

I ダイバージェンスに基づく独立低ランク行列分析の実験的評価*

☆最上伸一 (東大), 三井祥幹 (東大), 高宗典玄 (東大), 北村大地 (東大), 猿渡洋 (東大), 高橋祐 (ヤマハ), 近藤多伸 (ヤマハ), 中嶋広明 (ヤマハ), 亀岡弘和 (NTT)

1 はじめに

ブラインド音源分離 (blind source separation: BSS) とは, 複数の音源が混合された観測信号のみを用いて, 混合前の個々の音源を推定する技術である. 音源信号, 観測信号及び分離信号をそれぞれ

$$\mathbf{s}_{ij} = (s_{ij1}, \dots, s_{ijN})^\top \in \mathbb{C}^N \quad (1)$$

$$\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijM})^\top \in \mathbb{C}^M \quad (2)$$

$$\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijN})^\top \in \mathbb{C}^N \quad (3)$$

とする. ここで, $i = 1, \dots, I$ 及び $j = 1, \dots, J$ はそれぞれ周波数ビン及び時間フレームのインデックス, $n = 1, \dots, N$ 及び $m = 1, \dots, M$ はそれぞれ音源, チャネルのインデックス, $*^\top$ はベクトル及び行列の転置を表す. 今, 観測信号が

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

と表現できると仮定する. ここで, $\mathbf{A}_i \in \mathbb{C}^{M \times N}$ は周波数ごとの混合行列である. さらに $N = M$ かつ \mathbf{A}_i を正則と仮定すると, 分離信号は

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

と推定できる. ここで, $\mathbf{W}_i = (\mathbf{w}_{i1} \dots \mathbf{w}_{iN})^\mathbf{H} = \mathbf{A}_i^{-1}$, \mathbf{w}_{in} は n 番目の音源の分離フィルタ, $*^\mathbf{H}$ はエルミート転置である. 近年, 音源間の統計的独立性と各音源の低ランク性を仮定した分離行列 \mathbf{W}_i の推定法として, 独立低ランク行列分析 (independent low-rank matrix analysis: ILRMA) [1] が提案されており, 高精度な BSS を実現している.

一方, 非負値行列因子分解 (nonnegative matrix factorization: NMF) を多チャンネル信号に拡張した多チャンネル NMF (multichannel NMF: MNMF) [2] は, 全音源の低ランク時間周波数構造と音源毎の空間相関行列 (混合系に対応する空間パラメータ) を同時に推定する BSS である. 最適化の観点では, 分離系の推定を行う ILRMA の方が MNMF よりも高速かつ高精度であることが報告されている [1]. 従来の MNMF 及び ILRMA は, 板倉斎藤ダイバージェンスに基づく NMF に対応する原点对称複素正規分布生成モデルを仮定していたが, 近年では原点对称複素 Student's t 分布に基づく MNMF (t -MNMF) [3] 及び ILRMA (t -ILRMA) [4] に一般化されている. さらに著者らは, 複素 Poisson 分布 [5] を生成モデルに用いることで, 一般化 Kullback-Leibler (KL) ダイバージェンス (I ダイバージェンス) 規範で各音源を低ランク近似する ILRMA (KL-ILRMA) を提案している [6]. 本稿では区別のため, 従来の ILRMA を IS-ILRMA と表記する. 既存手法と KL-ILRMA の関係を Fig. 1 に示す.

KL-ILRMA の音源モデル推定は一般化 KL ダイバージェンスの最小化に基づくため, その凸性より初期値等への頑健性が期待されるが, 文献 [6] では十分な検証が行われていなかった. そこで本稿では, KL-ILRMA の性能を評価するためにより詳細な実験を行う. まず, トイモデルを用いて, 複素 Poisson 分布に従う音源分離に KL-ILRMA が適していること

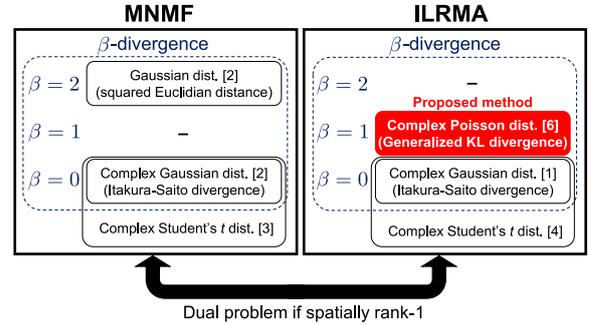


Fig. 1 Related methods of MNMF and ILRMA.

を示す. 次に, 音楽信号に対する音源分離実験を通して複数の従来手法と比較し, KL-ILRMA の有効性及び初期値に対する頑健性を示す.

2 I ダイバージェンスに基づく ILRMA

2.1 生成モデル及びコスト関数

KL-ILRMA では, 複素 Poisson 分布 [5] を音源モデルに仮定することで, 一般化 KL ダイバージェンスに基づいて音源を低ランク近似する. 複素 Poisson 分布の確率密度関数は, $\mathcal{D} = \{z \in \mathbb{C} \mid |z| \in \mathbb{N}\}$ を台とする複素数 z に対して次式で与えられる.

$$p(z) = \frac{p |z|^{p-2} \lambda |z|^p}{2\pi (|z|^p)!} e^{-\lambda} \quad (6)$$

NMF に基づく単一チャンネル音源分離では, 振幅スペクトログラムを用いた一般化 KL ダイバージェンス規範 NMF (KL-NMF) が音源分離に適している [7] ことから, 本稿でも $p = 1$ のみを考える.

KL-ILRMA では, 各音源のスペクトログラムの複素時間周波数成分 y_{ijn} が, 時間周波数ごとに異なるパラメータ λ_{ijn} を持つ独立な複素 Poisson 分布に従って生成されると仮定する. 従って, KL-ILRMA の音源モデルは以下の式で与えられる.

$$\prod_{i,j,n} p(y_{ijn}) = \prod_{i,j,n} \frac{|y_{ijn}|^{-1} \lambda_{ijn}^{|y_{ijn}|}}{2\pi |y_{ijn}|!} e^{-\lambda_{ijn}} \quad (7)$$

また, λ_{ijn} は低ランクな非負スペクトログラムとして, 以下のようにモデル化される.

$$\lambda_{ijn} = \sum_k t_{ikn} v_{kjn} \quad (8)$$

ここで, t_{ikn} 及び v_{kjn} は, それぞれ基底行列 $\mathbf{T}_n \in \mathbb{R}_+^{I \times K}$ 及びアクティベーション行列 $\mathbf{V}_n \in \mathbb{R}_+^{K \times J}$ の成分である. \mathbb{R}_+ は非負値実数の集合を表す. $k = 1, \dots, K$ は基底インデックスであり, 基底の本数 K は通常 I, J と比較して非常に小さい値とする. 式 (7) の負対数尤度は, $y_{ijn} = \mathbf{w}_{in}^\mathbf{H} \mathbf{x}_{ij}$ から

$$\mathcal{L} = \sum_{i,j,n} \left[\log(|\mathbf{w}_{in}^\mathbf{H} \mathbf{x}_{ij}|!) + \log|\mathbf{w}_{in}^\mathbf{H} \mathbf{x}_{ij}| \right]$$

*Experimental evaluation of independent low-rank matrix analysis based on I divergence by Shinichi Mogami, Yoshihiro Mitsui, Norihiro Takamune, Daichi Kitamura, Hiroshi Saruwatari (The University of Tokyo), Yu Takahashi, Kazunobu Kondo, Hiroaki Nakajima (Yamaha), Hirokazu Kameoka (NTT).

$$\begin{aligned}
& - |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \log \sum_k t_{ikn} v_{kjn} + \sum_k t_{ikn} v_{kjn} \Big] \\
& - 2J \sum_i \log |\det \mathbf{W}_i| + \text{const.} \quad (9)
\end{aligned}$$

となる。 $\log(|\mathbf{w}_{in}^H \mathbf{x}_{ij}|!)$ は $|\mathbf{w}_{in}^H \mathbf{x}_{ij}|$ に関する離散的な関数であるが、ここに Stirling の近似

$$\log(x!) \approx x \log x - x \quad (10)$$

を導入することにより、負対数尤度の式 (9) を初等関数で近似しつつ $\mathbf{w}_{in}^H \mathbf{x}_{ij}$ の変域を複素数平面全域に拡張した以下のコスト関数を得る。

$$\begin{aligned}
\mathcal{J} = \sum_{i,j,n} \Big[& |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \log |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \\
& - |\mathbf{w}_{in}^H \mathbf{x}_{ij}| + \log |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \\
& - |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \log \sum_k t_{ikn} v_{kjn} + \sum_k t_{ikn} v_{kjn} \Big] \\
& - 2J \sum_i \log |\det \mathbf{W}_i| + \text{const.} \quad (11)
\end{aligned}$$

これは、一般化 KL ダイバージェンス $D_{\text{KL}}(y | x) = y \log y - y \log x - y + x$ を用いて

$$\begin{aligned}
\mathcal{J} = \sum_{i,j,n} \Big[& D_{\text{KL}}(|\mathbf{w}_{in}^H \mathbf{x}_{ij}| | \sum_k t_{ikn} v_{kjn}) \\
& + \log |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \Big] \\
& - 2J \sum_i \log |\det \mathbf{W}_i| + \text{const.} \quad (12)
\end{aligned}$$

とも書けることから、音源間の独立性最大化と一般化 KL ダイバージェンスに基づく低ランク音源モデル推定を同時に達成するコストであると解釈できる。

2.2 音源モデルの更新式

式 (12) で与えられるコスト関数 \mathcal{J} を $\mathbf{T}_n, \mathbf{V}_n$ についてみれば、これらが関与するのは $D_{\text{KL}}(|\mathbf{w}_{in}^H \mathbf{x}_{ij}| | \sum_k t_{ikn} v_{kjn})$ の項のみである。従って、 \mathbf{T}_n 及び \mathbf{V}_n は KL-NMF [8] と同様に以下の乗算更新式を反復することで最適化できる。

$$t_{ikn} \leftarrow t_{ikn} \cdot \frac{\sum_j \frac{|\mathbf{w}_{in}^H \mathbf{x}_{ij}|}{\sum_{k'} t_{ik'n} v_{k'jn}} v_{kjn}}{\sum_j v_{kjn}} \quad (13)$$

$$v_{kjn} \leftarrow v_{kjn} \cdot \frac{\sum_i \frac{|\mathbf{w}_{in}^H \mathbf{x}_{ij}|}{\sum_{k'} t_{ik'n} v_{k'jn}} t_{ikn}}{\sum_i t_{ikn}} \quad (14)$$

上記の一般化 KL ダイバージェンス最小化問題は各変数に関して凸であることが知られており、それが従来の MNMF や ILRMA とは異なる利点であると期待される。

2.3 分離行列の更新式

従来手法の IS-ILRMA は、分離行列の更新に iterative projection (IP) [9] と呼ばれる手法を用いた。しかしながら、IP はコスト関数が $-\log |\det \mathbf{W}_i|$ と \mathbf{w}_{in} の二次形式の和で構成される場合にのみ適用でき、KL-ILRMA のコスト関数 (11) の形には適用できない。そこで、補助関数法を用いて分離行列 \mathbf{W}_i の反復更新式を導出する。まず、式 (11) の $\log |\mathbf{w}_{in}^H \mathbf{x}_{ij}|$

に関して次の接線不等式

$$\log |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \leq \frac{1}{\alpha_{ijn}} (|\mathbf{w}_{in}^H \mathbf{x}_{ij}| - \alpha_{ijn}) + \log \alpha_{ijn} \quad (15)$$

を適用すると、次の補助関数 \mathcal{J}_1 が得られる。

$$\begin{aligned}
\mathcal{J} \leq \mathcal{J}_1 = \sum_{i,j,n} \Big[& \frac{1}{\alpha_{ijn}} |\mathbf{w}_{in}^H \mathbf{x}_{ij}|^2 + d_{ijn} |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \Big] \\
& - 2J \sum_i \log |\det \mathbf{W}_i| + \text{const.} \quad (16)
\end{aligned}$$

$$d_{ijn} = \frac{1}{\alpha_{ijn}} + \log \alpha_{ijn} - \log \sum_k t_{ikn} v_{kjn} - 2 \quad (17)$$

ここで $\alpha_{ijn} > 0$ は補助変数であり、式 (15) は $\alpha_{ijn} = |\mathbf{w}_{in}^H \mathbf{x}_{ij}|$ のときに等号成立する。

続いて、 \mathbf{w}_{in} についての微分が可能となるように $d_{ijn} |\mathbf{w}_{in}^H \mathbf{x}_{ij}|$ の上界関数を設計することを考える。Fig. 2 が示す通り、 d の符号で場合分けを行うことにより、任意の y_0 に対して $d|y|$ に点 y_0 (Fig. 2 では黒の点線) で一致するような上界関数を、 y について 2 次 (放物面) または 1 次 (平面) の形で構成できる。よって、 $d_{ijn} |\mathbf{w}_{in}^H \mathbf{x}_{ij}|$ に対する上界関数の式は

$$d_{ijn} |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \leq \begin{cases} \frac{d_{ijn}}{2\beta_{ijn}} |\mathbf{w}_{in}^H \mathbf{x}_{ij}|^2 + \frac{1}{2} d_{ijn} \beta_{ijn} & (d_{ijn} \geq 0) \\ \frac{1}{2} d_{ijn} (\overline{\omega_{ijn}} \mathbf{w}_{in}^H \mathbf{x}_{ij} + \omega_{ijn} \overline{\mathbf{w}_{in}^H \mathbf{x}_{ij}}) & (d_{ijn} < 0) \end{cases} \quad (18)$$

と表される。ここで、 $\bar{*}$ は $*$ の複素共役を表す。 $\beta_{ijn} > 0$ は実数の補助変数であり、等号は $\beta_{ijn} = |\mathbf{w}_{in}^H \mathbf{x}_{ij}|$ にて成立する。また、 ω_{ijn} は $|\omega_{ijn}| = 1$ を満たす複素数の補助変数であり、等号は $\omega_{ijn} = \mathbf{w}_{in}^H \mathbf{x}_{ij} / |\mathbf{w}_{in}^H \mathbf{x}_{ij}|$ のときに成立する。式 (18) を式 (16) に適用すると、次の補助関数 \mathcal{J}_2 が得られる。

$$\begin{aligned}
\mathcal{J}_1 \leq \mathcal{J}_2 = J \sum_{i,n} \Big[& \mathbf{w}_{in}^H \mathbf{U}_{in} \mathbf{w}_{in} + \mathbf{w}_{in}^H \mathbf{r}_{in} + \mathbf{r}_{in}^H \mathbf{w}_{in} \Big] \\
& - 2J \sum_i \log |\det \mathbf{W}_i| + \text{const.} \quad (19)
\end{aligned}$$

$$\mathbf{U}_{in} = \frac{1}{J} \sum_j \left(\frac{1}{\alpha_{ijn}} + \frac{\max(0, d_{ijn})}{2\beta_{ijn}} \right) \mathbf{x}_{ij} \mathbf{x}_{ij}^H \quad (20)$$

$$\mathbf{r}_{in} = \frac{1}{J} \sum_j \frac{\overline{\omega_{ijn}}}{2} \min(d_{ijn}, 0) \mathbf{x}_{ij} \quad (21)$$

式 (19) は \mathbf{w}_{in} の一次項を含むため、依然として IP を適用することはできない。このような一次項を含む関数の最小化には、著者らの一部が開発した vectorwise coordinate descent (VCD) [10] が適用できる。VCD では、まず変数行列 \mathbf{W}_i の特定の行ベクトル \mathbf{w}_{in} に着目し、余因子展開によって式 (19) を以下の形式に変形する。

$$\begin{aligned}
\mathcal{J}_2 = J \sum_{i,n} \Big[& \mathbf{w}_{in}^H \mathbf{U}_{in} \mathbf{w}_{in} + \mathbf{w}_{in}^H \mathbf{r}_{in} + \mathbf{r}_{in}^H \mathbf{w}_{in} \Big] \\
& - J \sum_i \log |\mathbf{b}_{in}^H \mathbf{w}_{in}|^2 + \text{const.} \quad (22)
\end{aligned}$$

ここで、 \mathbf{b}_{in} は \mathbf{W}_i の余因子行列 $\mathbf{B}_i = (\mathbf{b}_{i1}, \dots, \mathbf{b}_{iN})$ の列ベクトルであり、 \mathbf{e}_n を n 番目の要素のみが 1 の

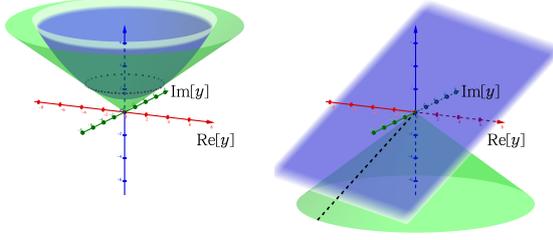


Fig. 2 Shapes of auxiliary function (blue) of original function $d|y|$ (green) with contact point y_0 (black dotted line). When $d > 0$, $d|y|$ can be majorized by paraboloid of revolution (left). When $d < 0$, $d|y|$ can be majorized by plane (right).

単位ベクトルとして $\mathbf{b}_{in} = (\det \mathbf{W}_i) \mathbf{W}_i^{-1} \mathbf{e}_n$ と書ける. $\mathbf{b}_{in}, \mathbf{U}_{in}$ 及び \mathbf{r}_{in} は \mathbf{w}_{in} と独立であるため, 式 (22) は他の行ベクトルを固定することで \mathbf{w}_{in} のみの関数となり, \mathbf{w}_{in} をブロック単位とした座標降下法が適用できる. 式 (22) の \mathbf{w}_{in} に関する停留点条件

$$\left(\frac{\partial \mathcal{J}_2}{\partial \mathbf{w}_{in}^H} \right) - \frac{\mathbf{b}_{in}}{\mathbf{w}_{in}^H \mathbf{b}_{in}} + \mathbf{U}_{in} \mathbf{w}_{in} + \mathbf{r}_{in} = \mathbf{0} \quad (23)$$

を解くと, 停留点は次式で与えられる [10].

$$\mathbf{w}_{in} = \begin{cases} \mathbf{U}_{in}^{-1} \begin{bmatrix} \frac{1}{\sqrt{u_{bb}}} \mathbf{b}_{in} - \mathbf{r}_{in} \end{bmatrix} & (\text{if } u_{br} = 0) \\ \mathbf{U}_{in}^{-1} \begin{bmatrix} \frac{u_{br}}{2u_{bb}} \left(1 - \sqrt{1 + \frac{4u_{bb}}{|u_{br}|^2}} \right) \mathbf{b}_{in} - \mathbf{r}_{in} \end{bmatrix} & (\text{otherwise}) \end{cases} \quad (24)$$

ここで, $u_{bb} = \mathbf{b}_{in}^H \mathbf{U}_{in}^{-1} \mathbf{b}_{in}$, $u_{br} = \mathbf{b}_{in}^H \mathbf{U}_{in}^{-1} \mathbf{r}_{in}$ である. 式 (24) に対して各補助変数の等号成立条件を代入することで, \mathbf{w}_{in} の更新式が次のように得られる.

$$d_{ijn} \leftarrow \frac{1}{|\mathbf{w}_{in}^H \mathbf{x}_{ij}|} + \log |\mathbf{w}_{in}^H \mathbf{x}_{ij}| - \log \sum_k t_{ikn} v_{kjn} - 2 \quad (25)$$

$$\mathbf{U}_{in} \leftarrow \frac{1}{J} \sum_j \left(\frac{1}{|\mathbf{w}_{in}^H \mathbf{x}_{ij}|} + \frac{\max(0, d_{ijn})}{2|\mathbf{w}_{in}^H \mathbf{x}_{ij}|} \right) \mathbf{x}_{ij} \mathbf{x}_{ij}^H \quad (26)$$

$$\tilde{\mathbf{r}}_{in} \leftarrow \frac{1}{J} \mathbf{U}_{in}^{-1} \sum_j \frac{1}{2} \frac{\mathbf{w}_{in}^H \mathbf{x}_{ij}}{|\mathbf{w}_{in}^H \mathbf{x}_{ij}|} \min(d_{ijn}, 0) \mathbf{x}_{ij} \quad (27)$$

$$\tilde{\mathbf{w}}_{in} \leftarrow (\mathbf{W}_i \mathbf{U}_{in})^{-1} \mathbf{e}_n \quad (28)$$

$$u_{ww} \leftarrow \tilde{\mathbf{w}}_{in}^H \mathbf{U}_{in} \tilde{\mathbf{w}}_{in} \quad (29)$$

$$u_{wr} \leftarrow \tilde{\mathbf{w}}_{in}^H \mathbf{U}_{in} \tilde{\mathbf{r}}_{in} \quad (30)$$

$$\mathbf{w}_{in} \leftarrow \begin{cases} \frac{\tilde{\mathbf{w}}_{in}}{\sqrt{u_{ww}}} - \tilde{\mathbf{r}}_{in} & (\text{if } u_{wr} \det \mathbf{W}_i = 0) \\ \frac{u_{wr}}{2u_{ww}} \left(1 - \sqrt{1 + \frac{4u_{ww}}{|u_{wr}|^2}} \right) \tilde{\mathbf{w}}_{in} - \tilde{\mathbf{r}}_{in} & (\text{otherwise}) \end{cases} \quad (31)$$

KL-ILRMA では, 音源モデル \mathbf{T}_n 及び \mathbf{V}_n の更新式 (13), (14) と, 分離行列 \mathbf{W}_i の更新式 (25)–(31) を交互に繰り返すことによって, 式 (11) で表されるコスト関数 \mathcal{J} の最小化を達成する. これらの更新式は全て補助関数法で導出されていることから, 更新における \mathcal{J} の単調非増加性が保証されている.

Table 1 Generating conditions of toy models

Size of spectrogram	$I = 257, J = 514$
Number of bases	$K = 10$
Scale parameter c	1, 2, 3, 4, 5, 10, 15, 20
Kurtosis of gamma distribution	$\tilde{\mathbf{T}}_n: 12, \tilde{\mathbf{V}}_n: 2$

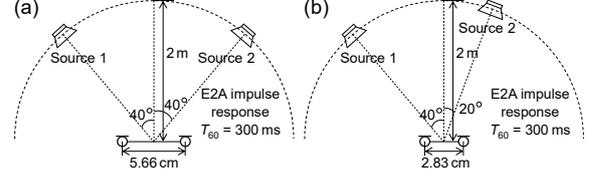


Fig. 3 Recording conditions of impulse responses E2A obtained from RWCP database [12]: (a) IR1 and (b) IR2.

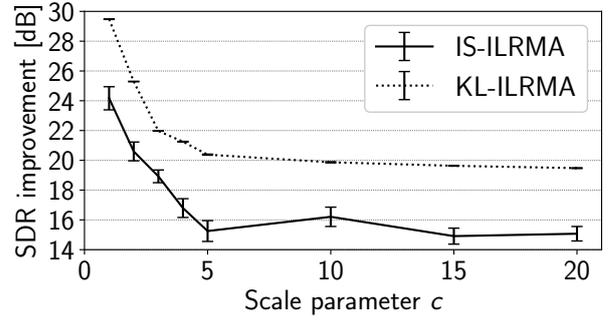


Fig. 4 Average SDR improvements of IS-ILRMA and KL-ILRMA for toy model separation.

3 性能比較実験

3.1 トイモデルを用いた性能評価

本稿では, 初めにトイモデルを用いて, KL-ILRMA が複素 Poisson 分布に従う音源の分離に適していることを確かめる. まず, 独立なガンマ分布に従って行列 $\tilde{\mathbf{T}}_n \in \mathbb{R}_+^{I \times K}$ 及び $\tilde{\mathbf{V}}_n \in \mathbb{R}_+^{K \times J}$ ($n = 1, 2$) の各成分をランダムに生成し, $\tilde{\mathbf{T}}_n \tilde{\mathbf{V}}_n$ の最大値を 1 に正規化することでモデル行列 \mathbf{R}_n を得る. 次に, \mathbf{R}_n をスケール倍した行列 $c\mathbf{R}_n$ の各 i, j 成分をパラメータ λ とする複素 Poisson 分布から, 分離信号のスペクトログラム $\mathbf{Y}_n = (y_{ijn})_{i,j}$ を生成する. 以上の手順により, 複素 Poisson 分布に従う音源のトイモデルを得た. トイモデルの生成条件の詳細を Table 1 に示す. 複素 Poisson 分布 (6) の形状はパラメータ λ のスケールによって変化し, λ が大きいほど振幅 $|z|$ が平均値 λ 付近の値を取りやすくなる. したがって, スケールパラメータ c の値が小さいとき \mathbf{Y}_n の振幅値はスパースになり, 逆に大きいときは \mathbf{Y}_n の振幅値は $c\mathbf{R}_n$ に近づき低ランク性の強い行列となる.

以上より得られる各音源のトイモデルをそれぞれ時間信号に変換し, Fig. 3 の (a) に示すインパルス応答 (IR1) を畳み込むことで, 本実験で用いる混合音を作成した. これを従来手法の IS-ILRMA と提案手法の KL-ILRMA で分離し, その性能を signal-to-distortion ratio (SDR) [11] 改善量によって評価した. なお, 各 ILRMA の音源モデルの基底数 K はトイモデル生成時のパラメータ行列 $\tilde{\mathbf{T}}_n \tilde{\mathbf{V}}_n$ のランクと同一の 10 本とし, 反復回数は 200 回とした. また, 分離行列の初期値には理想解に対して 5% の複素正規乱数を加えたものを用い, 音源モデル \mathbf{T}_n 及び \mathbf{V}_n の初期値には一様乱数を用いた.

Table 2 Music sources obtained from SiSEC2011

Index	Source (1/2)	Impulse response
No. 1	A. guitar/vocal	IR1
No. 2	A. guitar/vocal	IR2
No. 3	A. guitar/piano	IR1
No. 4	A. guitar/piano	IR2

Table 3 Experimental conditions

Sampling frequency	16 000 Hz
STFT	512-ms-long Hamming window with 128-ms shift
Number of iterations	200
Number of bases	10
Number of trials	20

スケールパラメータ c に対する各手法の性能を Fig. 4 に示す。ただし、 \mathbf{T}_n 及び \mathbf{V}_n の初期値の乱数シードを変えて 20 回試行した際の平均 SDR 改善量を示し、エラーバーは標準偏差を表す。KL-ILRMA の性能は IS-ILRMA の性能を任意の c で上回っており、KL-ILRMA が複素 Poisson 分布に従う信号の分離に有効であることが示されている。また、KL-ILRMA の標準偏差は最大で 0.03 dB 程度と IS-ILRMA に比べて小さく、KL-ILRMA はより頑健な音源モデルの推定が可能であることが示唆されている。なお、提案 KL-ILRMA においては、式 (10) の近似を導入しているためスケールの大きな所では生成分布との不整合が生じているが、20 dB 程度の SDR は確保されていることより実用上は妥当な近似であることも示されている。

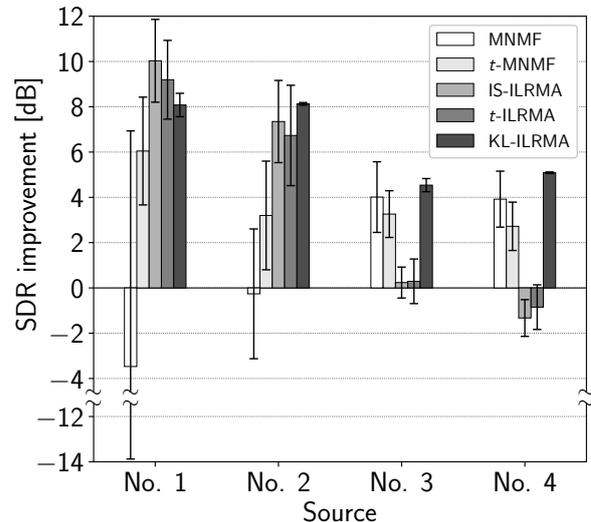
3.2 音楽信号の分離実験による既存手法との比較

次に音楽信号分離について、KL-ILRMA と従来手法である MNMF, t -MNMF, IS-ILRMA 及び t -ILRMA との性能比較を行う。音楽信号として、SiSEC2011 [13] で公開されている音楽信号 *berlin-roads* のうち A. guitar/vocal 及び A. guitar/piano の 2 種類の組合せを用いた。さらに各々に対して Fig. 3 の (a), (b) に示す 2 種類のインパルス応答 (IR1 及び IR2) を畳み込むことで、計 4 種類の混合音 No. 1–No. 4 を作成した。各混合音における実験条件の組合せを Table 2 に示す。IS-ILRMA, t -ILRMA 及び KL-ILRMA では、分離行列の初期値を単位行列、音源モデル \mathbf{T}_n 及び \mathbf{V}_n の初期値を一様乱数とした。MNMF 及び t -MNMF では、文献 [2] と同様の初期値を用いた。 t -MNMF の自由度パラメータは $\nu = 1$ 、 t -ILRMA の自由度パラメータ及びドメインパラメータはそれぞれ $\nu = 3, p = 2$ とし、 t -ILRMA では文献 [4] に基づいて ν に関する tempering 処理を行った。その他の実験条件を Table 3 に示す。

各音源における手法ごとの SDR 改善量の平均を Fig. 5 に示す。エラーバーは標準偏差を表す。No. 2–No. 4 の 3 種類の混合音に対して、KL-ILRMA は他手法と比較して平均的に優れた性能を示している。また、いずれの混合音においても KL-ILRMA の標準偏差は他手法と比較して小さい。これは、凸なダイバージェンスを用いて音源モデルを推定する KL-ILRMA が音源初期値に対して頑健であることを示している。

4 おわりに

本稿では、凸な一般化 KL ダイバージェンスを用いて音源モデルを低ランク近似する KL-ILRMA の性能を詳細な実験によって比較した。トイモデル実

Fig. 5 Average SDR improvements for MNMF, t -MNMF, IS-ILRMA, t -ILRMA, and KL-ILRMA.

験では、複素 Poisson 分布に従う音源に KL-ILRMA が適していることが示された。また、音楽信号の分離実験では、他の分離手法と比較して KL-ILRMA が主に頑健性の面で優れていることを示した。

謝辞 本研究は、セコム科学技術振興財団、JSPS 科研費 17H06572 の助成を受けた。

参考文献

- [1] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, 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.
- [2] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, “Multi-channel extensions of non-negative matrix factorization with complex-valued data,” *IEEE Trans. ASLP*, vol. 21, no. 5, pp. 971–982, 2013.
- [3] K. Kitamura, Y. Bando, K. Itoyama, and K. Yoshii, “Student’s t multichannel nonnegative matrix factorization for blind source separation,” in *Proc. IWAENC*, 2016.
- [4] 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,” in *Proc. MLSP*, 2017.
- [5] H. Kameoka, “Towards a statistical audio signal processing framework based on the I-divergence,” in *Proc. 2011 Spring Meeting of Acoustical Society of Japan*, 2011, pp. 813–814.
- [6] 最上伸一, 三井祥幹, 高宗典玄, 北村大地, 猿渡洋, 高橋祐, and 近藤多伸, “I ダイバージェンスを用いた独立低ランク行列分析,” in 第 32 回信号処理シンポジウム, 2017, pp. 354–359.
- [7] 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.
- [8] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Proc. NIPS*, 2000, pp. 556–562.
- [9] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” in *Proc. WASPAA*, 2011, pp. 189–192.
- [10] 三井祥幹, 高宗典玄, 北村大地, 猿渡洋, 高橋祐, and 近藤多伸, “空間事前情報を用いた独立低ランク行列分析,” in 第 32 回信号処理シンポジウム, 2017, pp. 360–365.
- [11] 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.
- [12] 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. LREC*, 2000, pp. 965–968.
- [13] S. Araki, F. Nesta, E. Vincent, Z. Koldovský, G. Nolte, A. Ziehe, and A. Benichoux, “The 2011 signal separation evaluation campaign (SiSEC2011): audio source separation,” in *Proc. LVA/ICA*, 2012, pp. 414–422.