一般社団法人 人工知能学会
Japanese Society for
Artificial Intelligence

人工知能学会資料
JSAI Technical Report
SIG-Challenge-049-6 (11/25)

ブラインド音源分離
〜時空間スモールデータの非ガウス・低ランクモデリングとその最適化の数理〜

Blind Source Separation – Non-Gaussian and Low-Rank Modeling for Time-Spatial Small Data and Its Optimization –

猿渡　洋
Hiroshi SARUWATARI
東京大学・大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo
hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

Abstract

ブラインド音源分離（BSS）技術は，音の空間伝搬及び時間周波数スペクトログラムをモデリングする音響分野において独自の発展を遂げてきた．本稿では，その歴史と潮流について振り返る．特に，優決定 BSS 問題における代表的なアルゴリズムとして周波数領域独立成分分析・独立ベクトル分析及び独立低ランク行列分析を取り上げ，それらが何をモデリングしどのような数理アルゴリズムによって信号推定するのか考察する．また，空間モデルの推定に関し，多チャネル非負値行列因子分解に見られるような「生成モデル型」と BSS で用いられる「分離モデル型」の違い・得失に関しても解説を行い，最適化パラメータのドメイン変更によって性能が大きく改善されていった歴史を紹介する．

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元信号を推定する信号処理技術である．優決定条件 (音源数 ≤ 観測チャネル数) における BSS では，独立成分分析 (independent component analysis: ICA) [1] に基づく手法が主流であり，盛んに研究されてきた [2]–[7]．一方，モノラル信号等を対象とした劣決定条件 (音源数 > 観測チャネル数) 下では，非負値行列因子分解 (nonnegative matrix factorization: NMF) [8] を応用した手法が注目を集めており，多チャネル信号用に拡張した多チャネル NMF (multichannel NMF: MNMF) [9]–[11] も提案されている．

優決定条件における周波数領域 ICA (frequency-domain ICA: FDICA) や ICA の多変量モデルである独立ベクトル分析 (independent vector analysis: IVA) [12]–[14] では，時間周波数領域での線形時不変混合を仮定する．この仮定は，多チャネル観測信号の空間相関行列のランクが 1 になることから,「ランク 1 空間近似」と呼ばれ，複素スペクトログラムの各時間フレーム内で複数の音源が瞬時混合されているという混合系を想定したものである．このような仮定は，各音源から各マイクロフォンまでのインパルス応答が，短時間フーリエ変換の窓関数と比べて十分に短い場合に成立する．Kitamura らは近年，IVA を拡張し，任意ランクの非負値行列積で音源スペクトログラムをモデリングする独立低ランク行列分析（independent low-rank matrix analysis: ILRMA） [15] を提案しており，従来の手法を凌駕する分離性能と高速な最適化アルゴリズムを実現している．本稿ではこれら優決定 BSS の歴史を踏まえ，ランク 1 空間近似を用いた 3 つの代表的な BSS アルゴリズム (FDICA, IVA, ILRMA) を取り上げ，それぞれの手法が仮定する音源モデルと空間モデル及びその最適化手法について概観する．

2　BSS 概観：何をモデリングするのか？

2.1　ランク 1 空間近似

音源数と観測チャネル数をそれぞれ N, M とし，各時間周波数における多チャネル音源信号，多チャネル観測信号，分離信号をそれぞれ

$$\boldsymbol{s}_{ij} = (s_{ij,1} \cdots s_{ij,N})^{\mathrm{T}} \tag{1}$$
$$\boldsymbol{x}_{ij} = (x_{ij,1} \cdots x_{ij,M})^{\mathrm{T}} \tag{2}$$
$$\boldsymbol{y}_{ij} = (y_{ij,1} \cdots y_{ij,N})^{\mathrm{T}} \tag{3}$$

と表す (要素はすべて複素数)．ここで，$i = 1, \cdots, I$ は周波数インデックス，$j = 1, \cdots, J$ は時間インデックス，$n = 1, \cdots, N$ は音源インデックス，$m = 1, \cdots, M$ はチャネルインデックスを示し，$^{\mathrm{T}}$ は転置を表す．

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，各時間フレームにおいて周波数毎の複素混合行列 $\boldsymbol{A}_i = (\boldsymbol{a}_{i,1} \cdots \boldsymbol{a}_{i,N})$ ($\boldsymbol{a}_{i,n}$ は各

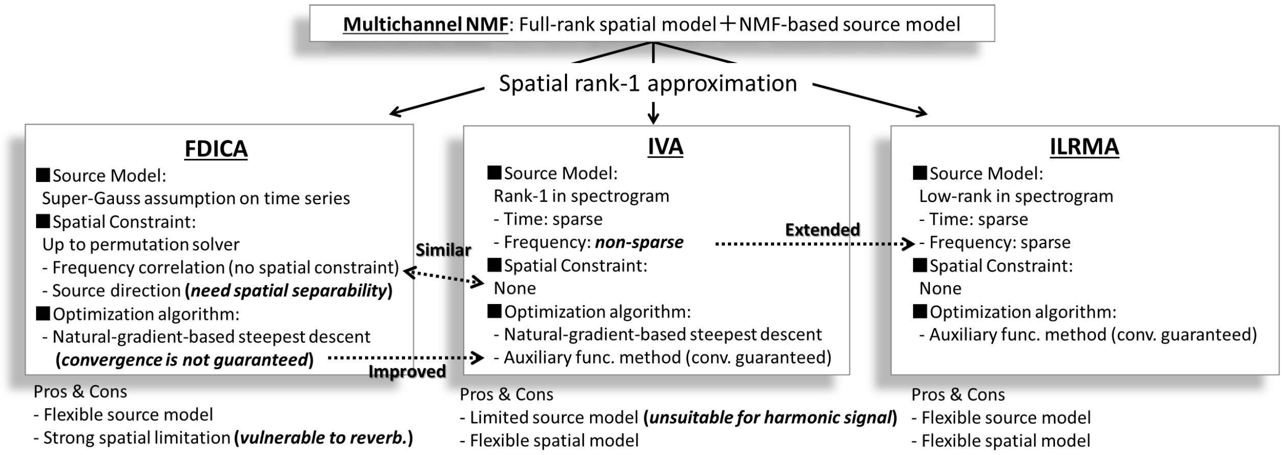

図 1: Overview and relationship between typical acoustic BSS algorithms.

音源のステアリングベクトル) が定義でき，多チャネル観測信号を次式で表現できる．

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

このとき，観測信号 $\boldsymbol{x}_{ij}$ に含まれる各音源の空間相関行列のランクは必ず 1 となる [16]．すなわち,「混合系が線形時不変かつ複素瞬時混合」という仮定は，ランク 1 空間近似と等価であり，各音源の伝達系が周波数毎の時不変なステアリングベクトル $\boldsymbol{a}_{i,n}$1 本で表現できるという近似を与えている．

式 (4) の混合系において $\boldsymbol{A}_i$ をフルランクとすれば，分離ベクトル $\boldsymbol{w}_{i,n}$ で表現される分離行列 $\boldsymbol{W}_i = (\boldsymbol{w}_{i,1} \cdots \boldsymbol{w}_{i,N})^{\mathrm{H}}$ が存在し，分離信号は次式となる．

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

但し，$^{\mathrm{H}}$ はエルミート転置を示す．

ランク 1 空間近似を用いた BSS では，式 (5) 中の分離行列 $\boldsymbol{W}_i$ を推定することが最終的な目標となる．様々なアルゴリズムが提案されているが，大きく分けて「複素時系列の非ガウス性に着目した FDICA」,「FDICA を多変量モデルへ拡張した IVA」及び「音源スペクトログラムを時間周波数低ランク非負値行列としてモデリングする ILRMA」の 3 種類が代表的である（Fig. 1 に相関図を示す）．以降の節では，これらのアルゴリズムについて解説する．

2.2 FDICA の仮定する音源及び空間モデル

FDICA における最適化問題は，以下のコストを最小化する分離行列 $\boldsymbol{W}_i$ を見つける問題に帰着する．これは，音源時系列（周波数ビン i 毎の時系列）の確率密度関数 $p(\cdot)$ をモデルとする対数尤度関数の負値である．

$$Q_{\mathrm{FDICA}} = -2\log|\det \boldsymbol{W}_i| - \frac{1}{J}\sum_{j}\sum_{n} \log p(y_{ij,n}) \tag{6}$$

本式と「非ガウス性」の関連について簡単に解説してみる．本式右辺第一項は音源間の関連度合を表す結合エントロピーを制御し，主に分離の精度に依存する．一方，右辺第二項は個々の音源に関する周辺エントロピー和を制御し，これを最小化するということは「より非ガウスな分離信号へ帰着させる」ことを意味する．つまり，FDICA は，音源の非ガウス信号モデリングと言うことが出来る．

時間周波数領域で各周波数成分に独立な ICA を施す FDICA では，パーミュテーション問題の解決が極めて重要であり，これまでに多くの手法が提案されてきた．代表的なパーミュテーション問題の解決法の一つとして，周波数成分間の相関を用いる手法 [4] がある．これは，後述の IVA と本質的に等価であり，IVA が分離行列の推定と同時にパーミュテーションを解くのに対して，本手法はポスト処理としてパーミュテーションを解いている．もう一つの代表的な解決法は，音源の到来方向 (direction of arrival: DOA) の違いを活用する手法 [3] である．本手法では，推定した周波数毎の分離行列から各音源のステアリングベクトルを逆算し，位相差及び振幅比から DOA を算出して音源毎にクラスタリングすることでパーミュテーションを解いている．この手法の音源モデルは，IVA や周波数間の相関を用いたパーミュテーション解決法とは異なり，時間方向の非ガウス性制約のみである．一方で，FDICA で推定した DOA をパーミュテーション解決に用いる為，空間モデルに関する制約を与えている．複数の音源位置が空間的に接近した場合や残響による拡散の影響が強い場合等，音源 DOA のクラスタリングが困難な状況では分離性能が劣化する．

2.3 IVA の仮定する音源及び空間モデル

IVA は複数の周波数成分を同時に取り扱う為に，ICA を多変量モデルへと拡張した手法である．周波数成分間の高次相関を考慮することで，FDICA におけるパーミュテーション問題 [3, 4] を解決しながら同時に分離行列 $\boldsymbol{W}_i$ を推定する．ICA が非ガウス性の分布を仮定するように，IVA も非ガウスな多変量分布を仮定する．このとき，変数間の高次相関を考慮する為に，球対称の多変量分布を仮定

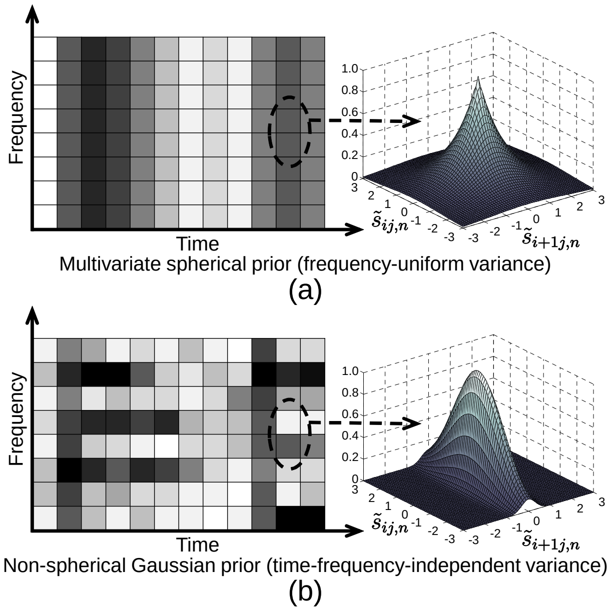

図 2: Illustration of source models (model spectrograms) for one source in (a) IVA and (b) ILRMA, where gray scale of each time-frequency slot indicates value of variance and $\tilde{s}$ denotes only real or imaginary part of complex-valued component s.

することが重要である [13]．最もよく用いられる分布は，Fig. 2 (a) の右側に示す球状ラプラス分布である．この図では，二つの周波数成分の同時分布を示しており，原点を中心に球対称となっている．この性質から，二つの変数間に高次の相関が保証される．

IVA が仮定している音源モデルは，球状多変量分布そのものと解釈できる．この音源モデルを Fig. 2 (a) の左側に示す．各音源は周波数方向に一定の分散値を持っており，それらが時間的に変化するようなパワースペクトログラムを仮定している．従って，複数の周波数で同時に生起する成分を同一音源としてまとめる傾向がある．さらに，音源モデルのパワースペクトログラムを行列とみたとき，1 本の基底ベクトルで表現できる．これは 1 つの音源に対して 1 本のスペクトル基底を与えた NMF と解釈することもできる．

一方，IVA は空間の性質に関して具体的なモデルを与えていない．音源やマイクの位置条件に関係なく，音源モデルの統計的独立性及び多チャネルの観測信号のみから分離行列の推定を行う．

2.4　ILRMA の仮定する音源及び空間モデル

ILRMA のコスト関数は以下で定義される [15]．

$$Q_{\mathrm{ILRMA}} = \sum_{i,j}\left[\sum_n \frac{|y_{ij,n}|^2}{\sum_l t_{il,n} v_{lj,n}} - 2\log|\det \boldsymbol{W}_i| + \sum_n \log \sum_l t_{il,n} v_{lj,n}\right] \tag{7}$$

ここで，$t_{il,n}, v_{lj,n}$ は n 番目の音源モデルに対応する非負値スペクトル基底とアクティベーションであり，$l=1,\cdots,L$ は基底のインデックスを示す．すなわち，$\sum_l t_{il,n} v_{lj,n}$ は n 番目の音源のモデルパワースペクトログラムとなる．また，観測チャネル数と音源数の関係は $M=N$ としている．このとき，ILRMA のコスト関数は IVA のコスト関数 (式 (7) の第一項及び第二項) と単一チャネル NMF のコスト関数 (式 (7) の第一項及び第三項) を重ね合わせた形をしている．これらの事実から，IVA は ILRMA においてスペクトル基底数が 1 の特殊ケースに相当しており，その意味で ILRMA は IVA の自然な拡張となっていると解釈できる．

ILRMA の仮定する音源モデルを Fig. 2 (b) に示す．IVA と比較して，1 つの音源に対して L 本のスペクトル基底を用いることができる為，より複雑なパワースペクトログラムを表現可能となっている．また，各時間周波数スロットで独立な複素ガウス分布を音源モデルとして仮定しており，コスト関数 (7) は板倉斎藤擬距離の行列版である log-determinant divergence となっている．従って，時間と周波数いずれの方向にも分散が変動する分布を定義でき，より複雑な時間周波数構造を，限られた基底数で低ランク分解される音源モデルとして表現できる．

一方，空間モデルに関して，ILRMA は，IVA と同様に具体的なモデルを与えていない．音源やマイクの位置に依存せず，観測信号と前述のモデルスペクトログラムの独立性から分離行列を推定する．

3　アルゴリズム概観：どう最適化するのか？

3.1　生成モデル or 分離モデル？

前章を眺めてお気づきの読者もいるとは思うが，先に示した BSS アルゴリズムは全て「音源のパラメタライズ」と「空間分離フィルタ」を推定する問題となっている．つまり,「分離モデル型」であると言える．しかし，一般に統計的モデル推定の観点から眺めると，生成モデル（音源モデル＋空間混合モデル $\boldsymbol{A}_i$）の同時推定問題を解く手法の方がポピュラーかもしれない．実際，MNMF をはじめとする他の音源分離手法は,「生成モデル型」であることが多い．では，BSS における「分離モデル型」の利点は何であろうか？実は，最適化アルゴリズムの発展と深く関係がある．以下の節にて具体的な例を挙げて説明を試みる．

3.2　FDICA における最適化

FDICA におけるコスト関数 Q_{FDICA} の最小化問題を直接解くことは出来ないため，反復法を用いて分離行列及び音源モデルパラメータを求める．様々なものが提案されたが，音声音響信号処理にて最も普及していたアルゴリズムは，分布 $p(\cdot)$ をラプラス分布等で固定化した最急降下法に基づくものであり，Amari らによって提案された自

然勾配 (natural gradient) [17] が有名である．

$$-\frac{\partial Q_{\mathrm{FDICA}}}{\partial \boldsymbol{W}_i}\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{W}_i=\left(\boldsymbol{W}^{-\mathrm{T}}-\frac{1}{J}\sum_j \boldsymbol{\Phi}(\boldsymbol{y}_{ij})\boldsymbol{x}_{ij}^{\mathrm{T}}\right)\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{W}_i$$
$$=\left(\boldsymbol{I}-\frac{1}{J}\sum_j \boldsymbol{\Phi}(\boldsymbol{y}_{ij})\boldsymbol{y}_{ij}^{\mathrm{T}}\right)\boldsymbol{W}_i \quad (8)$$

ここで $\boldsymbol{\Phi}(\cdot)$ は適当なベクトル関数であり，シグモイド関数等が用いられる．

この自然勾配法の意義について，音響信号処理の視点で考察する．式 (8) 左辺を見ると，単純な勾配（$\partial Q_{\mathrm{FDICA}}/\partial \boldsymbol{W}_i$）にリーマン計量 $\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{W}_i$ が乗じられていることが分かる．これにより，単純な勾配に現れる「分離行列の逆行列 $\boldsymbol{W}^{-\mathrm{T}}$（つまり生成系 $\boldsymbol{A}_i^{\mathrm{T}}$)」を打ち消し，一切の逆行列演算を行うことなく分離音源を求めることが出来る．この恩恵は，空間混合が畳み込みで表される音響信号処理においては非常に大きなものであった．また，分離行列そのものを（逆行列演算せずに）反復更新するので，それをビームフォーミング等の空間フィルタと解釈すれば，過去の音響信号処理研究で得られた事前情報を盛り込むことが容易となり，様々な融合手法が産み出されるに至った [5, 7]．

3.3 IVA 及び ILRMA における最適化

FDICA のパーミュテーション問題を解決するために提案された IVA においても，当初は自然勾配の形で分離行列を反復更新するアルゴリズムが使われていた．前述の通り，演算量の少なさは大きな魅力であったが，基本は最急降下法であるため，その収束性（コスト関数の単調減少性）は保証されないという問題があった．この収束性の保証問題に関し，2011 年 Ono らは補助関数法と Iterative Projection (IP) に基づく IVA 更新式を提案した [14]．これは分離行列自体を補助関数と IP で更新するものであり，分離モデル型ならではの特徴を活かしつつ，かつコスト関数の単調減少性を保証する画期的な発明であった．

この補助関数法及び IP に基づく IVA の発明は，続く ILRMA の発明にも大きく影響している．2.4 節にて述べた通り，ILRMA は IVA の自然な低ランク行列拡張であるが，そのコスト関数式 (7) は板倉斎藤擬距離基準 NMF と IVA との結合となっている．つまり，このコスト関数全体を補助関数法で最適化することができ，全パラメータ（分離行列 $\boldsymbol{W}_i$ 及び音源パラメータ $t_{il,n}, v_{lj,n}$）に関して収束性の保証が与えられた反復更新式を得ることが出来る．その場合，音源低ランクモデル $t_{il,n}, v_{lj,n}$ は乗算更新の形となり，非負値性も保証される．

以下，具体的な最適化アルゴリズム [15] について説明する．まず，音源パラメータ $t_{il,n}, v_{lj,n}$ が固定された元で，分離行列 $\boldsymbol{W}_i$ の更新を行う．ここでは $r_{ij,n}=\sum_k t_{ik,n}v_{kj,n}$ とおき，以下のステップに従って IP を実行する．

$$\boldsymbol{U}_{i,n}\leftarrow\frac{1}{J}\sum_j\frac{1}{r_{ij,n}}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}} \quad (9)$$

$$\boldsymbol{w}_{i,n}\leftarrow(\boldsymbol{W}_i\boldsymbol{U}_{i,n})^{-1}\boldsymbol{e}_n \quad (10)$$

$$\boldsymbol{w}_{i,n}\leftarrow\frac{\boldsymbol{w}_{i,n}}{\sqrt{\boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{U}_{i,n}\boldsymbol{w}_{i,n}}} \quad (11)$$

$$y_{ij,n}\leftarrow\boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{x}_{ij} \quad (12)$$

ここで，$\boldsymbol{e}_n$ は n 番目の要素が 1 である単位ベクトルである．このアルゴリズムを繰り返すことにより，コスト関数 (7) が単調に減少するよう分離行列 $\boldsymbol{W}_n$ が更新される．

次に，音源の低ランクモデルパラメータ $t_{il,n}$ 及び $v_{lj,n}$ の更新を行う．これは，その形式より，通常の板倉斎藤擬距離基準 NMF と同様な最適化アルゴリズムが適用できる．具体的には，補助関数法を用いて以下のように更新される．

$$t_{ik,n}\leftarrow t_{ik,n}\left[\frac{\sum_j\frac{|y_{ij,n}|^2}{\left(\sum_k t_{ik,n}v_{kj,n}\right)^2}v_{kj,n}}{\sum_j\frac{1}{\sum_k t_{ik,n}v_{kj,n}}v_{kj,n}}\right]^{\frac{1}{2}} \quad (13)$$

$$v_{kj,n}\leftarrow v_{kj,n}\left[\frac{\sum_i\frac{|y_{ij,n}|^2}{\left(\sum_k t_{ik,n}v_{kj,n}\right)^2}t_{ik,n}}{\sum_i\frac{1}{\sum_k t_{ik,n}v_{kj,n}}t_{ik,n}}\right]^{\frac{1}{2}} \quad (14)$$

$$r_{ij,n}\leftarrow\sum_k t_{ik,n}v_{kj,n} \quad (15)$$

上記は大規模な逆行列演算を多数回行う必要もなく，複雑な代数方程式を解く必要も無い．以上より，非常に少ない演算量で ILRMA の全パラメータを更新可能であることが分かる．

3.4 MNMF と ILRMA のミッシングリンク

本節では，BSS の拡張として提案された ILRMA と低ランクモデリングとして先に発展した MNMF とを比較し，各最適化アルゴリズムの関連性・相違について述べ，分離モデル型の優位性について概説する．まず，MNMF [11] においては，多チャネル観測信号の相関行列 $\mathsf{X}_{ij}=\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}}$ を定義し，それを個々の音源に関する空間相関 $\boldsymbol{R}_{i,n}^{(s)}$ と NMF 音源パラメータによる近似

$$\mathsf{X}_{ij}\approx\hat{\mathsf{X}}_{ij}=\sum_k\left(\sum_n\boldsymbol{R}_{i,n}^{(s)}d_{nk}\right)t_{ik}v_{kj} \quad (16)$$

により時空間のモデリングを行う．ここで d_{nk} は各音源へ基底を分配する変数である．本モデルに基づく MNMF のコスト関数は以下で与えられる．

$$Q_{\mathrm{MNMF}}=\sum_{i,j}\left[\mathrm{tr}\left(\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\right)+\log\det\hat{\mathsf{X}}_{ij}\right] \quad (17)$$

前述の通り，MNMF は信号の生成系を定義して観測信号を記述しており,「生成モデル型」と言うことが出来る．本

モデリングは柔軟であり，空間がランク 1 でない場合も記述することが出来るが，一方でその最適化は容易ではなく莫大な演算量と解の不安定性が大きな問題とされていた．

ここで大変興味深いことは，この MNMF のコスト関数式 (17) において空間ランク 1 仮説，つまり $\boldsymbol{R}_{i,n}^{(\boldsymbol{s})} = \boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathrm{H}}$ とおき，更に $\mathsf{D}_{ij} = \mathrm{diag}[\sum_k d_{1k}t_{ik}v_{kj}, \cdots, \sum_k d_{Nk}t_{ik}v_{kj}]$ とおけば，$\boldsymbol{A}_i = \boldsymbol{W}_i^{-1}$ より

$$
\begin{aligned}
Q = \sum_{i,j}\Big[&\mathrm{tr}\Big(\boldsymbol{W}_i^{-1}\boldsymbol{y}_{ij}\boldsymbol{y}_{ij}^{\mathrm{H}}\left(\boldsymbol{W}_i^{\mathrm{H}}\right)^{-1}\boldsymbol{W}_i^{\mathrm{H}}\mathsf{D}_{ij}^{-1}\boldsymbol{W}_i\Big) \\
&+\log\left(\det \boldsymbol{A}_i\right)\left(\det \mathsf{D}_{ij}\right)\left(\det \boldsymbol{A}_i^{\mathrm{H}}\right)\Big] \\
= \sum_{i,j}\Big[&\mathrm{tr}\Big(\boldsymbol{W}_i\boldsymbol{W}_i^{-1}\boldsymbol{y}_{ij}\boldsymbol{y}_{ij}^{\mathrm{H}}\left(\boldsymbol{W}_i^{\mathrm{H}}\right)^{-1}\boldsymbol{W}_i^{\mathrm{H}}\mathsf{D}_{ij}^{-1}\Big) \\
&+2\log|\det \boldsymbol{A}_i| + \log\det \mathsf{D}_{ij}\Big] \\
= \sum_{i,j}\Bigg[&\sum_n \frac{|y_{ij,n}|^2}{\sum_k d_{nk}t_{ik}v_{kj}} - 2\log|\det \boldsymbol{W}_i| \\
&+\sum_m \log\sum_k d_{nk}t_{ik}v_{kj}\Bigg] \qquad (18)
\end{aligned}
$$

となり，これは ILRMA のコスト関数式 (7) に基底分配変数 d_{nk} を含めたものと一致する．つまり，ILRMA のコスト関数は，MNMF に空間ランク 1 仮説を置き，更に空間モデルのパラメータを生成モデル型（$\boldsymbol{A}_i$）から分離モデル型（$\boldsymbol{W}_i$）へ変更したものだと言うことが出来る．たかが逆行列の関係と思えるかもしれないが，最適化問題においてどちらのドメインで計算するかは非常に大きな差異を生み出す．この空間パラメータ最適化ドメインの変更こそが ILRMA の本質的な新規性であり，それによって全変数の補助関数法（特に分離行列は IP による更新）による高速最適化及び収束性の保証が達成されたと言える．

3.5 実験的比較例

前節で考察した各手法の音源及び空間モデルの違いを例示する為に，人工的に作成した音源を用いた実験結果を Fig. 3 に示す．ここでは，音源スペクトログラムのランク R を変え，FDICA（パーミュテーション解決は DOA 利用），IVA，及び ILRMA による分離を行った（実験条件の詳細は [15] 参照）．本図より，FDICA は音源のランクに影響されにくいがパーミュテーションエラーにより音源分離性能が低く，IVA は逆に音源のランクに強く影響されることが分かる．一方で，ILRMA は適切な基底数を与えてやれば，高い分離性能を維持できることも示されている．

次に，実際の音響データ（SiSEC [19]）を用いた分離実験結果例を Table 1 に示す（実験条件の詳細は [15] 参照）．従来の手法と比べ，ILRMA の分離精度が高く，演算時間の面でも効率的であることが示されている．

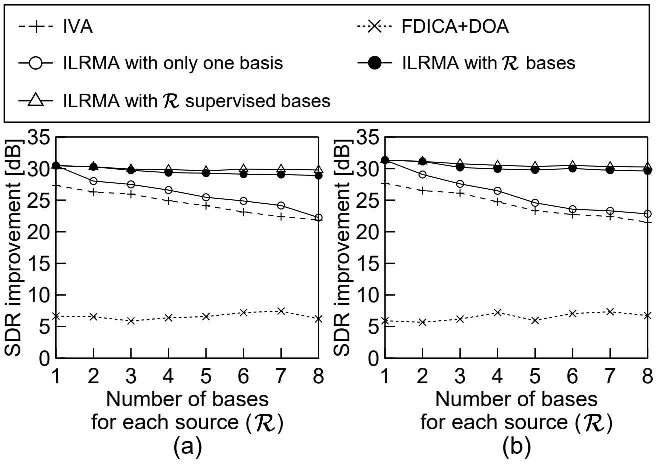

図 3: SDR [18] results of (a) source 1 and (b) source 2 for various numbers of bases.

表 1: Averaged SDR improvement in dB under SiSEC conditions and computational time normalized by IVA's one

Algorithm	SDR improv.	Comp. time
Soft masking [20]	-0.1	-
IVA [14]	2.6	1.0
Ozerov's MNMF [9]	1.2	-
Sawada's MNMF [11]	5.0	49.1
ILRMA	8.7	1.3

4 スパースな生成モデルに基づく BSS

4.1 複素 Student's t 分布に基づく ILRMA

前述の通り，従来の ILRMA はその生成確率モデルとして時変複素ガウス分布を仮定していた．一方で，音声や音楽信号等に関し，ガウス分布よりもさらに尖度の高い「スパース」な分布を仮定することも出来る．ここでは，その一例として，複素 Student's t 分布に基づくものを紹介する [21]．以降ではオリジナルな ILRMA[15] と区別するため，本アルゴリズムを t-ILRMA と呼ぶ．ここでは以下の生成モデルを考える．

$$
\prod_{i,j} p(y_{ij,n}) = \prod_{i,j}\frac{1}{\pi\sigma_{ij,n}^2}\left(1+\frac{2}{\nu}\frac{|y_{ijn}|^2}{\sigma_{ij,n}^2}\right)^{-\frac{2+\nu}{2}} \qquad (19)
$$

$$
\sigma_{ij,n}^p = \sum_l t_{il,n}v_{lj,n} \qquad (20)
$$

ここで，分布 $p(y_{ij,n})$ は球状（原点対称）複素 Student's t 分布であり，$\sigma_{ij,n}$ は時間周波数において振幅スペクトル $|y_{ij,n}|$ に対応する時変な非負値スケールである．また，ν は分布の形状を制御する自由度パラメータ，p はスペクトログラムの指数乗ドメインを定めるドメインパラメータであり，$1 \leq p \leq 2$ を満たす．ここで $\nu \to \infty$ かつ $p=2$ とするならば，式 (19) は時変複素ガウス分布に基づく生成モデルに一致し，$\nu=1$ かつ $p=1$ とするならば，式 (19) は時

変複素コーシー分布に基づく生成モデルに一致する．信号間の独立性の仮定とともに，式 (19) の負対数尤度は以下で与えられる．

$$\mathcal{L}_t = \text{const.} - 2J\sum_i \log|\det \boldsymbol{W}_i| + \sum_{i,j,n}\left[\left(1+\frac{\nu}{2}\right)\log\left(1+\frac{2}{\nu}\frac{|y_{ij,n}|^2}{\sigma_{ij,n}^2}\right) + 2\log\sigma_{ij,n}\right] \quad (21)$$

ここで，$\nu\to\infty$ かつ $p=2$ ならば，式 (21) は ILRMA のコスト関数式 (7) に一致する．

4.2　t-ILRMA における分離行列の更新式

ここでは分離行列 $\boldsymbol{W}_i$ の更新について考える．一般に，前述の複素ガウス分布由来の ILRMA において用いられていた IP は，logdet 項と $|y_{ij,n}|^2=|\boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{x}_{ij}|^2$ の項の和に対してのみ適用できたが，t-ILRMA におけるコスト関数 (21) においては $|y_{ij,n}|^2$ の項が対数関数内にあるため，IP を直接適用することが不可能である．そこで，補助関数法を適用するため，以下の接線不等式を考える．

$$\log\left(\sum_q z_q\right) \le \frac{1}{\lambda}\left(\sum_q z_q - \lambda\right) + \log\lambda \quad (22)$$

ここで，z_q は元の変数，$\lambda>0$ は補助変数である．式 (22) の等号は $\lambda=\sum_q z_q$ の時に限り成立する．式 (22) を式 (21) の第三項及び第四項に適用することにより，以下の補助関数 $\mathcal{L}_t^+$ を得ることが出来る．

$$\begin{aligned}\mathcal{L}_t \le\ & \text{const.} - 2J\sum_i \log|\det \boldsymbol{W}_i| \\ & + \sum_{i,j,n}\left[\left(1+\frac{\nu}{2}\right)\frac{1}{\alpha_{ij,n}}\left(1+\frac{2}{\nu}\frac{|y_{ij,n}|^2}{\sigma_{ij,n}^2}-\alpha_{ij,n}\right)\right. \\ & + \left(1+\frac{\nu}{2}\right)\log\alpha_{ij,n} + \frac{2}{p\beta_{ij,n}}\left(\sum_l t_{il,n}v_{lj,n} - \beta_{ij,n}\right) \\ & \left. + \frac{2}{p}\log\beta_{ij,n}\right] \\ \equiv\ & \mathcal{L}_t^+ \end{aligned} \quad (23)$$

ここでは $\sigma_{ij,n}=(\sum_l t_{il,n}v_{lj,n})^{1/p}$ を代入した．$\alpha_{ij,n},\beta_{ij,n}>0$ は補助変数であり，$\mathcal{L}_t$ と $\mathcal{L}_t^+$ は以下の条件の時に限り等しくなる．

$$\alpha_{ij,n} = 1+\frac{2}{\nu}\frac{|y_{ij,n}|^2}{\sigma_{ij,n}^2} \quad (24)$$

$$\beta_{ij,n} = \sum_l t_{il,n}v_{lj,n} \quad (25)$$

式 (23) においては $|y_{ij,n}|^2=|\boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{x}_{ij}|^2$ が対数関数の外にあるので，IP を直接適用することが出来る．さらに補助関数 (23) を整理し，以下の式を得る．

$$\begin{aligned}\mathcal{L}_t^+ =\ & \text{const.} - 2J\sum_i \log|\det \boldsymbol{W}_i| + J\sum_{i,n}\boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{U}_{i,n}\boldsymbol{w}_{i,n} \\ & + \sum_{i,j,n}\left[\left(1+\frac{\nu}{2}\right)\left(\alpha_{ij,n}^{-1}-1+\log\alpha_{ij,n}\right)\right. \\ & \left. + \frac{2}{p\beta_{ij,n}}\left(\sum_l t_{il,n}v_{lj,n}-\beta_{ij,n}\right) + \frac{2}{p}\log\beta_{ij,n}\right] \end{aligned} \quad (26)$$

$$\boldsymbol{U}_{i,n} = \frac{1}{J}\left(\frac{2}{\nu}+1\right)\sum_j \frac{1}{\alpha_{ij,n}\sigma_{ij,n}^2}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}} \quad (27)$$

式 (26) の停留点を $\boldsymbol{w}_{i,n}$ に関して求めることは，以下の連立方程式を解くことと等価である．

$$\boldsymbol{w}_{i,k}^{\mathrm{H}}\boldsymbol{U}_{i,n}\boldsymbol{w}_{i,n} = \delta_{kn} \quad (28)$$

ここで，$\delta_{kn}=1\ (k=n)$ および $\delta_{kn}=0\ (k\neq n)$ である．式 (28) に IP を適用することにより，分離行列に関して以下の反復更新式を得ることが出来る．

$$\boldsymbol{w}_{i,n} \leftarrow (\boldsymbol{W}_i\boldsymbol{U}_{i,n})^{-1}\boldsymbol{e}_n \quad (29)$$

$$\boldsymbol{w}_{i,n} \leftarrow \frac{\boldsymbol{w}_{i,n}}{\sqrt{\boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{U}_{i,n}\boldsymbol{w}_{i,n}}} \quad (30)$$

$\boldsymbol{W}_i$ の更新の後，分離信号 $\boldsymbol{y}_{ij}$ は $y_{ij,n}\leftarrow\boldsymbol{w}_{ij,n}^{\mathrm{H}}\boldsymbol{x}_{ij}$ のように更新される．

4.3　t-ILRMA における NMF パラメータの更新式

t-ILRMA の低ランク音源モデル（NMF）に関するパラメータ $t_{il,n}$ 及び $v_{lj,n}$ も，補助関数法を用いて求めることが可能である．式 (23) の NMF パラメータに関する補助関数を得るため，以下の Jensen の不等式を $\sigma_{ij,n}^{-2} = (\sum_l t_{il,n}v_{lj,n})^{-2/p}$ に適用する．

$$\left(\sum_q z_q\right)^{-2/p} = \left(\sum_q \mu_q\frac{z_q}{\mu_q}\right)^{-2/p} \le \sum_q \mu_q\left(\frac{z_q}{\mu_q}\right)^{-2/p} = \sum_q \mu_q^{\frac{2}{p}+1}z_q^{-\frac{2}{p}} \quad (31)$$

ここで $\mu_q>0$ は $\sum_q\mu_q=1$ を満たす補助変数である．なお，$1\le p\le 2$ を想定しているので，式 (31) は変数 z_q に関して凸関数であることに留意する．式 (31) の等号は $\mu_q=z_q/\sum_{q'}z_{q'}$ の時に限り成立する．式 (31) を式 (21) の $\sigma_{ij,n}^{-2}=(\sum_l t_{il,n}v_{lj,n})^{-2/p}$ に適用することにより，以下の補助関数 $\mathcal{L}_t^{++}$ を得る．

$$\begin{aligned}\mathcal{L}_t^+ \le\ & \text{const.} - 2J\sum_i \log|\det \boldsymbol{W}_i| \\ & + \sum_{i,j,n}\left[\left(1+\frac{\nu}{2}\right)\frac{1}{\alpha_{ij,n}}\left(1+\frac{2}{\nu}|y_{ij,n}|^2\sum_l \gamma_{ij,nl}^{\frac{2}{p}+1}t_{il,n}^{-\frac{2}{p}}v_{lj,n}^{-\frac{2}{p}}-\alpha_{ij,n}\right)\right. \\ & + \left(1+\frac{\nu}{2}\right)\log\alpha_{ij,n} + \frac{2}{p\beta_{ij,n}}\left(\sum_l t_{il,n}v_{lj,n}-\beta_{ij,n}\right) \\ & \left. + \frac{2}{p}\log\beta_{ij,n}\right] \\ \equiv\ & \mathcal{L}_t^{++} \end{aligned} \quad (32)$$

ここで，$\gamma_{ij,nl} > 0$ は補助変数であり，$\mathcal{L}_t^{+}$ と $\mathcal{L}_t^{++}$ は以下の条件の時に限り等しくなる．

$$\gamma_{ij,nl} = \frac{t_{il,n} v_{lj,n}}{\sum_{l'} t_{il',n} v_{l'j,n}} \tag{33}$$

式 (32) の最小点を求めるため $\partial\mathcal{L}_t^{++}/\partial t_{il,n} = 0$ を計算し，以下の式を得る．

$$t_{il,n} = \left[\frac{\left(\frac{2}{\nu}+1\right)\sum_j \frac{1}{\alpha_{ij,n}} |y_{ij,n}|^2 \gamma_{ij,nl}^{\frac{2}{p}+1} v_{lj,n}^{-\frac{2}{p}}}{\sum_j \frac{1}{\beta_{ij,n}} v_{lj,n}} \right]^{\frac{p}{p+2}} \tag{34}$$

式 (25) と (33) を式 (34) に代入することにより，基底 $t_{il,n}$ に関する以下の更新式を得る．

$$t_{il,n} \leftarrow t_{il,n} \left[\frac{\sum_j |y_{ij,n}|^2 \left(\frac{\nu}{\nu+2}\sigma_{ij,n}^2 + \frac{2}{\nu+2}|y_{ij,n}|^2\right)^{-1} \sigma_{ij,n}^{-p} v_{lj,n}}{\sum_j \sigma_{ij,n}^{-p} v_{lj,n}} \right]^{\frac{p}{p+2}} \tag{35}$$

式 (35) と同様にして，アクティベーション $v_{lj,n}$ に関する以下の更新式を得る．

$$v_{lj,n} \leftarrow v_{lj,n} \left[\frac{\sum_i |y_{ij,n}|^2 \left(\frac{\nu}{\nu+2}\sigma_{ij,n}^2 + \frac{2}{\nu+2}|y_{ij,n}|^2\right)^{-1} \sigma_{ij,n}^{-p} t_{il,n}}{\sum_i \sigma_{ij,n}^{-p} t_{il,n}} \right]^{\frac{p}{p+2}} \tag{36}$$

パラメータ $t_{il,n}$ と $v_{lj,n}$ を更新した後, 低ランクモデル $\sigma_{ij,n}^p$ が式 (20) に従って更新される．

4.4 *t*-ILRMA におけるスパース性と低ランク性の関係

複素 Student's *t* 分布を生成モデルに持つことによりスパースな信号を仮定することが出来たが，そのことと低ランクモデリングとの関係はどうなっているのであろうか．この疑問に答えるため，前節にて導出された低ランク音源モデル（NMF）パラメータ更新式を従来の板倉斎藤擬距離基準 NMF で解釈してみる．例えば，基底 $t_{il,n}$ に関する更新式 (35) は，以下のように書き換えることが出来る．

$$t_{ik,n} \leftarrow t_{ik,n} \left[\frac{\sum_j \frac{z_{ij,n}^p}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^2} v_{kj,n}}{\sum_j \frac{1}{\sum_{k'} t_{ik',n} v_{k'j,n}} v_{kj,n}} \right]^{\frac{p}{p+2}} \tag{37}$$

$$\begin{aligned} z_{ij,n} &= \left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^{1-\frac{2}{p}} \left[\frac{\nu}{\nu+2} |y_{ij,n}|^{-2} \right. \\ &\qquad \left. + \frac{2}{\nu+2} \left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^{-\frac{2}{p}} \right]^{-1} \\ &= \sigma_{ij,n}^{p-2} \left(\frac{\nu}{\nu+2} |y_{ij,n}|^{-2} + \frac{2}{\nu+2} \sigma_{ij,n}^{-2} \right)^{-1} \end{aligned} \tag{38}$$

式 (37) は,「板倉斎藤擬距離基準 NMF において指数乗に関して一般化された更新式 [22]」の形になるよう整式されている．ここでは，各パラメータの挙動に関して以下のように解釈することが出来る．

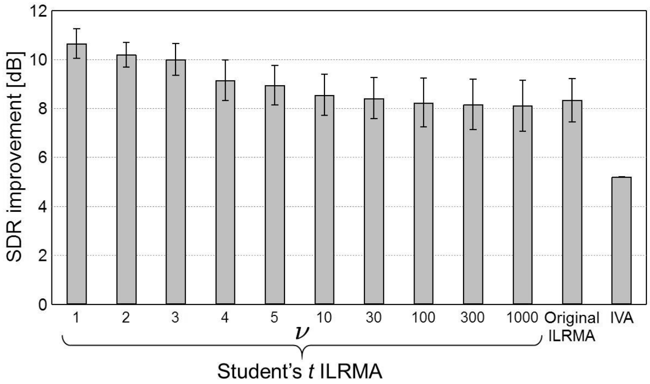

図 4: SDR results of *t*-ILRMA for various parameter ν.

- 式 (37) は，仮想的な観測信号 $z_{ij,n}$ に関する板倉斎藤擬距離基準 NMF だと見なせる．また，その $z_{ij,n}$ は，真の観測信号 $y_{ij,n}$ と低ランクモデル $\sigma_{ij,n}$ の調和平均で表される（式 (38) 参照）．上記の調和平均の比率は ν 対 2 であることより，自由度パラメータ ν を小さくする（よりスパースな信号を仮定する）につれて低ランク性が強調されることになる．
- ドメインパラメータ p は主に NMF における乗算更新量の指数 $p/(p+2)$ を制御する．p を小さくするほどこの指数は小さくなり，NMF の更新速度が遅くなる．一般に ILRMA においては，分離行列の更新速度とのバランスによって局所解への停留確率が変化することより，特に反復初期においてこの指数は小さい方が好ましいとされている [22].
- また，式 (38) には $\sigma_{ij,n}^{p-2}$ が含まれる．これは，$p < 2$ とおいた場合，低ランクモデルでの除算を意味し，低ランク性をディスカウントする効果が生じる．これにより，過度な低ランク性の強調が抑制される．

以上より，自由度パラメータ ν とドメインパラメータ p の設定についてまとめると，比較的小さな p 及び大きな ν で反復を開始し，徐々に ν を下げて低ランク性を強調していく tempering（焼き戻し）が有効であると言える．

4.5 実験的比較例

t-ILRMA におけるスパース音源モデルによる性能改善を示す為に，実際の音響データを用いた分離実験結果例を Fig. 4 に示す（実験条件の詳細は [21] 参照）．ここでは，前節で述べた tempering を行っており，パラメータの反復更新前半では $\nu = \infty$ とし，後半では図の横軸に示す値にしている．本図より，複素ガウス分布由来の手法と比べて ν の小さな *t*-ILRMA の分離精度が高く，スパースな音源モデルの導入が効果的であることが示されている．

5 事前分布・正則化の導入

2.4 節にて解説した ILRMA において，分離行列や低ランク音源モデルパラメータに事前分布を入れることで分離

性能を安定化させることも可能である．これは，各パラメータに関する正則化項を主コスト関数 Q_{ILRMA} へ加えることで実現される．以下にその実装例を示す．

● **低ランク音源モデル：** 従来の板倉斎藤擬距離基準 NMF と同様にして，スパース性を誘導する正則化項を付与することで音源モデルパラメータの事前分布を与える．例えば，文献 [23] では，アクティベーション $v_{lj,n}$ の事前分布としてラプラス分布を仮定し，L1 ノルムを正則化として加えるものが提案されている．また，Mitsui らによって，低ランク表現された音源スペクトログラムの事前分布として周波数毎に独立なカイ分布を仮定するものが提案されており，ILRMA における性能向上が報告されている [24]．

● **分離行列：** FDICA 等の最急降下法においてはその勾配に適切な正則化項の勾配を加算するだけで実装できるため，分離行列の正則化に関して多くの手法が提案されてきた [5]．一方で，IVA や ILRMA 等の補助関数法に基づく手法では，正則化項の加算によって補助関数の最小点が変化するため，解析解の導出が困難になるという問題があった．近年，三井らによって，分離行列の事前分布としてガウス分布を仮定する場合（分離行列に関する L2 正則化）に限り，それを分離行列の行ベクトル毎のブロック座標降下法で解くアルゴリズムが提案されており，ILRMA における有効性が確認されている [25]．

6 おわりに

本稿では，ランク 1 空間近似を用いた 3 つの BSS について，それらの音源及び空間モデルに関して考察した．特に，空間モデルパラメータの最適化に関し，MNMF 等との比較を通じて，分離モデル型に利があることを解説した．更に，実験的な比較例を通じ，ILRMA が分離精度及び演算量の両面において優れていることを示した．また，ILRMA の拡張として，スパース音源分布モデルを仮定した手法に関しても解説を行った．

謝辞 本稿を執筆するにあたり，貴重なご助言及び資料をご提供頂いた東京大・北村大地氏に深く感謝いたします．本研究の一部は，セコム科学技術振興財団及び総合科学技術・イノベーション会議による革新的研究開発推進プログラム（ImPACT）の支援を受けた．

References

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

[6] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, “The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech,” *IEEE Trans. SAP*, vol. 11, no. 2, pp. 109–116, 2003.

[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] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” *Proc. Advances in Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.

[9] A. Ozerov and C. Fevotte, “Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation,” *IEEE Trans. ASLP*, vol. 18, no. 3, pp. 550–563, 2010.

[10] 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. Int. Conf. Latent Variable Anal. Signal Separation*, pp. 245–253, 2010.

[11] 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.

[12] A. Hiroe, “Solution of permutation problem in frequency domain ICA using multivariate probability density functions,” *Proc. ICA*, pp. 601–608, 2006.

[13] 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.

[14] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” *Proc. WASPAA*, pp. 189–192, 2011.

[15] 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.

[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. ASLP*, vol. 18, no. 7, pp. 1830–1840, 2010.

[17] S. Amari, S. Douglas, A. Cichocki, H. Yang, “Multichannel blind deconvolution and equalization using the natural gradient,” *Proc. IEEE International Workshop on Wireless Commun.*, pp. 101–104, 1997.

[18] E. Vincent, R. Gribonval and C. Fevotte, “Performance measurement in blind audio source separation,” *IEEE Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.

[19] 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.

[20] S. Araki, H. Sawada, R. Mukai, and S. Makino, “Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors,” *Signal Processing*, vol. 87, no. 8, pp. 1833–1847, 2007.

[21] S. Mogami, D. Kitamura, Y. Mitsui, N. Takamune, H. Saruwatari, N. Ono, “Independent low-rank matrix analysis based on complex Student’s t-distribution for blind audio source separation,” *Proc. MLSP*, 2017.

[22] Y. Mitsui, D. Kitamura, N. Takamune, H. Saruwatari, Y. Takahashi, K. Kondo, “Independent low-rank matrix analysis based on parametric majorization-equalization algorithm,” *Proc. CAMSAP*, 2017.

[23] W. Liu, N. Zheng, X. Lu, “Non-negative matrix factorization for visual coding,” *Proc. ICASSP*, vol. 3, pp. 293–296, 2003.

[24] Y. Mitsui, D. Kitamura, S. Takamichi, N. Ono, H. Saruwatari, “Blind source separation based on independent low-rank matrix analysis with sparse regularization for timeseries activity,” *Proc. ICASSP*, pp. 21–25, 2017.

[25] 三井祥幹, 高宗典玄, 北村大地, 猿渡洋, 高橋祐, 近藤多伸, “空間事前情報を用いた独立低ランク行列分析,” 第 32 回 SIP シンポジウム, no. B8-2, pp. 360–365, 2017.