

I ダイバージェンスを用いた独立低ランク行列分析

ILRMA using I divergence criterion

最上伸一[†] 三井祥幹[†] 高宗典玄[†] 北村大地[†] 猿渡洋[†] 高橋祐[‡] 近藤多伸[‡]
[†]東京大学大学院情報理工学系研究科 [‡]ヤマハ株式会社

Shinichi MOGAMI[†] Yoshiki MITSUI[†] Norihiro TAKAMUNE[†] Daichi KITAMURA[†]
Hiroshi SARUWATARI[†] Yu TAKAHASHI[‡] Kazunobu KONDO[‡]
[†]The University of Tokyo [‡]YAMAHA Corporation

アブストラクト 本稿では、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) において、I ダイバージェンス規範を用いたアルゴリズムを新たに提案する。提案法では、分離信号の時間および周波数成分が、複素数領域に拡張された独立な Poisson 分布に従うと仮定し、得られる尤度関数を近似的に最大化する高速かつ頑健な手法を導出する。さらに、音楽信号に対する音源分離実験を行うことで、提案手法の有用性を示す。

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から、混合系に関する事前情報を用いずに混合前の音源信号を推定する技術である。特に、観測信号を短時間フーリエ変換 (short-time Fourier transform: STFT) して得られるスペクトログラムに対して独立成分分析 (independent component analysis: ICA) [1] を適用する周波数領域 ICA (frequency-domain ICA: FDICA) [2], [3] や、その多変量拡張の独立ベクトル分析 (independent vector analysis: IVA) [4], [5] がよく用いられる。

一方で、非負値行列因子分解 (nonnegative matrix factorization: NMF) [6] を用いたブラインドおよび教師あり音源分離手法も広く研究されている [7]–[9]。NMF は、音響信号のパワーまたは振幅スペクトログラムを非負行列とみなし、限られた数の非負基底を含む基底行列とそれらの非負係数を含むアクティベーション行列の行列積に低ランク近似する。このとき推定される基底は、音響信号中に頻出するスペクトルパタンに対応し、それらの係数は時間的な強度変化 (アクティベーション) を示す。この基底行列およびアクティベーション行列は、観測行列とモデル (基底行列とアクティベーション行列の行列積) の間のダイバージェンスを最小化することで推定される。また、特定のダイバージェンスの最小化問題は、観測行列の生成モデルを仮定した最尤推定問題と解釈することが

可能である。たとえば、板倉斎藤ダイバージェンスに基づく NMF (IS-NMF) [10] では、各時間周波数において独立に定義された原点对称の複素正規分布に従って複素スペクトログラムが生成されると仮定している。様々なダイバージェンスを用いた NMF が提案されているが、その中でも振幅スペクトログラムを観測行列とした場合の一般化 Kullback–Leibler (KL) ダイバージェンス (I ダイバージェンス) に基づく NMF (KL-NMF) が、音源分離タスクに対して頑健かつ高性能な結果をもたらすことが経験的に知られている [9], [11]。Kameoka は、Poisson 分布を複素領域に拡張した複素 Poisson 分布を新たに提案し、IS-NMF と同様に、KL-NMF の複素生成モデルとしての解釈を与え、最尤推定問題として再定式化した [12]。このような生成モデルの解釈は、NMF を ICA のような他の統計的信号処理に応用する場合に重要となる。

NMF は本来単一チャネルの信号に適用するものであったが、これを多チャネル音響信号の分離へと拡張した多チャネル NMF (multichannel NMF: MNMF) [13]–[15] が考案されている。MNMF は頻出スペクトルパタンの他に、各音源の空間的な伝達特性 (混合系) に対応する空間相関行列 [16] を推定して音源分離を達成する。最適化規範の観点では、これまでに二乗 Euclid 距離および板倉斎藤ダイバージェンスに基づく MNMF [15] や多次元複素 Student's t 分布生成モデルに基づく MNMF [17] が提案されているが、一般化 KL ダイバージェンス規範の MNMF は未だ提案されていない。MNMF は混合系を推定するため劣決定条件 (チャネル数 < 音源数) 音源分離にも適用できるが、分離精度が変数の初期値に強く依存することや、膨大な計算時間を必要とする欠点がある。

近年、MNMF よりも頑健かつ高速な BSS として、各音源の時間周波数構造の低ランク性と音源間の独立性の両性質を仮定した手法の独立低ランク行列分析 (independent low-rank matrix analysis: ILRMA) [18], [19] が提案されている。ILRMA は IVA のベクトルモデルを低ランク行

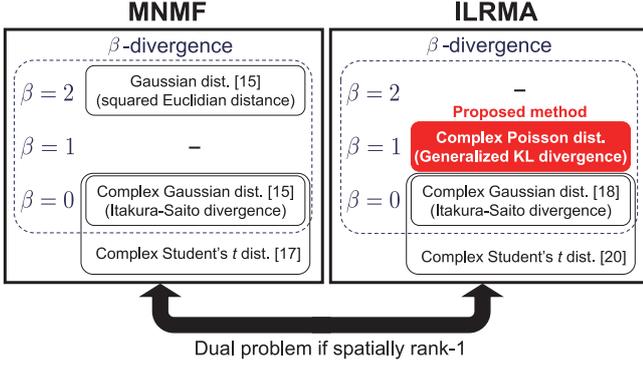


図 1: Overview of MNMF and ILRMA.

列モデルに拡張した手法と解釈できる。また、MNMF における空間相関行列のランクが 1 となる伝達系の近似制約を導入したモデルと等価である。ただし、MNMF は混合系を推定するのに対し、ILRMA は IVA と同様に周波数毎の分離行列を推定する。この違いに起因して、ILRMA は MNMF と比較して最適化が高速であり、変数の初期値に頑健という利点がある。さらに ILRMA の生成モデルを複素正規分布から複素 Student's t 分布に拡張した t -ILRMA [20] も提案されており、生成モデルを変えることによって分離精度が変化することが示唆されている。

以上に鑑み、本稿では ILRMA のさらなる分離精度向上および頑健性向上を目的として、一般化 KL ダイバージェンスを ILRMA の距離規範に取り入れた KL-ILRMA を提案する。図 1 は各生成モデルに基づく MNMF および ILRMA の俯瞰図である。本稿では、複素 Poisson 分布を新たに生成モデルとして用いることで、これまで提案されていなかった一般化 KL ダイバージェンスに基づく多チャンネル音源分離を新たに確立する。特に、一般化 KL ダイバージェンスの凸性より、本提案手法は従来の ILRMA における問題であった初期値依存性を大きく改善できることを示す。

2 従来手法

2.1 定式化

音源信号、観測信号およ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、 $i = 1, \dots, I$ および $j = 1, \dots, J$ はそれぞれ周波数ビンおよび時間フレームのインデックス、 $n = 1, \dots, N$ および $m = 1, \dots, M$ はそれぞれ音源、チャンネルのインデックスである。 $*^\top$ は転置を表す。

FDICA, IVA および ILRMA では、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、 $\mathbf{A}_i = (\mathbf{a}_{i1}, \dots, \mathbf{a}_{iN}) \in \mathbb{C}^{M \times N}$ は周波数ごとの混合行列であり、 \mathbf{a}_{in} は n 番目の音源に対するステアリングベクトルである。観測信号と音源との関係が式 (4) で与えられるという仮定は、MNMF において、各音源の空間相関行列のランクが 1 であると仮定することに相当する [18]。さらに $N = M$ かつ \mathbf{A}_i が正則の場合、分離信号は

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

として推定できる。ここで、 $\mathbf{W}_i = (\mathbf{w}_{i1} \dots \mathbf{w}_{iN})^\mathbf{H} = \mathbf{A}_i^{-1}$ である。 \mathbf{w}_{in} は n 番目の音源の分離フィルタを表し、 $*^\mathbf{H}$ はエルミート転置である。FDICA, IVA および ILRMA では分離信号間の独立性を仮定することで、 \mathbf{x}_{ij} のみを用いて \mathbf{W}_i および \mathbf{y}_{ij} の推定を可能にしている。

2.2 板倉斎藤ダイバージェンスに基づく ILRMA

板倉斎藤ダイバージェンス規範でモデル化された従来の ILRMA (IS-ILRMA) では、各音源の複素スペクトログラムが、各時間周波数で分散が変動する独立な複素正規分布に従って生成されると仮定する。従って、音源 n のスペクトログラムの生成モデルは

$$\prod_{i,j} p(y_{ijn}) = \prod_{i,j} \frac{1}{\pi r_{ijn}} \exp\left(-\frac{|y_{ijn}|^2}{r_{ijn}}\right) \quad (6)$$

と与えられる。また、上のモデルにおける分散 r_{ijn} は

$$r_{ijn} = \sum_k t_{ikn} v_{kjn} \quad (7)$$

と与えられる非負値とする。ここで、 $k = 1, \dots, K$ は NMF における基底のインデックスであり、 t_{ikn} および v_{kjn} は、それぞれ基底行列 $\mathbf{T}_n \in \mathbb{R}_+^{I \times K}$ およ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とする。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ことで、各音源スペクトログラムを低ランク近似できる。また、分散 r_{ijn} が時間フレーム j と周波数ビン i に依存することから、式 (6) を時間フレーム j に関して周辺化すると、一般に正規分布よりも尖度の大きい分布となる。式 (6),(7) および音源間の独立性を仮定すると、IS-ILRMA のコスト関数である全チャンネルの観測信号の負対数尤度は

$$\mathcal{L} = \text{const.} - 2J \sum_i \log |\det \mathbf{W}_i| + \sum_{i,j,n} \left[\log \sum_k t_{ikn} v_{kjn} + \frac{|\mathbf{w}_{in}^\mathbf{H} \mathbf{x}_{ij}|^2}{\sum_k t_{ikn} v_{kjn}} \right] \quad (8)$$

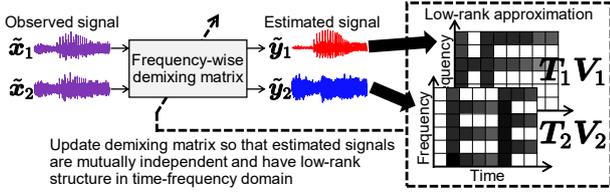


図 2: Principle of source separation in ILRMA, where $\tilde{\mathbf{x}}_m$ and $\tilde{\mathbf{y}}_n$ are time-domain signals of x_{ijn} and y_{ijn} , respectively.

となる。変数 t_{ikn}, v_{kjn} に注目すると、式 (8) の最小化は、各音源における IS-NMF の最適化、すなわち分離信号のパワースペクトログラム $|\mathbf{w}_{in}^H \mathbf{x}_{ij}|^2$ とモデル $\sum_k t_{ikn} v_{kjn}$ との板倉斎藤ダイバージェンスを最小化することと等価である。ILRMA では分離行列 \mathbf{W}_i と音源モデル $\mathbf{T}_n, \mathbf{V}_n$ を交互に反復更新することにより、コスト関数 (8) を最小化する。IS-ILRMA では、分離行列の更新時に iterative projection (IP) [5] と呼ばれる手法を用いることで、高速かつ頑健な反復更新を可能としている。また、音源モデルの更新には対応するダイバージェンス規範の NMF の乗算更新式をそのまま用いることができる。

ILRMA における音源分離の原理を図 2 に示す。混合前の音源信号のパワースペクトログラムが低ランクであると仮定すると、観測信号のパワースペクトログラムのランクは混合により増加する。ILRMA では、分離信号のパワースペクトログラムが十分低ランクな行列 $\mathbf{T}_n \mathbf{V}_n$ で表されるように、分離行列 \mathbf{W}_i を推定する。

3 提案手法

3.1 動機

NMF を用いた音源分離では、前述の通り振幅スペクトログラムを観測行列とした場合の KL-NMF が最も良い分離精度をもたらすことが経験的に知られている。なおかつ、板倉斎藤ダイバージェンスが非凸であるのに対して一般化 KL ダイバージェンスが凸であることから、KL-NMF は IS-NMF に比べて分離結果のばらつきを抑える効果が期待できる。ILRMA の音源モデルに一般化 KL ダイバージェンスに対応する複素生成モデルを導入できれば、このような KL-NMF の優位性を持つ独立性基準の BSS が実現できると考えられる。

3.2 複素 Poisson 分布に基づく ILRMA

本稿では、ILRMA における個々の音源のモデルとして複素 Poisson 分布を仮定することにより、一般化 KL ダイバージェンス規範の KL-ILRMA を提案する。複素 Poisson 分布は、 $\mathcal{D} = \{z \in \mathbb{C} \mid |z| \in \mathbb{N}\}$ を台とする複素

数 z の確率密度関数

$$p(z) = \frac{p |z|^{p-2} \lambda |z|^p}{2\pi (|z|^p)!} e^{-\lambda} \quad (9)$$

で与えられる確率分布である [12]。先行研究 [9], [11] では、振幅スペクトログラムを用いた KL-NMF が音源分離に適していることから、本稿でも $p = 1$ のみを考える。従って、KL-ILRMA の生成モデルは以下の式で与えられる。

$$\prod_{i,j,n} p(y_{ijn}) = \prod_{i,j,n} \frac{|y_{ijn}|^{-1} \lambda_{ijn} |y_{ijn}|}{2\pi |y_{ijn}|!} e^{-\lambda_{ijn}} \quad (10)$$

$$\lambda_{ijn} = \sum_k t_{ikn} v_{kjn} \quad (11)$$

IS-ILRMA と同様に、上記の確率密度関数の負対数尤度を考えると

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

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

$$\log(x!) \approx x \log x - x \quad (13)$$

を導入することにより、負対数尤度の式 (12) 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。

$$\begin{aligned} \mathcal{J} = \text{const.} &- 2J \sum_i \log |\det \mathbf{W}_i| \\ &+ \sum_{i,j,n} \left[|\mathbf{w}_{in}^H \mathbf{x}_{ij}| \log |\mathbf{w}_{in}^H \mathbf{x}_{ij}| - |\mathbf{w}_{in}^H \mathbf{x}_{ij}| + \log |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \right. \\ &\quad \left. - |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \log \sum_k t_{ikn} v_{kjn} + \sum_k t_{ikn} v_{kjn} \right] \end{aligned} \quad (14)$$

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

$$\begin{aligned} \mathcal{J} = \text{const.} &- 2J \sum_i \log |\det \mathbf{W}_i| \\ &+ \sum_{i,j,n} [D_{\text{KL}}(|\mathbf{w}_{in}^H \mathbf{x}_{ij}| \mid \sum_k t_{ikn} v_{kjn}) + \log |\mathbf{w}_{in}^H \mathbf{x}_{ij}|] \end{aligned} \quad (15)$$

とも書けることから、一般化 KL ダイバージェンス規範で音源モデルの低ランク性を高めるコストであると解釈できる。

3.3 分離行列の更新式

式 (14) を最小化する分離行列 \mathbf{W}_i を直接求めることは困難であるから、補助関数法 [21] 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 (14) の $\log|\mathbf{w}_{in}^H \mathbf{x}_{ij}|$ に関して次の接線不等式

$$\log|\mathbf{w}_{in}^H \mathbf{x}_{ij}| \leq \frac{1}{\alpha_{ijn}}(|\mathbf{w}_{in}^H \mathbf{x}_{ij}| - \alpha_{ijn}) + \log \alpha_{ijn} \quad (16)$$

を適用すると、次の補助関数 \mathcal{J}_1 が得られる。

$$\mathcal{J} \leq \mathcal{J}_1 = \text{const.} - 2J \sum_i \log |\det \mathbf{W}_i| + \sum_{i,j,n} \left[\frac{1}{\alpha_{ijn}} |\mathbf{w}_{in}^H \mathbf{x}_{ij}|^2 + d_{ijn} |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \right] \quad (17)$$

$$d_{ijn} = \frac{1}{\alpha_{ijn}} + \log \alpha_{ijn} - \log \sum_k t_{ikn} v_{kjn} - 2 \quad (18)$$

ここで $\alpha_{ijn} > 0$ は補助変数であり、式 (16) は $\alpha_{ijn} = |\mathbf{w}_{in}^H \mathbf{x}_{ij}|$ のときに等号成立する。さらに、 $d_{ijn} |\mathbf{w}_{in}^H \mathbf{x}_{ij}|$ に対しては不等式

$$d_{ijn} |\mathbf{w}_{in}^H \mathbf{x}_{ij}| \leq \begin{cases} \frac{d_{ijn}}{2\beta_{ijn}} |\mathbf{w}_{in}^H \mathbf{x}_{ij}|^2 + \frac{1}{2} d_{ijn} \beta_{ijn} & (d_{ijn} \geq 0) \\ \frac{1}{2} d_{ijn} (\overline{\omega}_{ijn} \mathbf{w}_{in}^H \mathbf{x}_{ij} + \omega_{ijn} \overline{\mathbf{w}_{in}^H \mathbf{x}_{ij}}) & (d_{ijn} < 0) \end{cases} \quad (19)$$

を考えることができる。ここで、 $\overline{\cdot}$ は \cdot 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 (19) を式 (17) に適用すると、次の補助関数 \mathcal{J}_2 が得られる。

$$\mathcal{J}_1 \leq \mathcal{J}_2 = \text{const.} - 2J \sum_i \log |\det \mathbf{W}_i| + J \sum_{i,n} [\mathbf{w}_{in}^H \mathbf{U}_{in} \mathbf{w}_{in} + \mathbf{w}_{in}^H \mathbf{r}_{in} + \mathbf{r}_{in}^H \mathbf{w}_{in}] \quad (20)$$

$$\mathbf{U}_{in} = \frac{1}{J} \sum_j \left(\frac{1}{\alpha_{ijn}} + \frac{\max(0, d_{ijn})}{2\beta_{ijn}} \right) \mathbf{x}_{ij} \mathbf{x}_{ij}^H \quad (21)$$

$$\mathbf{r}_{in} = \frac{1}{J} \sum_j \frac{\overline{\omega}_{ijn}}{2} \min(d_{ijn}, 0) \mathbf{x}_{ij} \quad (22)$$

式 (20) は \mathbf{w}_{in} の一次項を含むため、IS-ILRMA で用いた IP を適用することができない (IP は補助関数が $-\log |\det \mathbf{W}_i|$ と \mathbf{w}_{in} の二次形式の和で構成される場合のみ適用できる [5])。しかしながら、一次項を含む補助関数の最小化には本稿の著者である三井らの提案したブ

ロック座標降下法 [22] が適用できる。この手法では、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 (20) を以下の形式に変形する。

$$\mathcal{J}_2 = \text{const.} - J \sum_i \log |\mathbf{b}_{in}^H \mathbf{w}_{in}|^2 + J \sum_{i,n} [\mathbf{w}_{in}^H \mathbf{U}_{in} \mathbf{w}_{in} + \mathbf{w}_{in}^H \mathbf{r}_{in} + \mathbf{r}_{in}^H \mathbf{w}_{in}] \quad (23)$$

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。この \mathbf{b}_{in} と \mathbf{U}_{in} および \mathbf{r}_{in} はすべて \mathbf{w}_{in} と独立な係数であることに着目すると、ほかの行ベクトルを固定することによって式 (23) は \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。ここでは、式 (23) の \mathbf{w}_{in} に関する偏微分を解析的に解くことができる [22] ため、 \mathbf{w}_{in} に関する補助関数 \mathcal{J}_2 の極小解は次式で与えられる。

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

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。

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

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

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

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

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

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

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

3.4 音源モデルの更新式

式 (14) 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、これらに関するものは $D_{\text{KL}}(|\mathbf{w}_{in}^H \mathbf{x}_{ij}| | \sum_k t_{ikn} v_{kjn})$

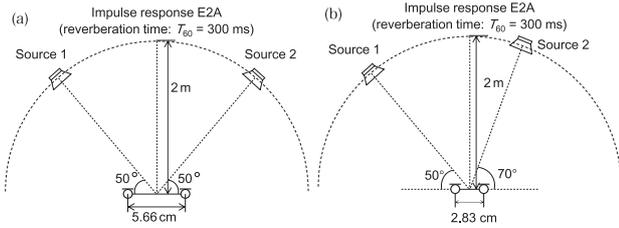


図 3: Recording conditions of impulse responses obtained from RWCP database: (a) IR1 and (b) IR2.

表 1: Music sources obtained from SiSEC2011

Index	Source (1/2)	Impulse response
No.1	acoustic_guit_main/vocals	IR1
No.2	acoustic_guit_main/vocals	IR2
No.3	acoustic_guit_main/piano	IR1
No.4	acoustic_guit_main/piano	IR2

の項のみである。従って、 $\mathbf{T}_n, \mathbf{V}_n$ は KL-NMF 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 [6].

$$t_{ikn} \leftarrow t_{ikn} \cdot \frac{\sum_j \frac{|\mathbf{w}_{in}^H \mathbf{x}_{ij}|}{\sum_{k'} t_{ik'n} v_{k'jn}} v_{k'jn}}{\sum_j v_{k'jn}} \quad (32)$$

$$v_{k'jn} \leftarrow v_{k'jn} \cdot \frac{\sum_i \frac{|\mathbf{w}_{in}^H \mathbf{x}_{ij}|}{\sum_{k'} t_{ik'n} v_{k'jn}} t_{ikn}}{\sum_i t_{ikn}} \quad (33)$$

KL-ILRMA では、分離行列 \mathbf{W}_i の更新式 (25)–(31) と、音源モデル $\mathbf{T}_n, \mathbf{V}_n$ の更新式 (32),(33) を交互に繰り返すことによって、式 (14) で表されるコスト関数 J の最小化を達成する。 \mathbf{W}_i の更新式および $\mathbf{T}_n, \mathbf{V}_n$ の更新式は全て補助関数法を用いて導出されていることから、 J の単調減少性 (非増加性) が保証されている。

4 実験

4.1 実験条件

本稿では、提案手法である KL-ILRMA の音源分離性能を音楽信号に対して確認し、従来の IS-ILRMA との実験的な性能比較を行う。実験では、SiSEC2011 [23] で公開されている音楽信号の berlin-roads を用いる。この音源信号に対し、RWCP データベース [24] から得た E2A インパルス応答 ($T_{60} = 300$ ms) を畳み込むことによって、2チャンネルかつ2音源の混合信号を生成した。本実験にて用いた2種類のインパルス応答の収録環境を図3に示す。用意した計4種類の混合信号について、使用されている音源パートおよび用いたインパルス応答を表1に示す。評価指標には signal-to-distortion ratio (SDR) [25] の改善量を用いた。分離行列 \mathbf{W}_i の初期値はすべて単位行列と

表 2: Experimental conditions

Sampling rate	16000
STFT	256-ms-long Hamming window with 128-ms shift
Number of bases	10
Number of iterations	200

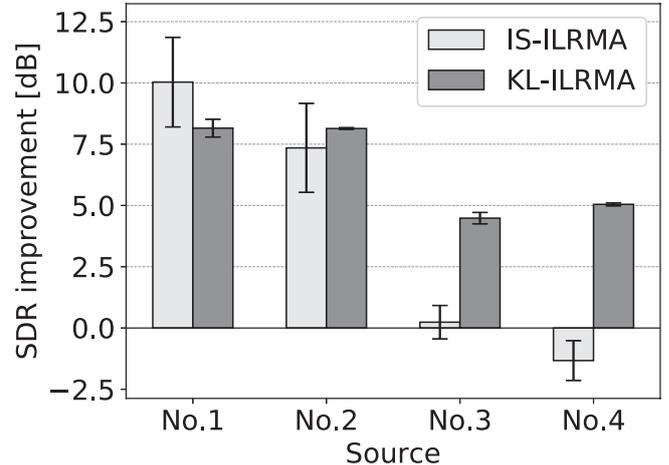


図 4: Average SDR improvements of IS-ILRMA and KL-ILRMA.

し、音源モデル $\mathbf{T}_n, \mathbf{V}_n$ の初期値は一様乱数で与えた。その他の実験条件を表2に示す。

4.2 結果

乱数シードを変えて各手法をそれぞれ20回試行したときの SDR 改善量の平均および標準偏差を図4に示す。全体の傾向として、分離性能は音源に大きく依存することが確認された。No.1 および No.2 に用いた音源は比較的複素正規分布のモデル (IS-ILRMA) がよく合致するのに対し、No.3 および No.4 に用いた音源は比較的複素 Poisson 分布のモデル (KL-ILRMA) がよく合致すると考えられる。また、2種類の ILRMA の結果の標準偏差に注目すると、いずれの音源でも KL-ILRMA は IS-ILRMA に比べて結果のばらつきが小さいことが確認された。これは、板倉斎藤ダイバージェンスが非凸であるのに対して一般化 KL ダイバージェンスが凸であることから、KL-ILRMA の音源モデル $\mathbf{T}_n, \mathbf{V}_n$ の最適化が初期値のばらつきに対して頑健になったことに起因していると考えられる。

5 まとめ

本稿では、複素正規分布を生成モデルとする従来の IS-ILRMA に対して、新たに複素 Poisson 分布を生成モデルとする KL-ILRMA を提案し、その更新式の導出および評価実験の結果を示した。評価実験の結果、KL-ILRMA

が従来の IS-ILRMA と比較して、特に頑健性（初期値依存性）の面において優れていることが示唆された。

謝辞

本研究は、総合科学技術・イノベーション会議による革新的研究開発推進プログラム (ImPACT)、セコム科学技術振興財団、JSPS 科研費 16H01735 および 17H06572 の助成を受けた。

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, no. 1, pp. 21–34, 1998.
- [3] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, “Blind source separation based on a fast-convergence algorithm combining ICA and beamforming,” *IEEE Trans. ASLP*, vol. 14, no. 2, pp. 666–678, 2006.
- [4] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, “Blind source separation exploiting higher-order frequency dependencies,” *IEEE Trans. ASLP*, vol. 15, no. 1, pp. 70–79, 2007.
- [5] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” in *Proc. WASPAA*, 2011, pp. 189–192.
- [6] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Proc. NIPS*, 2000, pp. 556–562.
- [7] 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.
- [8] 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,” in *Proc. ICASSP*, 2012, pp. 5365–5368.
- [9] 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.
- [10] 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.
- [11] D. FitzGerald, M. Cranitch, and E. Coyle, “On the use of the beta divergence for musical source separation,” in *Proc. ISSC*, 2009.
- [12] 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.
- [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] H. Kameoka, T. Yoshioka, M. Hamamura, J. L. Roux, and K. Kashino, “Statistical model of speech signals based on composite autoregressive system with application to blind source separation,” in *Proc. LVA/ICA*, 2010, pp. 245–253.
- [15] 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.
- [16] 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.
- [17] K. Kitamura, Y. Bando, K. Itoyama, and K. Yoshii, “Student’s t multichannel nonnegative matrix factorization for blind source separation,” in *Proc. IWAENC*, 2016.
- [18] 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.
- [19] Y. Mitsui, D. Kitamura, S. Takamichi, N. Ono, and H. Saruwatari, “Blind source separation based on independent low-rank matrix analysis with sparse regularization for time-series activity,” in *Proc. ICASSP*, 2017, pp. 21–25.
- [20] 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 (in press).
- [21] D. R. Hunter and K. Lange, “Quantile regression via an MM algorithm,” *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 60–77, 2000.
- [22] Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo, “ILRMA incorporating with a priori spatial information,” in *Proc. IEICE 32nd SIP Symposium*, no. B8-2, 2017 (in Japanese, in press).
- [23] 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.
- [24] 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.
- [25] 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.