (NMF) and can be interpreted as a natural extension of the source separation theory based on statistical independence. Also, ILRMA is identical to a special case of multichannel NMF (MNMF), and this fact reveals the relationship between the state-of-the-art BSS techniques including IVA and MNMF.

**Key words** blind source separation, independent component analysis, nonnegative matrix factorization

### 1. はじめに

ブラインド音源分離 (blind source separation: BSS) とは、音源位置や混合系が未知の条件で観測された信号のみから混合前の音源信号を推定する音響信号処理技術である。優決定条件 (音源数  $\leq$  観測チャンネル数) における BSS では、独立成分分析 (independent component analysis: ICA) [1] に基づく手法が歴史的に主流である。ICA は、音源間の統計的独立性と音源信号の非ガウスな生成モデルを仮定することで、混合行列の逆である分離行列を推定する理論である。音響信号では混合系が部屋の残響に起因する畳み込み混合となることから、短時間 Fourier 変換 (short-time Fourier transform: STFT) で得られる複素スペクトログラムの各周波数ビンに対して ICA を適用することで周波数毎の分離行列を推定する周波数領域

ICA (frequency-domain ICA: FDICA) [2] が提案された。しかしながら、ICA は分離信号の順序に任意性があるため、周波数ビン毎に推定される分離信号の順序を適切に並び替えるポスト処理が新たに必要となる。このポスト処理はパーミュテーション問題と呼ばれ、これまでに様々な解決法が提案されている [3]–[7]。やがて、パーミュテーション問題の解決と周波数毎の分離行列の推定を同時に達成する手法として、独立ベクトル分析 (independent vector analysis: IVA) [8]–[10] が提案された。IVA は ICA を多変量に拡張した理論であり、各音源の周波数成分を一つにまとめた周波数ベクトルの生成モデル (多変量確率分布) を仮定する。このとき、球対称な性質を持つ非ガウスな多変量分布を仮定することで、同一音源の周波数成分間の共変関係 (高次相関) [10] を考慮でき、パーミュテーション問題を避けながら分離行列を推定できる。また、ICA や IVA の

最適化に対して補助関数法 [11] を用いた高速かつ安定なアルゴリズムも提案されている [12], [13].

一方、モノラル信号等を対象とした劣決定条件（音源数 > 観測チャンネル数）下では、非負行列の低ランク近似理論である非負値行列因子分解（nonnegative matrix factorization: NMF）[14] を応用した音源分離手法 [15]–[17] が注目を集めている。NMF に基づく手法では、混合信号のパワースペクトログラム等を非負行列とみなして少数の非負パーツに分解することで、音源間のスペクトルの違いを用いた音源分離とが可能となる。しかしながら、分解された非負パーツを音源毎に選別することは一般に容易ではない。多チャンネルの観測信号に対しては、音源毎の空間的な伝達系（空間相関行列 [18]）の違いを用いて非負パーツを音源毎に選別することで分離を実現する多チャンネル NMF（multichannel NMF: MNMF）[19]–[21] が提案されている。MNMF は、劣決定条件における高精度な音源分離を実現しているが、最適化変数の初期値に対する頑健性や反復計算における膨大な計算コストが解決すべき問題となっている。

本稿では、優決定条件における BSS を対象として、IVA と NMF を融合した新しい音源分離手法の独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）[22], [23] を提案する<sup>(注1)</sup>。ILRMA は、FDICA や IVA の自然な拡張であり、周波数毎の分離行列をより高精度に推定するアルゴリズムである。また、MNMF に空間的な近似仮定を導入したモデルが ILRMA と等価であることを理論的に示し、近年の最先端音源分離アルゴリズム（IVA 及び MNMF）の関係性を俯瞰する。音声及び音楽信号を用いた実験により、ILRMA は従来手法よりも高速かつ高精度な音源分離を達成することを示す。

## 2. 従来手法

### 2.1 定式化

音源数と観測チャンネル数をそれぞれ  $N$  及び  $M$  とし、各時間周波数における音源信号、観測信号、分離信号をそれぞれ

$$\mathbf{s}_{ij} = (s_{ij,1} \cdots s_{ij,N})^T \in \mathbb{C}^{N \times 1} \quad (1)$$

$$\mathbf{x}_{ij} = (x_{ij,1} \cdots x_{ij,M})^T \in \mathbb{C}^{M \times 1} \quad (2)$$

$$\mathbf{y}_{ij} = (y_{ij,1} \cdots y_{ij,N})^T \in \mathbb{C}^{N \times 1} \quad (3)$$

と表す。ここで、 $i=1, \dots, I$  は周波数インデックス、 $j=1, \dots, J$  は時間インデックス、 $n=1, \dots, N$  は音源インデックス、 $m=1, \dots, M$  はチャンネルインデックスを示し、 $^T$  は転置を表す。また、各信号の複素スペクトログラム行列を  $\mathbf{S}_n \in \mathbb{C}^{I \times J}$ 、 $\mathbf{X}_m \in \mathbb{C}^{I \times J}$ 、及び  $\mathbf{Y}_n \in \mathbb{C}^{I \times J}$  で表す。これらの行列の要素はそれぞれ  $s_{ij,n}$ 、 $x_{ij,m}$ 、及び  $y_{ij,n}$  に一致する。混合系が線形時不変であり、時間周波数領域での複素瞬時混合で表現できると仮定すると、周波数毎の時不変な複素混合行列  $\mathbf{A}_i = (\mathbf{a}_{i,1} \cdots \mathbf{a}_{i,N}) \in \mathbb{C}^{M \times N}$  ( $\mathbf{a}_{i,n} = (a_{i,n,1} \cdots a_{i,n,M})^T$  は各音源のステアリングベクトル) が定義でき、観測信号を次式で表現できる。

$$\mathbf{x}_{ij} = \mathbf{A}_i \mathbf{s}_{ij} \quad (4)$$

この混合モデルは、時不変混合系の残響時間が STFT の窓長よりも十分短い場合に成立する。このとき、 $M=N$  かつ  $\mathbf{A}_i$  が正則であれば、分離ベクトル  $\mathbf{w}_{i,n} = (w_{i,n,1} \cdots w_{i,n,M})^T$  で構成される分離行列  $\mathbf{A}_i^{-1} \approx \mathbf{W}_i = (\mathbf{w}_{i,1} \cdots \mathbf{w}_{i,N})^H \in \mathbb{C}^{N \times N}$  が存在し、分離信号は次式で与えられる。

$$\mathbf{y}_{ij} = \mathbf{W}_i \mathbf{x}_{ij} \quad (5)$$

ここで、 $^H$  はエルミート転置を示す。優決定条件 BSS では、式 (5) 中の分離行列  $\mathbf{W}_i$  を全周波数において推定することが最終的な目標となる。本稿では、以後常に決定的な系 ( $M=N$ ) を考える。観測チャンネル数が音源数より多い場合は、事前に  $\mathbf{x}_{ij}$  に対して主成分分析を適用し  $M=N$  とする方法が用いられる。

### 2.2 FDICA と IVA

Figure 1 に、FDICA 及び IVA の音源モデルの違いを示す。FDICA [2] は残響による畳み込み混合を分離するために、周波数ビン  $i$  毎の複素時系列信号  $x_{i,1,m} \cdots x_{i,J,m}$  に対して ICA を適用する。このとき、音源信号の生成モデル（統計的な音源モデル）として音源間で独立かつ非ガウスな分布  $p(\mathbf{s}) \approx p(\mathbf{y})$ （複素ラプラス分布等）を仮定する。各周波数での分離行列  $\mathbf{W}_i$  は他の周波数に関わりなく推定されるため、正しい分離信号を得るためには、FDICA の後に周波数毎の分離信号の順序をそろえるパーミュテーション問題を解く必要がある [3]–[7].

IVA [8]–[10] は、パーミュテーション問題に対するエレガントな解法である。FDICA のように周波数毎に独立な ICA を適用するのではなく、全周波数成分を一つにまとめた周波数ベクトル  $\bar{\mathbf{x}}_{j,m} = (x_{j,m} \cdots x_{I,j,m})^T \in \mathbb{C}^{I \times 1}$  を考え、このベクトルの時系列信号  $\bar{\mathbf{x}}_{1,m} \cdots \bar{\mathbf{x}}_{J,m}$  に対して ICA を適用する。このとき、音源周波数ベクトル  $\bar{\mathbf{s}}_{j,n} = (s_{j,n} \cdots s_{I,j,n})^T$  の生成モデルには、 $I$  次元の非ガウスかつ球対称（確率値  $p(\bar{\mathbf{s}})$  がノルム  $\|\bar{\mathbf{s}}\|$  にのみ依存する性質）な多変量分布  $p(\bar{\mathbf{s}}) \approx p(\bar{\mathbf{y}})$  を仮定する。例えば、次式の球対称ラプラス分布がよく用いられる [10].

$$p(\bar{\mathbf{y}}_{j,n}) = \frac{1}{\pi \prod_i \sigma_{i,n}} \exp \left( -\sqrt{\sum_i \left| \frac{y_{ij,n}}{\sigma_{i,n}} \right|^2} \right) \quad (6)$$

ここで、 $\sigma_{i,n}$  はスケールパラメタである。この分布は球対称性を持つため、同一ベクトル内の成分（同一音源の全周波数成分）が高次相関を持つ。従って IVA は、基本周波数とその倍音のように同時に生起する周波数成分を一つの音源にまとめる傾向があり、パーミュテーション問題を回避しつつ分離行列を推定できる。以後、式 (6) に基づく IVA をラプラス IVA と表記する。

音源周波数ベクトル間の独立性  $p(\bar{\mathbf{y}}_{j,1}, \dots, \bar{\mathbf{y}}_{j,N}) = \prod_n p(\bar{\mathbf{y}}_{j,n})$  を仮定すると、IVA の観測信号に対する負対数尤度関数は次式のように得られる<sup>(注2)</sup>。

$$\mathcal{L} = -2J \sum_i \log |\det \mathbf{W}_i| + \sum_{j,n} G(\bar{\mathbf{y}}_{j,n}) \quad (7)$$

(注2) : [10] を含む多くの文献で、式 (7) の第一項の係数 2 が抜けているが、この係数は複素確率変数の変数変換（観測信号  $\mathbf{x}_{ij}$  から分離信号  $\mathbf{y}_{ij}$  への変換）におけるヤコビアンに由来するものである（複素変数のヤコビアンは実数変数の場合の二乗となる）。しかしながら、最適化の過程で分離信号が互いに無相関になると第一項は 0 になるため、この係数の間違いはあまり重要ではなくなる。

(注1) : 本手法は文献 [22], [23] では rank-1 MNMF と呼ばれている。

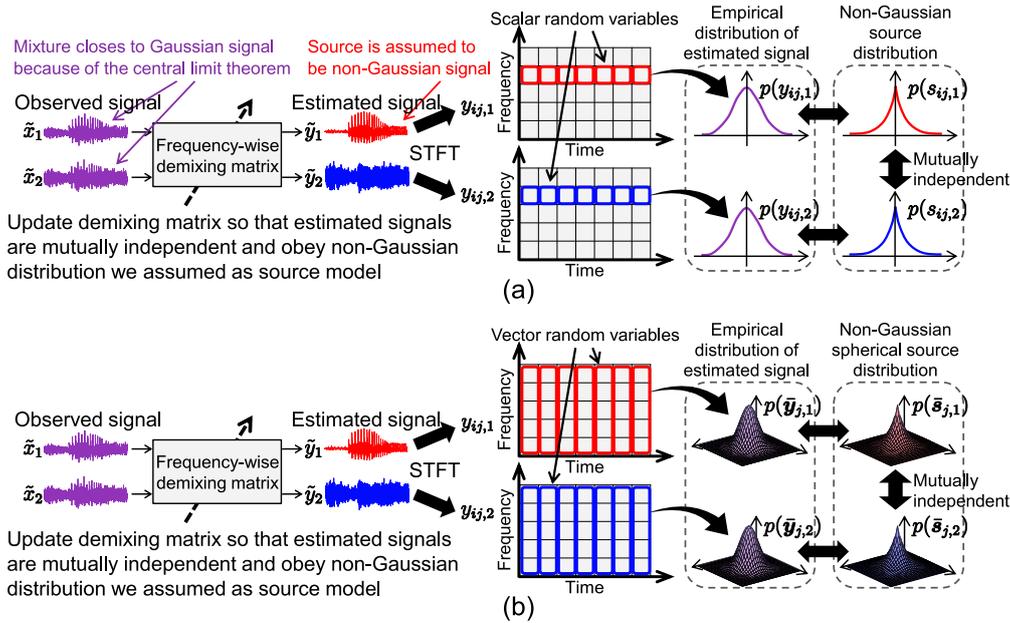


Fig. 1 Source models assumed in (a) FDICA and (b) IVA, where  $\tilde{x}_m$  and  $\tilde{y}_n$  shows time-domain signals of  $x_{ij,m}$  and  $y_{ij,n}$ , respectively.

ここで、 $G(\tilde{y}_{j,n})$  はコントラスト関数と呼ばれ  $G(\tilde{y}_{j,n}) = -\log p(\tilde{y}_{j,n})$  で定義される。ラプラス IVA の場合は  $\sigma_{i,n}=1$  とおくと  $G(\tilde{y}_{j,n}) = \log \pi + \|\tilde{y}_{j,n}\|_2$  となる ( $\|\cdot\|_2$  は  $L_2$  ノルム)。分離信号は  $y_{ij,n} = \mathbf{w}_{i,n}^H \mathbf{x}_{ij}$  として最適化変数  $\mathbf{W}_i$  を含む点に注意する。式 (7) の最適化には、補助関数法 [11] を用いた高速かつ安定な手法が提案されている [12], [13]。なお、FDICA や IVA のような独立性基準の音源分離では、分離信号  $y_{ij,n}$  のスケールが推定できない。最終的な分離信号のスケールは、プロジェクションバック法 [24] で復元できるため、ラプラス IVA のスケールパラメタは常に  $\sigma_{i,n}=1$  においてかまわない。

式 (6) の球対称ラプラス分布以外の優ガウス分布を音源モデルに用いた IVA として、次式のような時変分散  $r_{j,n}$  を持つ原点对称複素ガウス分布に基づく手法が提案されている [25]。

$$p(\tilde{\mathbf{y}}_{1,n}, \dots, \tilde{\mathbf{y}}_{J,n}) = \prod_j p(\tilde{\mathbf{y}}_{j,n}) = \prod_j \frac{1}{\pi r_{j,n}} \exp\left(-\frac{\|\tilde{\mathbf{y}}_{j,n}\|_2^2}{r_{j,n}}\right) \quad (8)$$

時変分散  $r_{j,n}$  は周波数に関しては共通の値である。式 (6) と同様に、式 (8) も  $\tilde{\mathbf{y}}_{j,n}$  のノルムのみ依存するため、球対称な分布である。以後、式 (8) に基づく IVA を時変ガウス IVA と表記する。式 (8) はガウス分布で構成されているが、分散  $r_{j,n}$  が時間的に変動することから、時間フレーム  $j$  に関する周辺分布は優ガウスとなり [26]、IVA の音源モデルとして活用できる。

### 2.3 単一チャネル信号を対象とした NMF

単一チャネルの音響信号を対象とした NMF では、複素スペクトログラム  $\mathbf{B} \in \mathbb{C}^{I \times J}$  を非負化した行列  $|\mathbf{B}|^p$  を別の二つの非負行列  $\mathbf{T} \in \mathbb{R}_{\geq 0}^{I \times K}$  (基底行列) 及び  $\mathbf{V} \in \mathbb{R}_{\geq 0}^{K \times J}$  (アクティベーション行列) の行列積に分解する。ここで、行列に対する絶対値記号は要素毎に絶対値を取った行列を表し、ドット付きの指数は要素毎の累乗を表す。従って  $|\mathbf{B}|^p$  は、 $p=1$  が振幅スペクトログラム、 $p=2$  がパワースペクトログラムに対応する。また、

$K$  は基底数であり、低ランク近似するために  $K \ll \min(I, J)$  と設定される。分解行列  $\mathbf{T}$  及び  $\mathbf{V}$  は、次式の最小化問題の解として推定される。

$$\min_{\mathbf{T}, \mathbf{V}} \mathcal{D}(|\mathbf{B}|^p \|\mathbf{T}\mathbf{V}\|) \quad \text{s.t. } t_{ik}, v_{kj} \geq 0 \quad \forall i, j, k \quad (9)$$

ここで、 $\mathcal{D}(\cdot)$  は 2 つの行列引数の要素間のダイバージェンスの総和、 $t_{ik}$  及び  $v_{kj}$  は行列  $\mathbf{T}$  及び  $\mathbf{V}$  の非負要素、 $k=1, \dots, K$  は基底インデックスを示す。行列  $\mathbf{T}\mathbf{V}$  が低ランクである制約から、基底行列  $\mathbf{T}$  の列ベクトルは、 $|\mathbf{B}|^p$  中に頻出するスペクトルパターンを表し、アクティベーション行列  $\mathbf{V}$  の行ベクトルは各スペクトルパターンの時間的な強度変化を表す。これまでに様々なダイバージェンスの活用が提案されているが、板倉斎藤ダイバージェンスに基づく NMF (Itakura-Saito NMF: ISNMF) によるパワースペクトログラムの低ランク近似には、次に示す特殊な生成モデルによる解釈が存在する [27]。

ISNMF のコスト関数は次式で定義される。

$$\mathcal{L}_{\text{ISNMF}} = \sum_{i,j} \left( \frac{|b_{ij}|^p}{\sum_k t_{ik} v_{kj}} - \log \frac{|b_{ij}|^p}{\sum_k t_{ik} v_{kj}} - 1 \right) \quad (10)$$

ここで、 $b_{ij}$  は  $\mathbf{B}$  の複素要素である。今、 $K$  個の複素スペクトログラム要素  $c_{ij1}, \dots, c_{ijK}$  の生成モデルを原点对称複素ガウス分布と仮定する。

$$p(c_{ijk}) = \frac{1}{\pi r_{ijk}} \exp\left(-\frac{|c_{ijk}|^2}{r_{ijk}}\right) \quad (11)$$

ここで、式 (11) は時間周波数の各グリッド  $i, j$  に対して互いに独立と仮定している。式 (11) は原点对称であるため、確率値は複素成分  $c_{ijk}$  の位相によらず振幅  $|c_{ijk}|$  あるいはパワー  $|c_{ijk}|^2$  のみ依存する。従って、分散はパワーの期待値  $r_{ijk} = \mathbb{E}[|c_{ijk}|^2]$  に対応する。また、行列  $\mathbf{B}$  の複素要素が  $b_{ij} = \sum_k c_{ijk}$  と成分分解できるとすると、ガウス分布の再生性より、 $\mathbf{B}$  に関する生成モデルも原点对称複素ガウス分布で表される。

$$\begin{aligned}
p(\mathbf{B}) &= \prod_{i,j} p(b_{ij}) \\
&= \prod_{i,j} \frac{1}{\pi \sum_k r_{ijk}} \exp\left(-\frac{|b_{ij}|^2}{\sum_k r_{ijk}}\right) \quad (12)
\end{aligned}$$

この事実は、一般に複素数のパワー値の加法性は成り立たないが、上記の生成モデルが成立する場合はパワースペクトル  $|c_{ijk}|^2$  の加法性が期待値の意味で保存されることを意味している<sup>(注3)</sup>。従って、式 (12) において  $r_{ijk} = t_{ik}v_{kj}$  とおくことで、パワースペクトル  $|b_{ij}|^2$  を NMF で  $\sum_k t_{ik}v_{kj}$  に加法分解することの妥当性が正当化される。このモデルの負対数尤度関数は

$$-\log p(\mathbf{B}) = \sum_{i,j} \left( \log \pi + \log \sum_k t_{ik}v_{kj} + \frac{|b_{ij}|^2}{\sum_k t_{ik}v_{kj}} \right) \quad (13)$$

となり、式 (13) の最小化問題は式 (10) のそれと等価である。なお、式 (12) に示す ISNMF の生成モデルは式 (8) と同様に、時間周波数に依存して分散が変動することから、時間フレーム  $j$  に関する周辺分布は優ガウスになる。

### 3. 独立低ランク行列分析

#### 3.1 動機

音声の BSS においては、ラプラス IVA や時変ガウス IVA はブラインドにパーミュテーション問題を解きつつ分離行列  $\mathbf{W}_i$  を比較的高い精度で推定できる。しかしながら、式 (6) や式 (8) で定義される周波数成分間の高次相関は、Fig. 2 (a) に示すように、全周波数成分に対して一律な強度の共変関係しか表現できない。実際の音響信号では、基本周波数とその倍音は強い高次相関を持つが、それ以外の成分や周波数差の大きい成分間の高次相関は弱くなる等、共変関係の度合いは時間周波数のグリッド  $i, j$  に依存して大きく変動する。このような理由から、IVA は特に楽器音信号等のように明確な時間周波数構造を持つ音響信号の分離には適していない。一方で NMF は、Fig. 2 (b) に示すように、限られた数のスペクトルパターンから成る低ランクな時間周波数構造の表現に良く適合しており、離散的な音階や音符の構造を持つ音楽信号等を効率的にモデル化できる。

時変ガウス IVA を用いた手法 [25] では、ユーザから与えられた分離したい音源の時間的なアクティビティを時変分散  $r_{j,n}$  の事前分布として用いている。また、時間と周波数両軸の分散変動を用いる手法として、スペクトル減算法やバイナリマスク法等何らかの手法で一時的に推定した分離信号を、式 (12) の時間周波数の分散  $r_{ij,n}$  の推定値として用いるモデルベース IVA [28] も提案されている。これらの手法では、与えられた分散値（時間周波数成分の共変関係の強度に関する構造）を考慮しながら分離行列  $\mathbf{W}_i$  を推定できる。さらに、正確な時間周波数毎の分散  $r_{ij,n}$ （即ち音源毎のパワースペクトログラム推定値）が分離行列の推定精度を向上させようことを示唆している。

(注3)：複素数  $c_1$  と  $c_2$  に対して  $|c_1 + c_2|^2 = |c_1|^2 + |c_2|^2$  は一般に成立しないが、式 (11) のモデルでは  $E[|b_{ij}|^2] = \sum_k E[|c_{ijk}|^2] = \sum_k r_{ijk}$  が成立する。即ち、 $E[|c_1 + c_2|^2] = E[|c_1|^2] + E[|c_2|^2]$  が成立している。

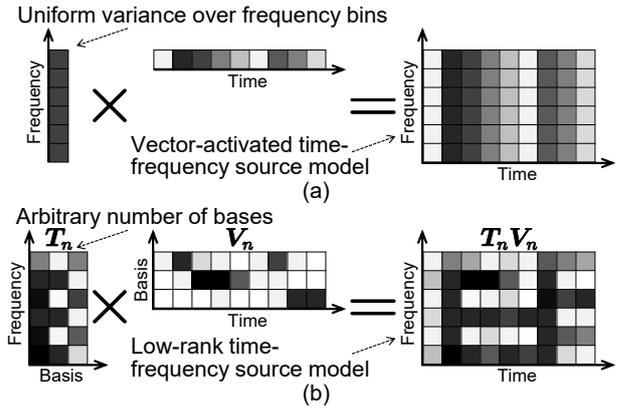


Fig. 2 Structure of source models in (a) Time-varying Gaussian IVA and (b) ILRMA, where grayscale depicts intensity of variance.

この事実に基づいて、本稿では、ISNMF を IVA の音源モデルとして導入することを提案する。本手法では、音源信号の生成モデルとして式 (12) を仮定し、その分散  $r_{ij,n}$  は、Fig. 2 (b) のように、ISNMF で  $r_{ij,n} = \sum_l t_{il,n}v_{lj,n}$  として低ランクな形で分離行列  $\mathbf{W}_i$  と同時にブラインドに推定される。ここで、 $t_{il,n}$  及び  $v_{lj,n}$  は各音源に対応する  $L$  本の基底を持つ基底行列  $\mathbf{T}_n \in \mathbb{R}_{\geq 0}^{I \times L}$  及びそのアクティベーション行列  $\mathbf{V}_n \in \mathbb{R}_{\geq 0}^{L \times J}$  の非負要素であり、 $l = 1, \dots, L$  は新たな基底のインデクスである。従って、本手法は時変ガウス IVA の自然な拡張であり、1本の基底であった時変ガウス IVA の音源モデルを、 $L$ 本の基底で構成される低ランク行列に拡張していると解釈できることから、独立低ランク行列分析 (ILRMA) と呼ぶ。

#### 3.2 尤度関数

ILRMA は ISNMF と同様に、次式の複素ガウス分布を音源信号の生成モデルとして仮定する。

$$\begin{aligned}
p(\bar{\mathbf{y}}_{j,1}, \dots, \bar{\mathbf{y}}_{j,N}) &= \prod_n p(\bar{\mathbf{y}}_{j,n}) \\
&= \prod_{n,i} \frac{1}{\pi r_{ij,n}} \exp\left(-\frac{|y_{ij,n}|^2}{r_{ij,n}}\right) \quad (14)
\end{aligned}$$

ここで、 $r_{ij,n}$  は音源毎の分散であり、 $r_{ij,n} = E[|y_{ij,n}|^2]$  である。式 (14) に基づく観測信号の負対数尤度関数は式 (7) より

$$\mathcal{L} = -2J \sum_i \log |\det \mathbf{W}_i| + \sum_{i,j,n} \left( \frac{|y_{ij,n}|^2}{r_{ij,n}} + \log r_{ij,n} \right) \quad (15)$$

で与えられる。式 (15) の第一項及び第二項はモデルベース IVA のコスト関数に対応し、第二項及び第三項は分散を  $r_{ij,n} = \sum_l t_{il,n}v_{lj,n}$  とおくと、ISNMF のコスト関数 (9) における  $b_{ij} = y_{ij,n}$  の場合に等しいことがわかる。但し、分散  $r_{ij,n}$  の低ランク分解については、次式の2通りを考える。

$$r_{ij,n} = \sum_l t_{il,n}v_{lj,n} \quad (16)$$

$$r_{ij,n} = \sum_k z_{nk}t_{ik}v_{kj} \quad (17)$$

ここで、 $z_{nk} \in [0, 1]$  は基底行列  $\mathbf{T}$  中の全音源を表現する  $K$  本

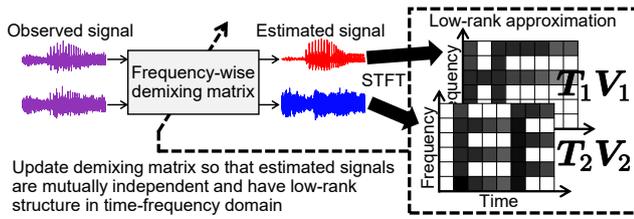


Fig. 3 Principle of source separation based on ILRMA.

の基底を、 $N$  個の音源に振り分ける潜在変数（分割関数）であり、 $\sum_n z_{nk} = 1$  を満たす。式 (16) を用いる場合、各音源の基底数は共通の  $L$  本となる。しかしながら、ボーカルとドラムの混合信号等のように、音源毎のスペクトログラムのランクが極端に異なる場合は最適な基底数が音源毎に異なる。そこで、式 (17) のように分割関数  $z_{nk}$  を導入することで、 $K$  本の基底を  $N$  個の音源に適応的に振り分けることができる。

ILRMA における音源分離の原理を Fig. 3 に示す。 $\mathbf{W}_i$  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  $|\mathbf{Y}_n|^2$  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  $\mathbf{W}_i$  を推定する。混合前の各音源のパワースペクトログラム  $|\mathbf{S}_n|^2$  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  $|\mathbf{X}_n|^2$  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問題を避けつつ、互いに独立となる分離信号を推定することができる。

### 3.3 更新式の導出

ICA や IVA における分離行列  $\mathbf{W}_i$  の最適化では、補助関数法に基づく手法が提案されており、従来の自然勾配法よりも高速かつ安定に収束することが示されている [12], [13]。式 (15) の  $\mathbf{W}_i$  に関する偏微分は、ラプラス IVA の補助関数の偏微分と同じであることから、分離ベクトル  $\mathbf{w}_{i,n}$  の更新式は反復射影法 (iterative projection: IP) [13] を用いて次式で得られる。

$$\mathbf{U}_{i,n} = \frac{1}{J} \sum_j \frac{1}{r_{ij,n}} \mathbf{x}_{ij} \mathbf{x}_{ij}^H \quad (18)$$

$$\mathbf{w}_{i,n} \leftarrow (\mathbf{W}_i \mathbf{U}_{i,n})^{-1} \mathbf{e}_n \quad (19)$$

$$\mathbf{w}_{i,n} \leftarrow \mathbf{w}_{i,n} (\mathbf{w}_{i,n}^H \mathbf{U}_{i,n} \mathbf{w}_{i,n})^{-\frac{1}{2}} \quad (20)$$

ここで、 $\mathbf{e}_n \in \mathbb{R}_{[0,1]}^{N \times 1}$  は  $n$  番目の要素が 1、他要素が 0 のベクトルである。分離ベクトル更新後は、分離信号を次式で更新する。

$$y_{ij,n} \leftarrow \mathbf{w}_{i,n}^H \mathbf{x}_{ij} \quad (21)$$

音源モデルの更新式は、分割関数の有無に応じて二種類導出される。分割関数の無い式 (16) を用いた場合、式 (15) の  $t_{il,n}$  及び  $v_{lj,n}$  に関する偏微分は ISNMF のコスト関数 (9) の偏微分と同じである。従って、 $t_{il,n}$  及び  $v_{lj,n}$  は次の ISNMF における乗算型反復更新式で最適化できる。

$$t_{il,n} \leftarrow t_{il,n} \sqrt{\frac{\sum_j |y_{ij,n}|^2 v_{lj,n} (\sum_{l'} t_{il',n} v_{lj,n})^{-2}}{\sum_j v_{lj,n} (\sum_{l'} t_{il',n} v_{lj,n})^{-1}}} \quad (22)$$

$$v_{lj,n} \leftarrow v_{lj,n} \sqrt{\frac{\sum_i |y_{ij,n}|^2 t_{il,n} (\sum_{l'} t_{il',n} v_{lj,n})^{-2}}{\sum_i t_{il,n} (\sum_{l'} t_{il',n} v_{lj,n})^{-1}}} \quad (23)$$

$t_{il,n}$  及び  $v_{lj,n}$  の更新後は、推定分散を次式で更新する。

$$r_{ij,n} \leftarrow \sum_l t_{il,n} v_{lj,n} \quad (24)$$

また、分割関数を含む式 (17) を用いた場合は、 $z_{nk}$ 、 $t_{ik}$ 、及び  $v_{kj}$  の更新式を補助関数法で導出できる [23]。

$$z_{nk} \leftarrow z_{nk} \sqrt{\frac{\sum_{i,j} |y_{ij,n}|^2 t_{ik} v_{kj} (\sum_{k'} z_{nk'} t_{ik'} v_{kj'})^{-2}}{\sum_{i,j} t_{ik} v_{kj} (\sum_{k'} z_{nk'} t_{ik'} v_{kj'})^{-1}}} \quad (25)$$

$$z_{nk} \leftarrow \frac{z_{nk}}{\sum_{n'} z_{n'k}} \quad (26)$$

$$t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_{j,n} |y_{ij,n}|^2 z_{nk} v_{kj} (\sum_{k'} z_{nk'} t_{ik'} v_{kj'})^{-2}}{\sum_{j,n} z_{nk} v_{kj} (\sum_{k'} z_{nk'} t_{ik'} v_{kj'})^{-1}}} \quad (27)$$

$$v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_{i,n} |y_{ij,n}|^2 z_{nk} t_{ik} (\sum_{k'} z_{nk'} t_{ik'} v_{kj'})^{-2}}{\sum_{i,n} z_{nk} t_{ik} (\sum_{k'} z_{nk'} t_{ik'} v_{kj'})^{-1}}} \quad (28)$$

ここで、式 (26) は  $\sum_n z_{nk} = 1$  を保証するために計算される。 $z_{nk}$ 、 $t_{ik}$ 、及び  $v_{kj}$  の更新後は、推定分散を次式で更新する。

$$r_{ij,n} \leftarrow \sum_k z_{nk} t_{ik} v_{kj} \quad (29)$$

以上より、これらの分離行列と音源モデルの更新式を交互に反復することで、式 (15) を最小化できる。しかしながら、 $\mathbf{W}_i$  と  $r_{ij,n}$  はいずれも分離信号  $y_{ij,n}$  のスケールを決定できるため任意性が存在し、数値計算が不安定になる。そこで、次式の正規化を反復毎に行うことで、この問題を避ける。

$$\mathbf{w}_{i,n} \leftarrow \mathbf{w}_{i,n} \lambda_n^{-1}, y_{ij,n} \leftarrow y_{ij,n} \lambda_n^{-1}, r_{ij,n} \leftarrow r_{ij,n} \lambda_n^{-2} \quad (30)$$

音源モデルの正規化は、分割関数がない場合

$$t_{il,n} \leftarrow t_{il,n} \lambda_n^{-2} \quad (31)$$

であり、分割関数を含む場合

$$t_{ik} \leftarrow t_{ik} \sum_n z_{nk} \lambda_n^{-2} \quad (32)$$

$$z_{nk} \leftarrow \frac{z_{nk} \lambda_n^{-2}}{\sum_{n'} z_{n'k} \lambda_{n'}^{-2}} \quad (33)$$

である。ここで、 $\lambda_n$  は任意の正規化係数であるが、 $\lambda_n = [(IJ)^{-1} \sum_{i,j} |y_{ij,n}|^2]^{1/2}$  等が適当である。これらの正規化はコスト関数 (15) の値を変えない点に注意する。全変数の推定後は次式のプロジェクションバック法 [24] を適用する。

$$\hat{\mathbf{y}}_{ij,n} = \mathbf{W}_i^{-1} (\mathbf{e}_n \circ \mathbf{y}_{ij}) \quad (34)$$

ここで、 $\hat{\mathbf{y}}_{ij,n} = (\hat{y}_{ij,n1} \cdots \hat{y}_{ij,nM})^T$  は推定された多チャネル分離信号であり、 $\circ$  は要素毎の積を示す。なお、一連の最適化アルゴリズムの具体的な実装は [29] で公開されている。

### 3.4 ILRMA, IVA, 及び MNMF の関係性

MNMF [19]–[21] は NMF を多チャネル観測信号用に拡張した音源分離手法であり、特に Sawada et al. による定式化 [21]

が最も一般化されている。本手法は、各時間周波数の瞬時空間相関行列  $\mathbf{X}_{ij} = \mathbf{x}_{ij} \mathbf{x}_{ij}^H$  を観測信号とし、次のモデルを仮定する。

$$\hat{\mathbf{X}}_{ij} = \sum_n r_{ij,n} \mathbf{R}_{i,n} \quad (35)$$

ここで、 $\mathbf{R}_{i,n} \in \mathbb{C}^{M \times M}$  は音源毎の空間相関行列 [18] であり、音源からマイクロホンまでの伝達系及び空間的な拡散を表す時不変な特徴量である。また、分散  $r_{ij,n}$  は時変な変動成分、即ち音源毎のパワースペクトログラムであり、式 (17) の形で低ランク近似される。MNMF では、 $\hat{\mathbf{X}}_{ij}$  を多変量複素ガウス分布の共分散行列として多チャンネル観測信号の生成モデル [18], [21] を考える。そのときの最尤推定は次式の最小化問題となる。

$$\mathcal{L}_{\text{MNMF}} = \sum_{i,j} [\log \det \hat{\mathbf{X}}_{ij} + \text{tr}(\mathbf{X}_{ij} \hat{\mathbf{X}}_{ij}^{-1})] \quad (36)$$

MNMF では、音源毎の空間相関行列  $\mathbf{R}_{i,n}$  と全音源の低ランクなパワースペクトログラム  $\mathbf{TV}$  を同時に推定し、 $\mathbf{TV}$  を分割関数  $z_{nk}$  で選別することで音源分離を達成する。空間相関行列  $\mathbf{R}_{i,n}$  はフルランク行列として推定されるが、ここに混合系 (4) の近似モデルを導入すると、伝達系は全てステアリングベクトル  $\mathbf{a}_{i,n}$  で記述でき、空間相関行列はランク 1 行列

$$\mathbf{R}_{i,n} = \mathbf{a}_{i,n} \mathbf{a}_{i,n}^H \quad (37)$$

に制約される。式 (37) を式 (35) に代入すると、

$$\begin{aligned} \hat{\mathbf{X}}_{ij} &= \sum_n r_{ij,n} \mathbf{a}_{i,n} \mathbf{a}_{i,n}^H \\ &= \mathbf{A}_i \mathbf{D}_{ij} \mathbf{A}_i^H \end{aligned} \quad (38)$$

$$\mathbf{D} = \text{diag}(r_{ij,1}, \dots, r_{ij,N}) \quad (39)$$

と表現でき、これをコスト関数 (36) に代入すると、次式を得る。

$$\begin{aligned} \mathcal{L}_{\text{MNMF}} &= \sum_{i,j} [\log \det \mathbf{A}_i \mathbf{D}_{ij} \mathbf{A}_i^H + \text{tr}(\mathbf{X}_{ij} \mathbf{A}_i^{-H} \mathbf{D}_{ij}^{-1} \mathbf{A}_i^{-1})] \\ &= \sum_{i,j} [\log (\det \mathbf{A}_i) (\det \mathbf{D}_{ij}) (\det \mathbf{A}_i)^* \\ &\quad + \text{tr}(\mathbf{W}_i^{-1} \mathbf{y}_{ij} \mathbf{y}_{ij}^H \mathbf{W}_i^{-H} \mathbf{W}_i^H \mathbf{D}_{ij}^{-1} \mathbf{W}_i)] \\ &= \sum_{i,j} [\log |\det \mathbf{A}_i|^2 + \log \det \mathbf{D}_{ij} \\ &\quad + \text{tr}(\mathbf{W}_i \mathbf{W}_i^{-1} \mathbf{y}_{ij} \mathbf{y}_{ij}^H \mathbf{D}_{ij}^{-1})] \\ &= -2J \sum_i \log |\det \mathbf{W}_i| \\ &\quad + \sum_{i,j} \left[ \log \prod_n r_{ij,n} + \text{tr}(\mathbf{y}_{ij} \mathbf{y}_{ij}^H \mathbf{D}_{ij}^{-1}) \right] \\ &= -2J \sum_i \log |\det \mathbf{W}_i| + \sum_{i,j,n} \left( \frac{|y_{ij,n}|^2}{r_{ij,n}} + \log r_{ij,n} \right) \end{aligned} \quad (40)$$

ここで、 $\mathbf{x}_{ij} = \mathbf{W}_i^{-1} \mathbf{y}_{ij}$  及び  $\mathbf{W}_i = \mathbf{A}_i^{-1}$  を用いた。式 (40) は ILRMA のコスト関数 (15) と一致することから、MNMF に対して空間相関行列のランクが 1 となる制約 (混合系 (4) の

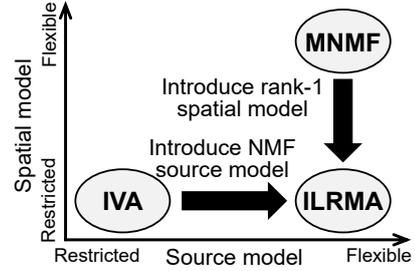


Fig. 4 Relationship between ILRMA, IVA, and MNMF from viewpoint of flexibility of spatial and source models.

仮定)を導入したモデルが、本質的に ILRMA と等価であることを示している。しかしながら、MNMF は混合系を推定するのに対して ILRMA は分離系を推定する双対な関係となっており、この違いが IP に基づく効率的な最適化を実現している。

ILRMA, IVA, 及び MNMF の関係性を Fig. 4 に示す。3.1 節で述べたように、IVA のベクトル音源モデルを NMF による低ランク行列音源モデルに拡張した手法が ILRMA であり、それは MNMF に対して混合系 (4) を仮定したモデルと等価である。この事実は、独立に発展してきた IVA 及び MNMF の両理論が、実は密接に関連していることを初めて明らかにしている。

## 4. 音声及び音楽信号の音源分離実験

### 4.1 実験条件

ILRMA による音源分離の妥当性を示すために、音声及び音楽信号を用いた音源分離実験を示す。比較手法として、補助関数法に基づくラプラス IVA [13], Ozerov et al. の MNMF [20], Sawada et al. の MNMF [21], 分割関数無し ILRMA, 分割関数有りの ILRMA の 5 手法を用いた。音声信号の実験で用いるデータは、SiSEC2011 [30] の UND タスクで公開されている実録音信号 (dev1 及び dev2) である。このデータセットは男性及び女性の 12 個の音声信号を含み、残響時間  $T_{60}$  は 130 ms 及び 250 ms である。このデータセットは劣決定条件 BSS を想定しているため、本実験では常に 1 番目及び 2 番目の音声信号のみを用いて  $M = N = 2$  となる混合信号を生成した。音楽信号の実験で用いるデータは、同じく SiSEC2011 の MUS タスクで公開されている 4 楽曲を、RWCP データベース [31] に収録されている E2A ( $T_{60} = 300$  ms) 及び JR2 ( $T_{60} = 470$  ms) のインパルス応答で畳み込み混合して生成した。音声信号は音楽信号よりも時間周波数構造が高ランクなため、正確なモデル化にはより多くの基底数が必要であるが、ILRMA での音声信号分離では、基底数が多いと分離性能が劣化することが確認されている [23]。従って、音声信号の分離では、ILRMA の基底数を  $L = 2$  及び  $K = 4$  とし、音楽信号の分離では  $L = 30$  及び  $K = 60$  とした。客観評価尺度には信号対歪み比 (SDR) [32] の改善量を用いた。その他の実験条件は [23] に記載されている。

### 4.2 実験結果

音声信号及び音楽信号の典型的な分離結果例を Figs. 5 及び 6 に示す。また、Tables 1 及び 2 は、全観測信号の平均分離性能をまとめている。なお、これらの結果は変数の乱数初期値を変えて 10 回試行した際の平均及び標準偏差である。この結果

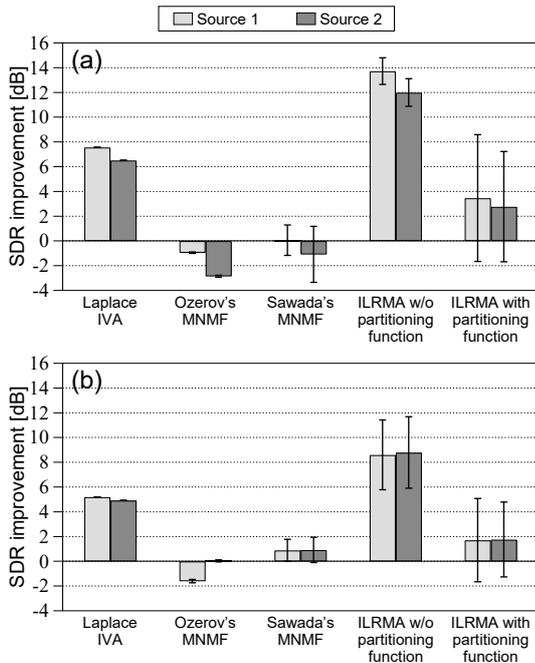


Fig. 5 Examples of average SDR improvements for female speech, where reverberation time is (a) 130 ms and (b) 250 ms.

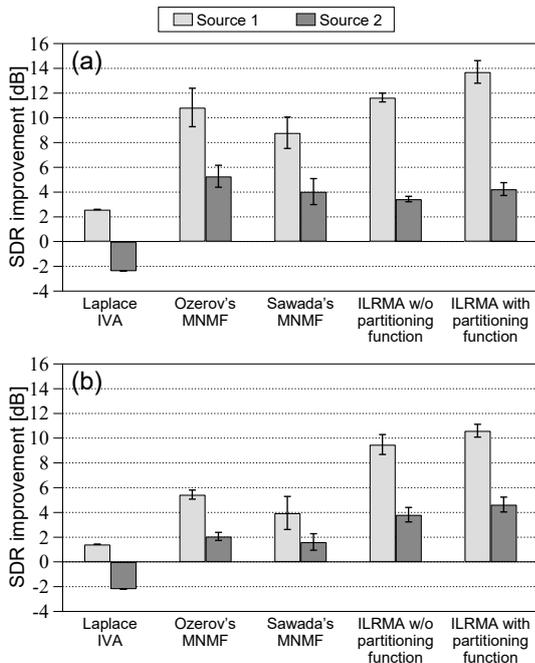


Fig. 6 Examples of average SDR improvements for music signal with impulse response (a) E2A and (b) JR2.

より、ラプラス IVA はシンプルな音源モデルに起因して音楽信号の分離には不向きであることが確認できる。2種の MNMF は、低ランクな音源モデルを用いることで、音楽信号を高精度で分離しているが、音声信号に対しては十分な分離精度が得られていない。一方で、分割関数無し of ILRMA は、音声信号に対して高精度かつ頑健な分離を達成しており、分割関数有りの ILRMA は音楽信号に対するさらなる性能改善に有効であることが確認できる。

さらに、Fig. 7 は、各手法の反復更新に対する SDR 改善量の収束例と実際の計算時間例を示している。計算には MATLAB 8.3 (64 bit) 環境で、Intel Core i7-4790 (3.6 GHz) の

Table 1 Total average SDR improvements (dB) of speech signals

Recording conditions (rev. time and mic. spacing)	Laplace IVA	Ozerov's MNMF	Sawada's MNMF	ILRMA w/o partitioning function	ILRMA with partitioning function
130 ms and 1 m	2.98	1.35	0.68	<b>11.91</b>	4.88
130 ms and 5 cm	2.86	2.13	1.13	<b>8.97</b>	3.48
250 ms and 1 m	2.03	0.49	0.48	<b>7.34</b>	2.09
250 ms and 5 cm	2.43	0.91	0.47	<b>6.43</b>	1.91

Table 2 Total average SDR improvements (dB) of music signals

Impulse response	Laplace IVA	Ozerov's MNMF	Sawada's MNMF	ILRMA w/o partitioning function	ILRMA with partitioning function
E2A	5.72	5.73	10.32	<b>12.29</b>	<b>12.29</b>
JR2	1.77	2.37	6.11	6.62	<b>7.40</b>

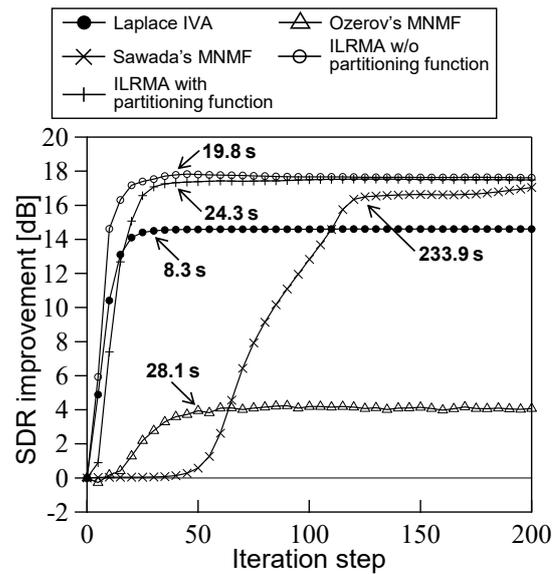


Fig. 7 SDR convergence and examples of actual calculation time with impulse response E2A, where signal length is 18.6 s.

CPU を用いている。ラプラス IVA と ILRMA は MNMF よりも高速な収束を示しており、計算時間も比較的短い。Sawada et al. の MNMF は、1 回の更新において  $IN$  回のサイズ  $2M$  行列の固有値分解や  $IJ$  回のサイズ  $M$  行列の逆行列演算が必要であり、より多くの計算時間が必要となる。

## 5. まとめ

本稿では、IVA と MNMF を融合した新しい BSS の ILRMA を提案した。ILRMA は、IVA の音源モデルをベクトルから低ランク行列に拡張した手法であり、効率的な最適化とより高精度な音源分離を達成している。また、MNMF に対して周波数領域での複素瞬時混合仮定を導入したモデルと ILRMA が本質的に等価であることを示し、IVA 及び MNMF の二大音源分離理論の関係性を理論的に明らかにした。本稿に記載した内容以外にも、観測チャンネル数が十分多い状況における残響のモデル化手法 [33]、適応的なスパース正則化を付与した ILRMA [34]、STFT の窓長に関する実験的な解析 [35]、生成モデルを複素 Student's  $t$  分布へと一般化した ILRMA [36]、分離行列と音源モデルの最適化速度を調整し初期値への頑健性を向上させるパラメトリックな最適化手法 [37] 等が検討されている。

謝辞 本研究は、セコム科学技術振興財団及び JSPS 科研費

## 文 献

- [1] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol.36, no.3, pp.287–314, 1994.
- [2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol.22, pp.21–34, 1998.
- [3] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," *Proc. ICASSP*, vol.5, pp.3140–3143, 2000.
- [4] N. Murata, S. Ikeda and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol.41, no.1–4, pp.1–24, 2001.
- [5] 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. ASLP*, vol.12, no.5, pp.530–538, 2004.
- [6] H. Sawada, R. Mukai, S. Araki and S. Makino, "Convulsive blind source separation for more than two sources in the frequency domain," *Proc. ICASSP*, pp.III-885–III-888, 2004.
- [7] 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. ASLP*, vol.14, no.2, pp.666–678, 2006.
- [8] A. Hiroe, "Solution of permutation problem in frequency domain ICA using multivariate probability density functions," *Proc. ICA*, pp.601–608, 2006.
- [9] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: an extension of ICA to multivariate components," *Proc. ICA*, pp. 165–172, 2006.
- [10] T. Kim, H. T. Attias, S.-Y. Lee and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. ASLP*, vol.15, no.1, pp.70–79, 2007.
- [11] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.
- [12] N. Ono and S. Miyabe, "Auxiliary-function-based independent component analysis for super-Gaussian sources," *Proc. LVA/ICA*, pp.165–172, 2010.
- [13] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," *Proc. WASPAA*, pp.189–192, 2011.
- [14] 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.
- [15] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino and S. Sagayama, "Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints," *Proc. ICASSP*, pp.5365–5368, 2012.
- [16] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, and K. Kondo, "Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties," *IEICE Trans. Fundamentals*, vol. E97-A, no. 5, pp. 1113–1118, 2014.
- [17] D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo, and S. Nakamura, "Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration," *IEEE/ACM Trans. ASLP*, vol. 23, no. 4, pp. 654–669, 2015.
- [18] N. Q. K. Duong, E. Vincent, and R. Gribonval, "Underdetermined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Trans. ASLP*, vol. 18, no. 7, pp. 1830–1840, 2010.
- [19] H. Kameoka, T. Yoshioka, M. Hamamura, J. Le Roux, and K. Kashino, "Statistical model of speech signals based on composite autoregressive system with application to blind source separation," *Proc. LVA/ICA*, pp. 245–253, 2010.
- [20] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. ASLP*, vol. 18, no. 3, pp. 550–563, 2010.
- [21] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complex-valued data," *IEEE Trans. ASLP*, vol. 21, no. 5, pp. 971–982, 2013.
- [22] D. Kitamura, N. Ono, H. Sawada, H. Kameoka and H. Saruwatari, "Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model," *Proc. ICASSP*, pp.276–280, 2015.
- [23] 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. ASLP*, vol. 24, no. 9, pp. 1626–1641, 2016.
- [24] K. Matsuoaka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proc. ICA*, pp. 722–727, 2001.
- [25] T. Ono, N. Ono, and S. Sagayama, "User-guided independent vector analysis with source activity tuning," *Proc. ICASSP*, pp. 2417–2420, 2012.
- [26] N. Ono, "Auxiliary-function-based independent vector analysis with power of vector-norm type weighting functions," *Proc. APSIPA ASC*, 2012.
- [27] C. Févotte, N. Bertin and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computation*, vol.21, no.3, pp.793–830, 2009.
- [28] A. R. López, N. Ono, U. Remes, K. Palomáki, and M. Kurimo, "Designing multichannel source separation based on single-channel source separation," *Proc. ICASSP*, pp. 469–473, 2015.
- [29] D. Kitamura, "Algorithms for independent low-rank matrix analysis," [Online]. <http://d-kitamura.net/pdf/misc/AlgorithmsForIndependentLowRankMatrixAnalysis.pdf>
- [30] S. Araki, F. Nesta, E. Vincent, Z. Koldovsky, G. Nolte, A. Ziehe and A. Benichoux, "The 2011 signal separation evaluation campaign (SiSEC2011):-audio source separation," *Proc. LVA/ICA*, pp.414–422, 2012.
- [31] 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," *Proc. LREC*, pp.965–968, 2000.
- [32] E. Vincent, R. Gribonval and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. ASLP*, vol.14, no.4, pp.1462–1469, 2006.
- [33] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Relaxation of rank-1 spatial constraint in overdetermined blind source separation," *Proc. EUSIPCO*, pp. 1271–1275, 2015.
- [34] 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," *Proc. ICASSP*, pp. 21–25, 2017.
- [35] D. Kitamura, N. Ono, and H. Saruwatari, "Experimental analysis of optimal window length for independent low-rank matrix analysis," *Proc. EUSIPCO*, pp. 1210–1214, 2017.
- [36] S. Mogami, D. Kitamura, Y. Mitsui, N. Takamune, H. Saruwatari, and N. Ono, "Independent low-rank matrix analysis based on complex Student's *t*-distribution for blind audio source separation," *Proc. MLSP*, 2017.
- [37] Y. Mitsui, D. Kitamura, N. Takamune, H. Saruwatari, Y. Takahashi, and K. Kondo, "Independent low-rank matrix analysis based on parametric majorization-equalization algorithm," *Proc. CAMSAP*, 2017 (in press).