Vol.2015-MUS-107 No.31
2015/5/23

多チャネル非負値行列因子分解における ランク1空間モデルの音源分離性能評価

北村 大地[1]　小野 順貴[2,1]　澤田 宏[3]　亀岡 弘和[4,3]　猿渡 洋[4]

概要：高精度なブラインド音源分離手法の一つとして多チャネル非負値行列因子分解 (MNMF) がある．MNMF は，音源の混合系を音源毎の空間相関行列として推定し分離を行うが，そのモデルの複雑さから最適化が困難となり，強い初期値依存性や高い 計算コストが問題となる．これらの問題を解決するために，空間相関行列をランク 1 に近似して推定するランク 1 空間モデルを用いた MNMF が提案さ れているが，本手法とフルランクの空間相関行列を推定する従来手法の分離性能の比較は行われていない．本稿では，音楽信号のブラインド音源分離を対象として，従来の MNMF とランク 1 空間モデルを用いた MNMF の分離性能比較を行い，ランク 1 空間モデルの妥当性に関して議論する．

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．過決定条件 (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]–[6]．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) [7] を応用した手法が注目を集めている [8], [9]．BSS は一般的に，話者分離や雑音抑圧が目的であるが，音楽を対象とした音源分離の研究も増加している [10]．

単一チャネルにおける NMF を用いた BSS では，分解されたスペクトル基底及びアクティベーションを音源毎にクラスタリングする必要があり，これは容易ではない．このようなスペクトル基底のクラスタリングを解決する最も単純な手法として，分離目的音源のサンプル信号を事前に学習する教師あり NMF [11], [12] が提案されている．しかし，BSS の枠組みでは，分離対象となる音源情報は未知であるため，このような教師あり手法の適用は不可能である．そこで，従来の NMF を多チャネル信号用に拡張した多チャネル NMF (multichannel NMF: MNMF) [13]–[15] が提案された．MNMF では，多チャネルで観測された信号を対象とし，チャネル間の空間的な情報 (音量比や位相差) を音源分離に活用することが可能である．文献 [13], [14] では，混合系と音源を別々にモデル化したうえで，EM アルゴリズムを用いて最適化を行う手法が提案されている．さらに文献 [15] では，混合系と音源を統合した形で定式化している．この MNMF モデルでは，単一チャネル NMF と同様に，乗法型反復更新式による最適化手法が提案されている．しかし，これらの手法は，モデルの自由度が高い反面，最適解を見つけることが非常に困難であり，反復更新回数の増加や極端な初期値依存性をまねいている．その結果，分離精度が不安定となる問題がある．

上記の MNMF の最適化問題を解決する手法として，混合系に対応する空間相関行列をランク 1 行列でモデル化するランク 1 空間制約付き MNMF (Rank-1 MNMF) [16], [17] が提案された．この手法では，過決定条件 BSS に問題を限定したうえで，ランク 1 空間相関行列を複素瞬時混合行列で再表現し，その逆数である分離行列を求める問題に変数変換することで，空間モデルの最適化を容易にしている．さらに，変換されたコスト関数は，独立ベクトル分析 (independent vector analysis: IVA) [18], [19] と単一チャネル NMF のコスト関数を重ね合わせた形となっており，両手法の更新式を交互に反復することで全変数を高速かつ安定に最適化することができる．

本稿では，音楽信号の BSS を対象として，従来の MNMF と Rank-1 MNMF の分離性能の実験的な比較を行う．さらに，Rank-1 MNMF で推定された分離行列から各音源の空

1 総合研究大学院大学, SOKENDAI (The Graduate University for Advanced Studies)
2 国立情報学研究所, National Institute of Informatics
3 日本電信電話株式会社, Nippon Telegraph and Telephone Corporation
4 東京大学, The University of Tokyo

間相関行列を再構成し，従来の MNMF の初期値として与えた場合の結果についても比較し，ランク 1 空間モデルの妥当性に関して議論する．

2. 定式化と従来手法

2.1 定式化

音源数と観測チャネル数をそれぞれ N 及び M とし，各時間周波数の多チャネルの音源信号，観測信号，分離信号をそれぞれ

$$\boldsymbol{s}_{ij} = (s_{ij,1} \cdots s_{ij,N})^{\mathrm{t}} \quad (1)$$

$$\boldsymbol{x}_{ij} = (x_{ij,1} \cdots x_{ij,M})^{\mathrm{t}} \quad (2)$$

$$\boldsymbol{y}_{ij} = (y_{ij,1} \cdots y_{ij,N})^{\mathrm{t}} \quad (3)$$

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

混合系が時不変かつ短時間フーリエ変換の窓長が音源とマイク間のインパルス応答よりも十分長い場合には，観測信号は混合行列 $A_i=(\boldsymbol{a}_{i,1} \cdots \boldsymbol{a}_{i,N})$ ($\boldsymbol{a}_{i,n}$ は各音源のステアリングベクトル) を用いて次式で表現できる．

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

特に，$M=N$ の過決定条件においては，混合行列の逆行列である分離行列 $\boldsymbol{W}_i=(\boldsymbol{w}_{i,1} \cdots \boldsymbol{w}_{i,M})^{\mathrm{h}}$ が定義され，分離信号は次式で表現できる．

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

但し，$^{\mathrm{h}}$ はエルミート転置を表す．

2.2 MNMF

MNMF では，基底とアクティベーションによる音源モデルに加えて，チャネル間情報として得られる空間 (混合系) のモデル化が可能である．本稿では，文献 [15] で定式化された MNMF について取り扱う．本手法は，チャネル数 $M \geq 2$ の劣決定条件 ($M < N$) にも適用することができる．入力となる多チャンネル観測信号は次式のように表現される．

$$\mathsf{X}_{ij} = \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathrm{h}} \quad (6)$$

$M \times M$ のエルミート半正定値行列となる X_{ij} は，その対角要素が各マイクロホンで観測した i, j 成分のパワー (実数) を示し，非対角要素がマイクロホン間の相関 (位相差) を示す複素数となる．この X_{ij} を，すべての i と j に対して近似する分解モデル $\hat{\mathsf{X}}_{ij}$ は以下で定義される．

$$\mathsf{X}_{ij} \approx \hat{\mathsf{X}}_{ij} = \textstyle\sum_k \left(\sum_n \mathsf{H}_{i,n} z_{nk}\right) t_{ik} v_{kj} \quad (7)$$

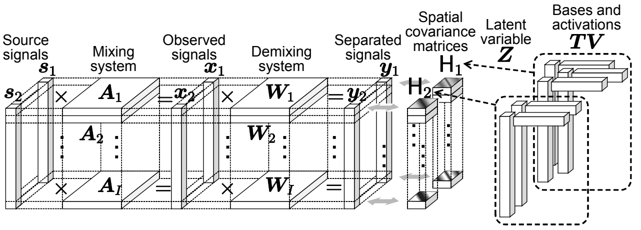

Fig. 1 Conceptual model of MNMF ($N=M=2$).

ここで，$k=1,\ldots,K$ は NMF における基底 (スペクトルパターン) のインデックスを示し，$\mathsf{H}_{i,n}$ は周波数 i における音源 n の空間相関行列を表す $M \times M$ のエルミート半正定値行列である．また，$z_{nk} \in \mathbb{R}_{[0,1]}$ は k 番目の基底を n 番目の音源に対応付ける潜在変数に相当し，$\sum_n z_{nk}=1$ であり，$z_{nk}=1$ のとき，k 番目の基底は n 番目の音源のみに寄与する．さらに，$t_{ik} \in \mathbb{R}_{\geq 0}$ 及び $v_{kj} \in \mathbb{R}_{\geq 0}$ はそれぞれ単一チャネル NMF の基底行列 $\boldsymbol{T}$ ($\in \mathbb{R}_{\geq 0}^{I \times K}$) 及びアクティベーション行列 $\boldsymbol{V}$ ($\in \mathbb{R}_{\geq 0}^{K \times J}$) の要素と等価である．MNMF のモデルの概念を Fig. 1 に示す．BSS においては Fig. 1 に示す混合系や分離系は未知である．MNMF では，観測信号中の全ての音源を K 本の基底 $\boldsymbol{T}$ 及びアクティベーション $\boldsymbol{V}$ でモデル化し，音源毎の空間モデルを空間相関行列 H でモデル化する．さらに，得られた K 本の基底を，潜在変数 $\boldsymbol{Z}$ ($\in \mathbb{R}_{\geq 0}^{N \times K}$) で N 個の空間相関行列にクラスタリングすることで，音源毎に分離された信号 $\boldsymbol{y}$ を得る．

X_{ij} と $\hat{\mathsf{X}}_{ij}$ 間の板倉斎藤擬距離は，定数項を省略すると

$$Q_{\mathrm{MNMF}} = \textstyle\sum_{i,j} \left[\mathrm{tr}(\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}) + \log\det\hat{\mathsf{X}}_{ij}\right] \quad (8)$$

で表される．MNMF においても単一チャネル NMF と同様に補助関数法に基づく最適化が適用されており，乗法型の反復更新式が導出されている [15]．しかし，更新の過程でサイズ $2M$ の行列の固有値分解が必要であり，高い計算コストが要求される．さらに，H はフルランクで推定されるため，空間モデルの自由度が高く，最適化変数の数も非常に多い．結果，局所解が増え，最適化が極めて困難となり，分離精度が初期値に強く依存してしまう問題がある．

3. Rank-1 MNMF

3.1 ランク 1 空間モデル

Figure 1 に示す混合系が，式 (4) のように混合行列 $\boldsymbol{A}_i=(\boldsymbol{a}_{i,1} \cdots \boldsymbol{a}_{i,N})$ で表現できる場合を考える．このとき，各音源の伝達系はステアリングベクトル $\boldsymbol{a}_{i,n}$ で与えられ，その外積となるランク 1 の半正定値エルミート行列 $\boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathrm{h}}$ は，MNMF における空間相関行列 $\mathsf{H}_{i,n}$ に相当する．

$$\mathsf{H}_{i,n} = \boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathrm{h}} \quad (9)$$

このランク 1 空間モデルは，各音源が空気中をコヒーレントに伝播し，混合系が式 (4) に示す時間周波数領域での複素瞬時混合で表現できるという仮定に相当する．

3.2 コスト関数

ランク 1 空間モデルを従来の MNMF に導入するために，式 (9) を分解モデル式 (7) に代入すると，次式を得る．

$$
\begin{aligned}
\hat{\mathsf{X}}_{ij} &= \textstyle\sum_k \left(\sum_n \boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathsf{h}} z_{nk}\right) t_{ik} v_{kj} \\
&= \textstyle\sum_n \boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathsf{h}} \sum_k z_{nk} t_{ik} v_{kj} \\
&= \boldsymbol{A}_i \mathsf{D}_{ij} \boldsymbol{A}_i^{\mathsf{h}}
\end{aligned} \tag{10}
$$

但し，D_{ij} は次式で定義される．

$$\mathsf{D}_{ij} = \mathrm{diag}\left(d_{ij,1}, \ldots, d_{ij,N}\right) \tag{11}$$

$$d_{ij,n} = \textstyle\sum_k z_{nk} t_{ik} v_{kj} \tag{12}$$

式 (10) をコスト関数 (8) に代入すると，次式を得る．

$$Q = \textstyle\sum_{i,j} \left[\mathrm{tr}\left(\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{h}} \left(\boldsymbol{A}_i^{\mathsf{h}}\right)^{-1} \mathsf{D}_{ij}^{-1} \boldsymbol{A}_i^{-1} \right) + \log\det \boldsymbol{A}_i \mathsf{D}_{ij} \boldsymbol{A}_i^{\mathsf{h}} \right] \tag{13}$$

ここで，過決定条件下 (簡便のため $N=M$ とする) では分離行列 $\boldsymbol{W}_i$ が存在するため，$\boldsymbol{W}_i = \boldsymbol{A}_i^{-1}$ 及び $\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij}$ を用いて混合行列から分離行列へ，観測信号から分離信号へ変数変換を行うと，最終的に下記のコスト関数が得られる．

$$
\begin{aligned}
Q &= \textstyle\sum_{i,j} \Big[\mathrm{tr}\left(\boldsymbol{W}_i^{-1} \boldsymbol{y}_{ij}\boldsymbol{y}_{ij}^{\mathsf{h}} \left(\boldsymbol{W}_i^{\mathsf{h}}\right)^{-1} \boldsymbol{W}_i^{\mathsf{h}} \mathsf{D}_{ij}^{-1} \boldsymbol{W}_i \right) \\
&\qquad\qquad + \log\left(\det \boldsymbol{A}_i\right)\left(\det \mathsf{D}_{ij}\right)\left(\det \boldsymbol{A}_i^{\mathsf{h}}\right) \Big] \\
&= \textstyle\sum_{i,j} \Big[\mathrm{tr}\left(\boldsymbol{W}_i \boldsymbol{W}_i^{-1} \boldsymbol{y}_{ij}\boldsymbol{y}_{ij}^{\mathsf{h}} \left(\boldsymbol{W}_i^{\mathsf{h}}\right)^{-1} \boldsymbol{W}_i^{\mathsf{h}} \mathsf{D}_{ij}^{-1} \right) \\
&\qquad\qquad + 2\log|\det \boldsymbol{A}_i| + \log\det \mathsf{D}_{ij} \Big] \\
&= \textstyle\sum_{i,j} \left[\mathrm{tr}\left(\boldsymbol{y}_{ij}\boldsymbol{y}_{ij}^{\mathsf{h}} \mathsf{D}_{ij}^{-1} \right) - 2\log|\det \boldsymbol{W}_i| + \sum_m \log d_{ij,m} \right] \\
&= \textstyle\sum_{i,j} \left[\sum_m \dfrac{|y_{ij,m}|^2}{\sum_k z_{mk} t_{ik} v_{kj}} - 2\log|\det \boldsymbol{W}_i| + \sum_m \log \sum_k z_{mk} t_{ik} v_{kj} \right]
\end{aligned} \tag{14}
$$

但し，$y_{ij,m} = \boldsymbol{w}_{i,m}^{\mathsf{h}} \boldsymbol{x}_{ij}$ である．このコスト関数は，第一項と第二項が空間モデル $\boldsymbol{W}_i$，第一項と第三項が音源モデル $\boldsymbol{TV}$ にそれぞれ対応しているが，これらは下記に示す IVA と単一チャネル NMF のコスト関数と本質的に等価である．

$$Q_{\mathrm{IVA}} = \textstyle\sum_m \dfrac{1}{J} \sum_j G(\boldsymbol{y}_{j,m}) - \sum_i \log|\det \boldsymbol{W}_i| \tag{15}$$

$$Q_{\mathrm{NMF}} = \textstyle\sum_{i,j} \left[\dfrac{|y_{ij}|^2}{\sum_l t_{il} v_{lj}} + \log \sum_l t_{il} v_{lj} \right] \tag{16}$$

但し，$\boldsymbol{y}_{j,m} = (y_{1j,m} \cdots y_{Ij,m})^{\mathsf{t}}$ であり，$G(\boldsymbol{y}_{j,m}) = -\log p(\boldsymbol{y}_{j,m})$ はコントラスト関数と呼ばれる ($p(\boldsymbol{y}_{j,m})$ は $\boldsymbol{y}_{j,m}$ の多変量確率密度関数)．IVA は多変量生成モデルとして，球状ラプラス分布のように球対称かつ優ガウス性の分布を仮定することが一般的である [18], [19]．これは，球対称という性質から，周波数方向に一様な分散を仮定しており，音源モデルとしてはフラットなスペクトル基底を各音源に 1 本ずつ与えていることに対応する．一方，式 (14) では，板倉斎藤擬距離に基づいているため，時間周波数の各スロットで独立な複素ガウス分布を仮定している [22]．また，それらの時間周波数で変動する音源毎の分散値 $r_{ij,m}$ が，基底及びアクティベーションから成る音源モデルとして推定される．

3.3 更新式の導出

ICA や IVA の分離行列の更新については，補助関数法を用いた高速かつ安定な更新式が提案されている [20], [21]．特に，文献 [21] 中のコントラスト関数を $G = |y_{ij,m}|^2 / r_{ij,m}$ ($r_{ij,m}$ は複素ガウス分布の推定分散) とし G に関する期待値演算を省くと，式 (14) 中の分離行列に関する項は文献 [21] と等価となる．以上より，分離行列の補助関数法に基づく更新式は次のように得られる．

$$V_{i,m} = \frac{1}{J} \textstyle\sum_j \dfrac{1}{r_{ij,m}} \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{h}} \tag{17}$$

$$\boldsymbol{w}_{i,m} \leftarrow \left(\boldsymbol{W}_i V_{i,m}\right)^{-1} \boldsymbol{e}_m \tag{18}$$

$$\boldsymbol{w}_{i,m} \leftarrow \boldsymbol{w}_{i,m} \left(\boldsymbol{w}_{i,m}^{\mathsf{h}} V_{i,m} \boldsymbol{w}_{i,m} \right)^{-\frac{1}{2}} \tag{19}$$

但し，$\boldsymbol{e}_m$ は m 番目の要素のみが 1 の単位ベクトルである．分離行列の更新後は，分離信号を $y_{ij,m} \leftarrow \boldsymbol{w}_{i,m}^{\mathsf{h}} \boldsymbol{x}_{ij}$ として更新し，続いて音源モデルの更新を行う．

音源モデルに関する NMF 変数 t_{ik}, v_{kj} 及び潜在変数 z_{mk} の更新式も，補助関数法により導出できる．まず，式 (14) の第一項及び第三項に着目し，補助関数を設計する．凸関数である式 (14) 第一項に対して，$\alpha_{ijk} \geq 0$ かつ $\sum_k \alpha_{ijk} = 1$ を満たす補助変数 α_{ijk} を用いて Jensen の不等式を適用すると，次式が得られる．

$$\frac{1}{\sum_k z_{mk} t_{ik} v_{kj}} \leq \textstyle\sum_k \dfrac{\alpha_{ijk}^2}{z_{mk} t_{ik} v_{kj}} \tag{20}$$

さらに，凹関数である式 (14) 第三項に対して，$\beta_{ij} \geq 0$ を満たす補助変数 β_{ij} を用いて接線不等式を適用すると，次式が得られる．

$$\log \textstyle\sum_k z_{mk} t_{ik} v_{kj} \leq \dfrac{1}{\beta_{ij}} \left(\sum_k z_{mk} t_{ik} v_{kj} - \beta_{ij} \right) + \log \beta_{ij} \tag{21}$$

不等式 (20) 及び (21) の等号成立条件は以下である．

$$\alpha_{ijk} = \frac{z_{mk} t_{ik} v_{kj}}{\sum_{k'} z_{mk'} t_{ik'} v_{k'j}} \tag{22}$$

$$\beta_{ij} = \textstyle\sum_k z_{mk} t_{ik} v_{kj} \tag{23}$$

以上より，式 (14) の補助関数 Q^+ が次式のように得られる．

$$
\begin{aligned}
Q \leq Q^+ = \textstyle\sum_{i,j} \Big[& \textstyle\sum_{m,k} \dfrac{|y_{ij,m}|^2 \alpha_{ijk}^2}{z_{mk} t_{ik} v_{kj}} - 2\log|\det \boldsymbol{W}_i| \\
& + \dfrac{1}{\beta_{ij}} \left(\textstyle\sum_k z_{mk} t_{ik} v_{kj} - \beta_{ij} \right) + \log \beta_{ij} \Big]
\end{aligned} \tag{24}
$$

次に，補助関数 Q^+ を各変数で偏微分する．$\partial Q^+ / \partial z_{mk} = 0$ より，次式が得られる．

$$\textstyle\sum_{i,j} \left[-\dfrac{|y_{ij,m}|^2 \alpha_{ijk}^2}{z_{mk}^2 t_{ik} v_{kj}} + \dfrac{1}{\beta_{ij}} t_{ik} v_{kj} \right] = 0 \tag{25}$$

非負性を保つため上式の第一項を移項し，両辺に z_{mk}^2 を掛けると，次のように変形できる．

Table 1 Music sources

ID	Song	Source (1/2)
1	bearlin-roads__snip_85_99	acoustic_guit_main/vocals
2	another_dreamer-the_ones_we_love	guitar/vocals
3	fort_minor-remember_the_name__snip_54_78	violins_synth/vocals
4	ultimate_nz_tour__snip_43_61	guitar/synth

$$z_{mk}^2 \sum_{i,j} \frac{1}{\beta_{ij}} t_{ik} v_{kj} = \sum_{i,j} \frac{|y_{ij,m}|^2 \alpha_{ijk}^2}{t_{ik} v_{kj}} \tag{26}$$

式 (26) に等号成立条件 (22) 及び (23) を代入し変形すると，z_{mk} に関する更新式が得られる．

$$z_{mk} \leftarrow z_{mk} \sqrt{\frac{\sum_{i,j} |y_{ij,m}|^2 t_{ik} v_{kj} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-2}}{\sum_{i,j} t_{ik} v_{kj} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-1}}} \tag{27}$$

但し，$\sum_m z_{mk} = 1$ を保証するため $z_{mk} \leftarrow z_{mk} / \sum_{m'} z_{m'k}$ を反復の度に計算する．同様にして，t_{ik} 及び v_{kj} の更新式も導出できる．

$$t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_{j,m} |y_{ij,m}|^2 z_{mk} v_{kj} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-2}}{\sum_{j,m} z_{mk} v_{kj} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-1}}} \tag{28}$$

$$v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_{i,m} |y_{ij,m}|^2 z_{mk} t_{ik} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-2}}{\sum_{i,m} z_{mk} t_{ik} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-1}}} \tag{29}$$

これらの変数の更新後は，推定分散を $r_{ij,m} \leftarrow \sum_k z_{mk} t_{ik} v_{kj}$ として更新し，再び分離行列の更新を行う．

以上より，Rank-1 MNMF では，IVA と NMF の更新式を交互に反復することで，全変数を容易に最適化できる．さらに，全変数の最適化が補助関数法に基づいているため，高速で安定な最適化が可能となる．尚，本手法では分離信号のスケールを決める推定分散と分離行列がともに変数となっており，両者の間でスケールの任意性が存在する．そのため，更新の過程でいずれかの変数が発散する危険がある．これを防ぐ為に，次式の正規化を更新の度に施す．

$$\boldsymbol{w}_{i,m} \leftarrow \boldsymbol{w}_{i,m} \lambda_m^{-1}, \quad y_{ij,m} \leftarrow y_{ij,m} \lambda_m^{-1}, \quad r_{ij,m} \leftarrow r_{ij,m} \lambda_m^{-2} \tag{30}$$

但し，λ_m はチャネル毎に求めた正規化係数で，分離信号のパワースペクトル $|y_{ij,m}|^2$ の時間周波数平均値等を用いる．また，式 (30) の正規化はコスト関数 (14) の値を変えないことに留意する．最終的な分離信号のスケールは Projection back [23] で復元することができる．

4. 分離性能比較実験

4.1 実験条件

本稿ではランク 1 空間モデルの妥当性を評価するため，残響時間の異なる 2 種のインパルス応答を用いて従来の MNMF と Rank-1 MNMF の分離性能の比較を行う．実験では，IVA，MNMF，Rank-1 MNMF の 3 手法に加え，Rank-1 MNMF で推定した分離行列から各音源の空間相関行列を逆算し，MNMF の初期値に用いる MNMF with rank-1 initialization を比較する．音源は SiSEC [24] で公開されているプロ音楽信号を用いた (Table 1 参照)．また，RWCP database [25] に収録されている E2A 及び JR2 のインパルス応答 (Fig. 2 参照) を用いて畳み込み，2 チャネル 2 音源の観測信号を作成した．その他の実験条件は Table 2 に示す通りである．分離精度を示す客観評価尺度には，文献 [26] で定義されている signal-to-distortion ratio (SDR) を用いた．SDR は分離度合いと音質を加味した総合的な分離性能を示す良い指標となる．

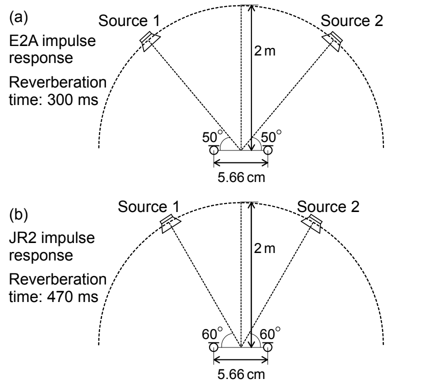

Fig. 2 Recording conditions of room impulse response.

Table 2 Experimental conditions

Sampling frequency	Downsampled from 44.1 kHz to 16 kHz
FFT length	512 ms
Window shift	128 ms
Number of bases K	60 for all sources
Number of iterations	200

4.2 実験結果

Figures 3–6 は，全変数の初期値を変えて 10 回試行した際の平均 SDR 改善量とその標準偏差を示している．但し，MNMF with rank-1 initialization では，Rank-1 MNMF の 10 回試行の最高性能時の分離行列 $\boldsymbol{W}$ から逆算した値を空間相関行列 H の初期値に用いており，その他の変数には乱数を与えている．E2A を用いたデータに対して，MNMF と Rank-1 MNMF では，後者がより高い分離性能を示している．また，初期値依存による性能のばらつきは Rank-1 MNMF の方が小さい場合が多く，頑健に最適化できていることがわかる．MNMF with rank-1 initialization は，より高い分離精度を示しており，ばらつきも大きく改善していることが確認できる．一方，残響が長い JR2 のデータでは，全手法において全体的に分離精度が低下しており，性能のばらつきも大きくなっている．しかし，MNMF with rank-1 initialization では他の手法と比較して分離精度が大きく向上している．これは，長い残響に起因して混合系のランク 1 モデルが成り立たなくなっている事を示している．しかしながら，そのような状況においても，一度ランク 1 モデルで推定した初期値を用いて，改めてフルランクで空間相関行列を推定することで，より高精度かつ頑健な音源分離を実現することが可能となっている．

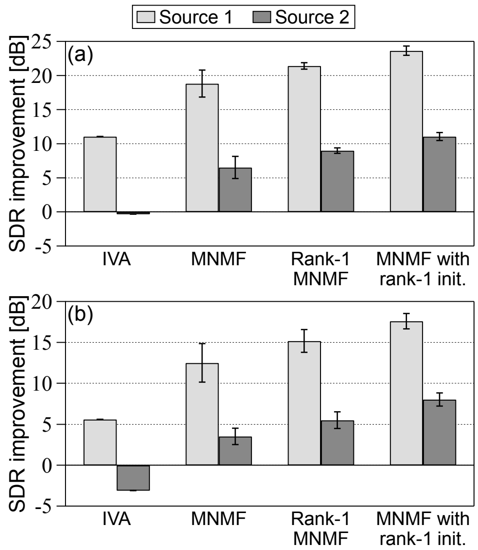

Fig. 3 Scores for ID1 song with (a) E2A and (b) JR2 impulse responses.

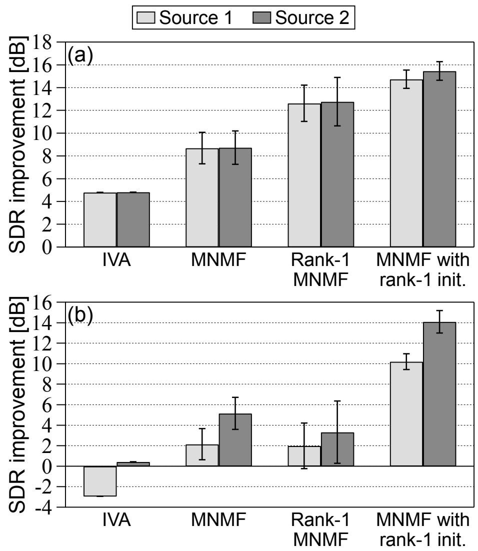

Fig. 4 Scores for ID2 song with (a) E2A and (b) JR2 impulse responses.

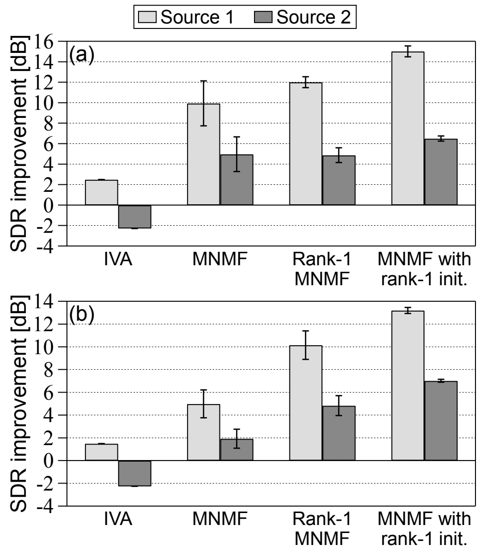

Fig. 5 Scores for ID3 song with (a) E2A and (b) JR2 impulse responses.

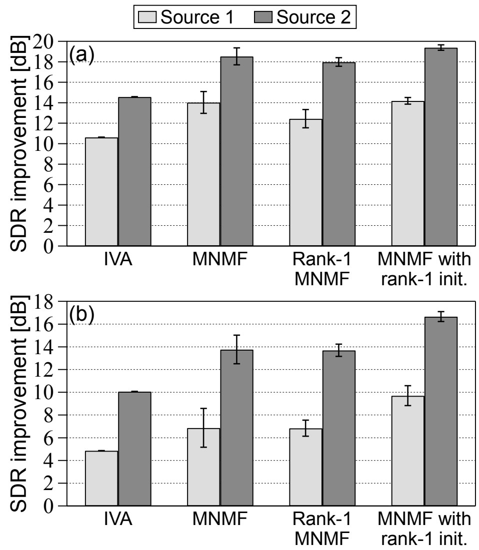

Fig. 6 Scores for ID4 song with (a) E2A and (b) JR2 impulse responses.

Figures 7–8 は，ID1 の楽曲に対する各手法の SDR 改善量の収束例を示している．この結果から，Rank-1 MNMF 及び MNMF with rank-1 initialization は E2A 及び JR2 の両データに対して少ない反復回数で高い SDR に到達していることが確認できる．しかしながら，MNMF はより多くの反復を必要とし，JR2 のデータに対してはさらに顕著となっている．この結果から，MNMF におけるフルランクの空間相関行列の推定の困難さがうかがえる．

Table 3 は，ID1 の楽曲に対する各手法の 200 回更新時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 (3.60 GHz) の CPU を用いている．この結果から，Rank-1 MNMF の計算時間は IVA の 2 倍程度であり，従来の MNMF と比較して効率的であることが確認できる．

4.3 まとめ

本稿では，従来のフルランク空間相関行列を推定する MNMF と，ランク 1 近似を導入した Rank-1 MNMF を実験的に比較し，ランク 1 空間モデルを用いた最適化の有用性及びその妥当性に関して考察を加えた．残響等の影響でランク 1 近似が成り立たなくなった場合，Rank-1 MNMF は従来の MNMF と同程度の分離性能となり，頑健性が失われる場合も確認できた．しかしながら，一度ランク 1 近似で推定した空間モデルを従来の MNMF の初期に用いた場合は，高精度な分離を頑健に達成することが確認できた．

謝辞 本研究の一部は JSPS 特別研究員奨励費 26·10796 の助成を受けたものである．

参考文献

[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. Araki, R. Mukai, S. Makino, T. Nishikawa and H. Saruwatari, “The fundamental limitation of frequency do-

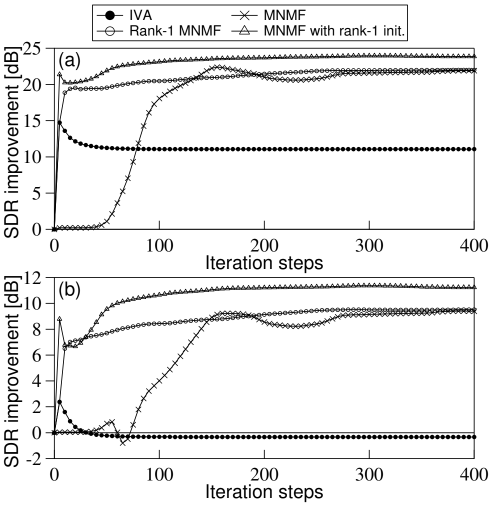

Fig. 7 Example of SDR convergence for ID1 song with E2A impulse response: (a) source 1 and (b) source 2.

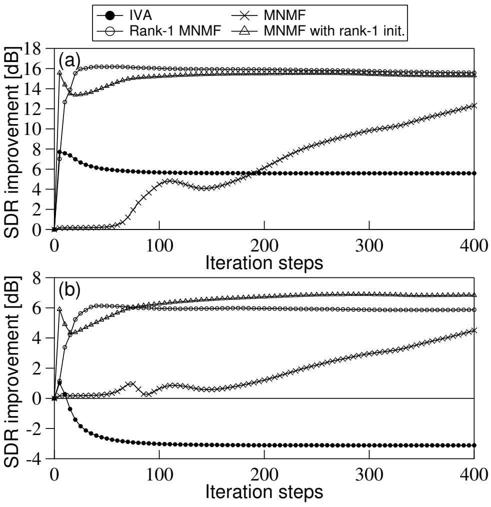

Fig. 8 Example of SDR convergence for ID1 song with JR2 impulse response: (a) source 1 and (b) source 2.

Table 3 Computational times for separation of ID1 (s)

IVA	MNMF	Rank-1 MNMF
41.8	304.1	87.9

main blind source separation for convolutive mixtures of speech," *IEEE Trans. SAP*, vol.11, no.2, pp.109–116, 2003.

[4] H. Sawada, R. Mukai, S. Araki and S. Makino, "Convolutive blind source separation for more than two sources in the frequency domain," *Proc. ICASSP*, pp.Ⅲ-885–Ⅲ-888, 2004.

[5] H. Buchner, R. Aichner and W. Kellerman, "A generalization of blind source separation algorithms for convolutive mixtures based on second order statistics," *IEEE Trans. SAP*, vol.13, no.1, pp.120–134, 2005.

[6] 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.

[7] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Proc. NIPS*, vol.13, pp.556–562, 2001.

[8] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. ASLP*, vol.15, no.3, pp.1066–1074, 2007.

[9] A. Ozerov, C. Févotte and M. Charbit, "Factorial scaled hidden Markov model for polyphonic audio representation and source separation," *Proc. WASPAA*, pp.121–124, 2009.

[10] 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.

[11] P. Smaragdis, B. Raj and M. Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," *Proc. ICA*, pp.414–421, 2007.

[12] 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 of Electronics, Communications and Computer Sciences*, vol.E97-A, no.5, pp.1113–1118, 2014.

[13] 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.

[14] S. Arberet, A. Ozerov, N.Q.K. Duong, E. Vincent, R. Gribonval, R. Bimbot and P. Vandergheynst, "Nonnegative matrix factorization and spatial covariance model for under-determined reverberant audio source separation," *Proc. ISSPA*, pp.1–4, 2010.

[15] 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.

[16] D. Kitamura, N. Ono, H. Sawada, H. Kameoka and H. Saruwatari, "Efficient multichannel nonnegative matrix factorization with rank-1 spatial model," *Proc. 2014 Autumn Meeting of ASJ*, pp.579–582, 2014 (in Japanese).

[17] D. Kitamura, N. Ono, H. Sawada, H. Kameoka and H. Saruwatari, "Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model," *Proc. ICASSP*, 2015 (in press).

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

[19] 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.

[20] N. Ono and S. Miyabe, "Auxiliary-function-based independent component analysis for super-Gaussian sources," *Proc. LVA/ICA*, pp.165–172, 2010.

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

[22] 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.

[23] 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.

[24] 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*, pp.414–422, 2012.

[25] 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.

[26] 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.