LETTER

Independent Low-Rank Matrix Analysis Based on Generalized Kullback–Leibler Divergence*

Shinichi MOGAMI[†a)], Yoshiki MITSUI[†], Norihiro TAKAMUNE[†], Daichi KITAMURA[††], *Nonmembers*, Hiroshi SARUWATARI[†b)], *Member*, Yu TAKAHASHI[†††], *Nonmember*, Kazunobu KONDO[†††], *Member*, Hiroaki NAKAJIMA[†††], *Nonmember, and* Hirokazu KAMEOKA[††††], *Member*

SUMMARY In this letter, we propose a new blind source separation method, independent low-rank matrix analysis based on generalized Kullback–Leibler divergence. This method assumes a time-frequency-varying complex Poisson distribution as the source generative model, which yields convex optimization in the spectrogram estimation. The experimental evaluation confirms the proposed method's efficacy.

key words: *blind source separation, nonnegative matrix factorization, Poisson distribution, Kullback–Leibler divergence*

1. Introduction

Blind source separation (BSS) [1]–[3] is a technique for extracting specific sources from an observed multichannel mixture signal without knowing a priori information about the mixing system. Let N and M be the numbers of sources and channels, respectively. The short-time Fourier transforms (STFTs) of the multichannel source, observed, and estimated signals are defined as $\boldsymbol{s}_{ij} = (s_{ij1}, \ldots, s_{ijN})^\top \in \mathbb{C}^N$, $\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijM})^\top \in \mathbb{C}^M$, and $\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijN})^\top \in \mathbb{C}^N$, where $i = 1, \ldots, I$; $j = 1, \ldots, J$; $n = 1, \ldots, N$; and $m = 1 \ldots, M$ are the integral indices of the frequency bins, time frames, sources, and channels, respectively, and $^\top$ denotes the transpose. We assume the mixing system $\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij}$, where $\boldsymbol{A}_i = (\boldsymbol{a}_{i1}, \ldots, \boldsymbol{a}_{iN}) \in \mathbb{C}^{M \times N}$ is a frequency-wise mixing matrix and $\boldsymbol{a}_{in}$ is the steering vector for the nth source. When $M = N$ and $\boldsymbol{A}_i$ is not a singular matrix, the estimated signal $\boldsymbol{y}_{ij}$ can be expressed as $\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij}$, where $\boldsymbol{W}_i = \boldsymbol{A}_i^{-1} = (\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{iN})^\mathsf{H}$ is the demixing matrix, $\boldsymbol{w}_{in}$ is the demixing filter for the nth source, and $^\mathsf{H}$ denotes the Hermitian transpose. Multichannel nonnegative matrix factorization (MNMF) [4], [5] is a BSS method that simultaneously estimates both the low-rank time-frequency structure and the spatial covariance matrix for each source, indirectly identifying the mixing system. Recently, *independent low-rank matrix analysis (ILRMA)* [6], which is a unification of direct estimation of the *demixing* matrix [7] and simple nonnegative matrix factorization (NMF) [8], [9], was proposed as a state-of-the-art BSS method. In terms of optimization, ILRMA is faster and more stable than MNMF.

Conventional MNMF and ILRMA assume the time-frequency-varying complex Gaussian distribution in the generative model of each source spectrogram, which corresponds to NMF based on Itakura–Saito (IS) divergence (IS-NMF). We refer to the conventional ILRMA based on IS divergence as IS-ILRMA. Recently, the source generative models assumed in ILRMA and MNMF were generalized to the time-frequency-varying complex Student's t-distribution (t-MNMF [10] and t-ILRMA [11]). These conventional ILRMAs always include a non-convex optimization in the spectrogram modeling that is sensitive to the initialization and diversifies the separation quality. In this letter, we propose ILRMA based on generalized Kullback–Leibler (KL) divergence (*KL-ILRMA*) by using a *time-frequency-varying complex Poisson distribution* as the source generative model. In KL-ILRMA, the estimation of a low-rank spectrogram model results in KL-divergence-based NMF (KL-NMF) [8], which is a *convex* optimization in terms of each decomposed matrix variable. To our knowledge, the proposed method is the world's first attempt to realize convex-optimization-based ILRMA w.r.t. source modeling. Owing to this property, we can perform separation robust against the initialization of the source model.

2. Proposed Method

2.1 ILRMA Based on Time-Frequency-Varying Complex Poisson Distribution

KL-ILRMA approximates the source spectrogram with a nonnegative low-rank matrix by minimizing their generalized KL divergence. Here we assume that the source model follows the time-frequency-varying complex Poisson distribution [12], which is the extension of the real-valued Poisson distribution to complex values. The probability density function of the complex Poisson distribution is defined as

$$p(z) = \frac{|z|^{-1} \lambda^{|z|}}{2\pi(|z|)!} e^{-\lambda}, \qquad (1)$$

where z is in the set $\mathcal{D} = \{z \in \mathbb{C} \mid |z| \in \mathbb{N}\}$ and λ is a shape

Manuscript received August 2, 2018.
Manuscript revised October 10, 2018.

[†] The authors are with the University of Tokyo, Tokyo, 113-8656 Japan.

[††] The author is with National Institute of Technology, Kagawa College, Takamatsu-shi, 761-8058 Japan.

[†††] The authors are with Yamaha Corporation, Hamamatsu-shi, 430-8650 Japan.

[††††] The author is with Nippon Telegraph and Telephone Corporation, Atsugi-shi, 243-0198 Japan.

[*] This work was supported by SECOM Science and Technology Foundation and JSPS KAKENHI Grant Number JP17H06572.

a) E-mail: shinichi_mogami@ipc.i.u-tokyo.ac.jp
b) E-mail: hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

DOI: 10.1587/transfun.E102.A.458

parameter characterizing its distribution.

In KL-ILRMA, the following time-frequency-varying complex Poisson source generative model is assumed:

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

$$\lambda_{ijn} = \sum_k t_{ikn} v_{kjn}, \tag{3}$$

where λ_{ijn} is the shape parameter as well as the *mean* of the amplitude of the complex Poisson distribution and is interpreted as the ith row and jth column value of the source spectrogram model as in Eq. (3). The variables t_{ikn} and v_{kjn} are the elements of the basis matrix $\boldsymbol{T}_n \in \mathbb{R}_{\geq 0}^{I \times K}$ and the activation matrix $\boldsymbol{V}_n \in \mathbb{R}_{\geq 0}^{K \times J}$, respectively, where $\mathbb{R}_{\geq 0}$ denotes the set of nonnegative real numbers. $k = 1, \ldots, K$ is the index of the bases, and the number of bases K is usually much less than I and J. Note that the validity of this generative model is discussed in Appendix.

Since $y_{ijn} = \boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}$ in Eq. (2), the negative log-likelihood of $\boldsymbol{x}_{ij} = \boldsymbol{W}_i^{-1} \boldsymbol{y}_{ij}$ is given by

$$\begin{aligned}\mathcal{L} = \sum_{i,j,n} \Bigg[& \log(|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|!) + \log|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| \\ & - |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| \log \sum_k t_{ikn} v_{kjn} + \sum_k t_{ikn} v_{kjn} \Bigg] \\ & - 2J \sum_i \log|\det \boldsymbol{W}_i| + \text{const.} \end{aligned} \tag{4}$$

To express the term $\log(|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|!)$ via elementary functions and extend the domain of the definition to an analog domain, we apply Stirling's approximation (its practical validity will be assessed in Sect. 3.1)

$$\log(|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|!) \approx |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| \log|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| - |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|. \tag{5}$$

Hence, we obtain the cost function to be minimized as

$$\begin{aligned}\mathcal{J} = \sum_{i,j,n} \Bigg[& |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| \log|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| - |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| + \log|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| \\ & - |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| \log \sum_k t_{ikn} v_{kjn} + \sum_k t_{ikn} v_{kjn} \Bigg] \\ & - 2J \sum_i \log|\det \boldsymbol{W}_i| + \text{const.} \end{aligned} \tag{6}$$

$$\begin{aligned} = & \sum_{i,j,n} [D_{\mathrm{KL}}(|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| \mid \textstyle\sum_k t_{ikn} v_{kjn}) + \log|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|] \\ & - 2J \sum_i \log|\det \boldsymbol{W}_i| + \text{const.}, \end{aligned} \tag{7}$$

where $D_{\mathrm{KL}}(y \mid x) = y \log y - y \log x - y + x$ is the generalized KL divergence. Therefore, the minimization of the cost function $\mathcal{J}$ simultaneously achieves high independence between the sources and the low-rank modeling of each source spectrogram based on the generalized KL divergence.

2.2 Update Rule for Source Model

In the cost function Eq. (7), the only term related to $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$ is $D_{\mathrm{KL}}(|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| \mid \sum_k t_{ikn} v_{kjn})$. Therefore, $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$ can be optimized by minimizing the divergence via the KL-NMF update rules [8]

$$t_{ikn} \leftarrow t_{ikn} \left(\sum_j \frac{|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| v_{kjn}}{\sum_{k'} t_{ik'n} v_{k'jn}} \right) \left(\sum_j v_{kjn} \right)^{-1}, \tag{8}$$

$$v_{kjn} \leftarrow v_{kjn} \left(\sum_i \frac{|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| t_{ikn}}{\sum_{k'} t_{ik'n} v_{k'jn}} \right) \left(\sum_i t_{ikn} \right)^{-1}. \tag{9}$$

Since minimization of the generalized KL divergence is a convex problem w.r.t. either $\boldsymbol{T}_n$ or $\boldsymbol{V}_n$ [9], KL-ILRMA is expected to more stably estimate the source spectrograms than conventional ILRMAs, which do not involve convex problems w.r.t. either $\boldsymbol{T}_n$ or $\boldsymbol{V}_n$.

2.3 Update Rule for Demixing Matrix

In conventional IS-ILRMA, the demixing matrix $\boldsymbol{W}_i$ can be updated by applying iterative projection (IP) [3], which is a fast and stable optimization algorithm that can be applied to the sum of $|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2$ and $-\log|\det \boldsymbol{W}_i|$. In KL-ILRMA, however, IP cannot be applied, i.e., the cost function Eq. (6) does not satisfy the necessary condition for the use of IP. Instead, we apply a majorization-minimization (MM) algorithm [13] to derive the update rule of $\boldsymbol{w}_{in}$. First, we apply the tangent line inequality

$$\log|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| \leq \frac{1}{\alpha_{ijn}} (|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| - \alpha_{ijn}) + \log \alpha_{ijn} \tag{10}$$

to the term $\log|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|$ in Eq. (6), where $\alpha_{ijn} > 0$ is an auxiliary variable. Thus, the majorization function can be designed as

$$\begin{aligned}\mathcal{J} \leq \mathcal{J}_1 = & \sum_{i,j,n} \left[\frac{1}{\alpha_{ijn}} |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2 + d_{ijn} |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}| \right] \\ & - 2J \sum_i \log|\det \boldsymbol{W}_i| + \text{const.}, \end{aligned} \tag{11}$$

$$d_{ijn} = \frac{1}{\alpha_{ijn}} + \log(\alpha_{ijn} / \textstyle\sum_k t_{ikn} v_{kjn}) - 2, \tag{12}$$

where $\mathcal{J}$ and $\mathcal{J}_1$ become equal only when $\alpha_{ijn} = |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|$.

Second, we design the further majorization function of $d_{ijn} |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|$ to make it differentiable w.r.t. $\boldsymbol{w}_{in}$. As shown in Fig. 1, we can compose a majorization function of $d\,|y|$ that coincides with $d\,|y|$ at an arbitrary point y_0 (black dotted line in Fig. 1) by branching into "paraboloid type" ($d \geq 0$) and "plane type" ($d < 0$) cases. Thus, the majorization functions of $d_{ijn} |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|$ are obtained as

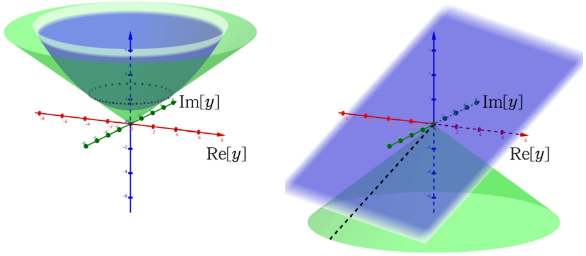


Fig. 1 Shapes of majorization function (blue) of original function $d\,|y|$ (green) with contact point y_0 (black dotted line). When $d > 0$, $d\,|y|$ can be majorized by paraboloid of revolution (left). When $d < 0$, $d\,|y|$ can be majorized by plane (right).

$$d_{ijn}|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}| \leq \begin{cases} \dfrac{d_{ijn}}{2\beta_{ijn}}|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2 + \dfrac{1}{2}d_{ijn}\beta_{ijn} \\ \hfill (d_{ijn} \geq 0) \\ \dfrac{1}{2}d_{ijn}(\overline{\omega_{ijn}}\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij} + \omega_{ijn}\overline{\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}}) \\ \hfill (d_{ijn} < 0) \end{cases}, \tag{13}$$

where $\overline{*}$ is the complex conjugate of $*$, $\beta_{ijn} > 0$ is a real auxiliary variable, and ω_{ijn} is a complex auxiliary variable that satisfies $\left|\omega_{ijn}\right| = 1$. The equality of Eq. (13) holds only when $\beta_{ijn} = |\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|\,(d_{ijn} \geq 0)$ and $\omega_{ijn} = \boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}/|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|\,(d_{ijn} < 0)$. Applying Eq. (13) to Eq. (11), we obtain the majorization function $\mathcal{J}_2$ as follows:

$$\begin{aligned} \mathcal{J}_1 \leq \mathcal{J}_2 = J\sum_{i,n}\left[\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{U}_{in}\boldsymbol{w}_{in} + \boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{r}_{in} + \boldsymbol{r}_{in}^{\mathsf{H}}\boldsymbol{w}_{in}\right] \\ - 2J\sum_{i}\log|\det \boldsymbol{W}_i| + \text{const.}, \end{aligned} \tag{14}$$

$$\boldsymbol{U}_{in} = \frac{1}{J}\sum_j\left(\frac{1}{\alpha_{ijn}} + \frac{\max(0, d_{ijn})}{2\beta_{ijn}}\right)\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}, \tag{15}$$

$$\boldsymbol{r}_{in} = \frac{1}{J}\sum_j\frac{\overline{\omega_{ijn}}}{2}\min(d_{ijn}, 0)\boldsymbol{x}_{ij}. \tag{16}$$

Since Eq. (14) contains a linear term in $\boldsymbol{w}_{in}$, we still cannot apply IP to Eq. (14). Instead, we apply another type of vectorwise coordinate descent to minimize functions such as Eq. (14). In this algorithm, we focus on $\boldsymbol{w}_{in}$, namely, the Hermitian transpose of the particular row vector of $\boldsymbol{W}_i$. Equation (14) can be transformed as follows by cofactor expansion:

$$\begin{aligned} \mathcal{J}_2 = J\sum_{i,n}\left[\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{U}_{in}\boldsymbol{w}_{in} + \boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{r}_{in} + \boldsymbol{r}_{in}^{\mathsf{H}}\boldsymbol{w}_{in}\right] \\ - J\sum_{i}\log|\boldsymbol{b}_{in}^{\mathsf{H}}\boldsymbol{w}_{in}|^2 + \text{const.}, \end{aligned} \tag{17}$$

where $\boldsymbol{b}_{in}$ is the column vector of $\boldsymbol{B}_i = (\boldsymbol{b}_{i1}, \ldots, \boldsymbol{b}_{iN})$, which is the adjugate matrix of $\boldsymbol{W}_i$. $\boldsymbol{b}_{in}$ can also be written as $\boldsymbol{b}_{in} = (\det \boldsymbol{W}_i)\boldsymbol{W}_i^{-1}\boldsymbol{e}_n$, where $\boldsymbol{e}_n$ is an N-dimensional vector whose nth element is one and whose other elements are zero. Since $\boldsymbol{b}_{in}$ only depends on $\boldsymbol{w}_{in'}$ $(n' \neq n)$ and is independent of $\boldsymbol{w}_{in}$, Eq. (17) can be regarded as a function of $\boldsymbol{w}_{in}$ by fixing the other row vectors of $\boldsymbol{W}_i$. The partial derivative of Eq. (17) w.r.t. $\boldsymbol{w}_{in}^{\mathsf{H}}$ is

$$\frac{\partial\mathcal{J}_2}{\partial\boldsymbol{w}_{in}^{\mathsf{H}}} = \boldsymbol{U}_{in}\boldsymbol{w}_{in} + \boldsymbol{r}_{in} - \frac{\boldsymbol{b}_{in}}{\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{b}_{in}}. \tag{18}$$

From $\partial\mathcal{J}_2/\partial\boldsymbol{w}_{in}^{\mathsf{H}} = \boldsymbol{0}$, the stationary point is given in [14] as

$$\boldsymbol{w}_{in} = \begin{cases} \boldsymbol{U}_{in}^{-1}\left[\dfrac{1}{\sqrt{u_{bb}}}\boldsymbol{b}_{in} - \boldsymbol{r}_{in}\right] & (\text{if } u_{br} = 0) \\ \boldsymbol{U}_{in}^{-1}\left[\dfrac{u_{br}}{2u_{bb}}\left(1 - \sqrt{1 + \dfrac{4u_{bb}}{|u_{br}|^2}}\right)\boldsymbol{b}_{in} - \boldsymbol{r}_{in}\right] & \\ & (\text{otherwise}) \end{cases}, \tag{19}$$

where $u_{bb} = \boldsymbol{b}_{in}^{\mathsf{H}}\boldsymbol{U}_{in}^{-1}\boldsymbol{b}_{in}$ and $u_{br} = \boldsymbol{b}_{in}^{\mathsf{H}}\boldsymbol{U}_{in}^{-1}\boldsymbol{r}_{in}$. Substituting the equality condition of the auxiliary variables for Eq. (19), the update rule of $\boldsymbol{w}_{in}$ can be obtained as follows:

$$d_{ijn} \leftarrow \frac{1}{|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|} + \log(|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|/\textstyle\sum_k t_{ikn}v_{kjn}) - 2, \tag{20}$$

$$\boldsymbol{U}_{in} \leftarrow \frac{1}{J}\sum_j\left(\frac{1}{|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|} + \frac{\max(0, d_{ijn})}{2|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|}\right)\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}, \tag{21}$$

$$\tilde{\boldsymbol{r}}_{in} \leftarrow \frac{1}{J}\boldsymbol{U}_{in}^{-1}\sum_j\frac{1}{2}\frac{\overline{\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}}}{|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|}\min(d_{ijn}, 0)\boldsymbol{x}_{ij}, \tag{22}$$

$$\tilde{\boldsymbol{w}}_{in} \leftarrow \boldsymbol{U}_{in}^{-1}\boldsymbol{W}_i^{-1}\boldsymbol{e}_n, \tag{23}$$

$$u_{ww} \leftarrow \tilde{\boldsymbol{w}}_{in}^{\mathsf{H}}\boldsymbol{U}_{in}\tilde{\boldsymbol{w}}_{in}, \tag{24}$$

$$u_{wr} \leftarrow \tilde{\boldsymbol{w}}_{in}^{\mathsf{H}}\boldsymbol{U}_{in}\tilde{\boldsymbol{r}}_{in}, \tag{25}$$

$$\boldsymbol{w}_{in} \leftarrow \begin{cases} \dfrac{\tilde{\boldsymbol{w}}_{in}}{\sqrt{u_{ww}}} - \tilde{\boldsymbol{r}}_{in} & (\text{if } u_{wr}\det\boldsymbol{W}_i = 0) \\ \dfrac{u_{wr}}{2u_{ww}}\left(1 - \sqrt{1 + \dfrac{4u_{ww}}{|u_{wr}|^2}}\right)\tilde{\boldsymbol{w}}_{in} - \tilde{\boldsymbol{r}}_{in} & \\ & (\text{otherwise}) \end{cases}. \tag{26}$$

In KL-ILRMA, the cost function Eq. (6) is minimized by alternately repeating the update of the source spectrograms $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$ using Eqs. (8) and (9), and the update of the demixing matrix $\boldsymbol{W}_i$ using Eqs. (20)–(26). Since all update rules in KL-ILRMA are derived by the MM algorithm, a monotonic decrease in the cost is guaranteed.

3. Numerical Simulation

3.1 Evaluation of KL-ILRMA with Toy Model

We confirmed that KL-ILRMA is valid for separating sources that follow a time-frequency-varying complex Poisson distribution by using an artificial sound source model, i.e., a toy model. We created the toy model via the following three steps. First, we generated each entry of $\tilde{\boldsymbol{T}}_n \in \mathbb{R}_{\geq 0}^{I\times K}$ and $\tilde{\boldsymbol{V}}_n \in \mathbb{R}_{\geq 0}^{K\times J}$ with independent gamma distributions. Second, we calculated the normalized low-rank matrix $\boldsymbol{R}_n$ by

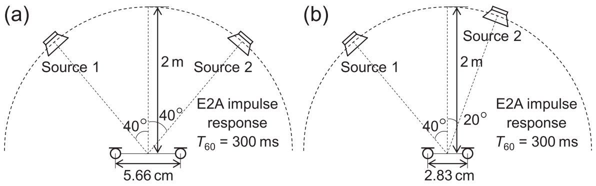


Fig. 2 Recording conditions of impulse responses E2A (T_{60} = 300 ms) obtained from RWCP database [16]: (a) IR1 and (b) IR2.

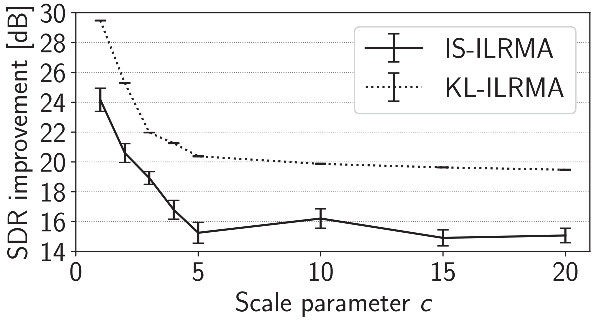


Fig. 3 Average SDR improvements of IS-ILRMA and KL-ILRMA for toy model separation.

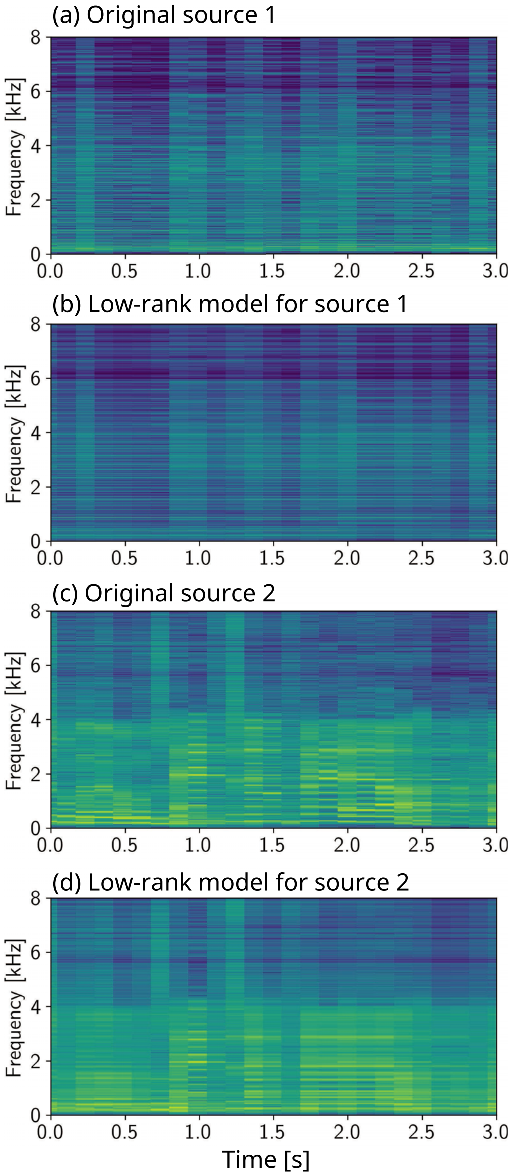
dividing $\tilde{\boldsymbol{T}}_n\tilde{\boldsymbol{V}}_n$ by the maximum value of $\tilde{\boldsymbol{T}}_n\tilde{\boldsymbol{V}}_n$. Third, we obtained the source spectrogram matrix $\boldsymbol{S}_n$ whose ith row and jth column entry is generated from the independent complex Poisson distribution with the shape (mean) parameter λ_{ijn} given by each entry of $c\boldsymbol{R}_n$, where c is the scale parameter of the toy model. In this experiment, the toy model was created with a size of $I = 257$ and $J = 514$. The number of bases was $K = 10$, and the kurtosis of each gamma distribution was set to 12 in $\tilde{\boldsymbol{T}}_n$ and 2 in $\tilde{\boldsymbol{V}}_n$. Note that in the time-frequency-varying complex Poisson distribution, λ_{ijn} controls both the shape of the distribution and the spectrogram strength, and consequently, various settings for the scale parameter c should be tested; we set $c = 1, 2, 3, 4, 5, 10, 15$, and 20.

We produced the two-channel observed signals by convoluting the IR1 impulse response (shown in Fig. 2(a)) with each source. We used the source-to-distortion ratio (SDR) [15] as the total separation performance. The number of bases of the source model in each ILRMA was 10, which is the same as the number of bases used to generate the toy model, and the number of iterations was 200. The initial demixing matrices $\boldsymbol{W}_i$ were set to the ideal value with 5% noise, and the initial source model matrices $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$ were set to uniformly distributed random values.

Figure 3 shows the SDR improvement for each method plotted against the scale parameter c, where the plotted values are the average of 20 trials with different initial values of $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$, and the error bar represents the standard deviation. KL-ILRMA outperforms IS-ILRMA for all values of c, and the standard deviation of KL-ILRMA is within 0.03 dB and much less than that of IS-ILRMA. This result clearly implies the robustness of KL-ILRMA against the initialization of the source model, which is due to the convex property of KL-NMF. Although KL-ILRMA causes a mismatch between the cost function Eq. (6) and the log-likelihood function Eq. (4) of the generative model due to the Stirling approximation Eq. (5), the approximation has no practical effect because the SDR improvement of KL-ILRMA is greater than or around 20 dB.

Table 1 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


(a) Original source 1

(b) Low-rank model for source 1

(c) Original source 2

(d) Low-rank model for source 2

Fig. 4 Original and separated spectrograms for No. 2 in KL-ILRMA. We only show truncated spectrograms with three-second duration.

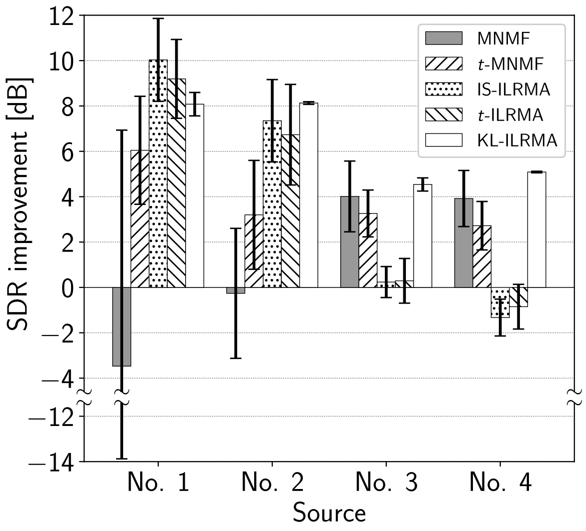


Fig. 5 Average SDR improvements for MNMF, *t*-MNMF, IS-ILRMA, *t*-ILRMA, and KL-ILRMA.

3.2 Evaluation of KL-ILRMA with Audio Source Separation

We compared the separation performance of KL-ILRMA with those of other conventional methods: MNMF [5], *t*-MNMF [10], IS-ILRMA [6], and *t*-ILRMA [11]. We used the music signal `bearlin-roads` in SiSEC2011 [17] as the dry sources, where acoustic guitar, vocal, and piano were used. We produced four different two-channel observed signals (No. 1–No. 4) by convoluting the IR1 and IR2 impulse responses (respectively shown in Figs. 2(a) and (b)) with each source. Table 1 shows the pair of instruments and impulse response for each source. In IS-ILRMA, *t*-ILRMA, and KL-ILRMA, the initial demixing matrices were set to identity matrices and the entries of the initial source model matrices were set to uniformly distributed random values. In MNMF and *t*-MNMF, the initial values of the parameters were set as in [5]. The degree of freedom parameter in *t*-MNMF was set to $\nu = 1$ and the degree of freedom parameter and the domain parameter in *t*-ILRMA were set to $\nu = 3$ and $p = 2$, respectively, which were the best settings for this experiment. In *t*-ILRMA, we applied a tempering technique based on [11]. The sampling frequency was 16 kHz. An STFT was performed using a 512-ms-long Hamming window with a 128-ms-long shift. The numbers of iterations, bases, and trials were set to 200, 10, and 20, respectively.

Figure 4 shows an example of separated spectrograms for No. 2, where we depict the original source spectrograms as reference and their low-rank models estimated in KL-ILRMA. It is confirmed that the proposed method can approximate the source spectrograms by low-rank matrices appropriately.

Figure 5 shows the average SDR improvements for each method, where the error bar represents the standard deviation. In terms of the average improvements, KL-ILRMA outperforms the other methods for No. 2–No. 4, and the standard deviation of KL-ILRMA is the smallest among the methods for all mixed signals. This result confirms the ability of KL-ILRMA for the stable estimation of sources.

4. Conclusion

We proposed a new BSS method, KL-ILRMA, which estimates the low-rank source model via a convex optimization based on generalized KL divergence. We derived the update rule in KL-ILRMA using the MM algorithm to guarantee a monotonic decrease in the cost function. From the experimental evaluation, we confirmed the robustness of KL-ILRMA against the parameter initialization and its ability for the stable estimation of sources.

References

[1] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” Neurocomputing, vol.22, no.1, pp.21–34, 1998.

[2] 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. Audio, Speech, Language Process., vol.14, no.2, pp.666–678, 2006.

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

[4] A. Ozerov and C. Févotte, “Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation,” IEEE Trans. Audio, Speech, Language Process., vol.18, no.3, pp.550–563, 2010.

[5] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, “Multichannel extensions of non-negative matrix factorization with complex-valued data,” IEEE Trans. Audio, Speech, Language Process., vol.21, no.5, pp.971–982, 2013.

[6] 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. Audio, Speech, Language Process., vol.24, no.9, pp.1626–1641, 2016.

[7] 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,” Proc. LVA/ICA, pp.245–253, 2010.

[8] D.D. Lee and H.S. Seung, “Algorithms for non-negative matrix factorization,” Proc. NIPS, pp.556–562, 2000.

[9] A. Cichocki, R. Zdunek, A.H. Phan, and S. Amari, Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation, John Wiley & Sons, UK, 2009.

[10] K. Kitamura, Y. Bando, K. Itoyama, and K. Yoshii, “Student’s *t* multichannel nonnegative matrix factorization for blind source separation,” Proc. IWAENC, 2016.

[11] 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.

[12] H. Kameoka, “Towards a statistical audio signal processing framework based on the I-divergence,” Proc. 2011 Spring Meeting of Acoustical Society of Japan, pp.813–814, 2011.

[13] D.R. Hunter and K. Lange, “Quantile regression via an MM algorithm,” J. Comput. Graph. Stat., vol.9, no.1, pp.60–77, 2000.

[14] Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo, “Vectorwise coordinate descent algorithm for

spatially regularized independent low-rank matrix analysis," Proc. ICASSP, pp.746–750, 2018.

[15] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," IEEE Trans. Audio, Speech, Language Process., vol.14, no.4, pp.1462–1469, 2006.

[16] 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.

[17] 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 -," Proc. LVA/ICA, pp.414–422, 2012.

[18] Y. Takahashi, T. Takatani, K. Osako, H. Saruwatari, and K. Shikano, "Blind spatial subtraction array for speech enhancement in noisy environment," IEEE Trans. Audio, Speech, Language Process., vol.17, no.4, pp.650–664, 2009.

[19] H. Saruwatari, Y. Ishikawa, Y. Takahashi, T. Inoue, K. Shikano, and K. Kondo, "Musical noise controllable algorithm of channelwise spectral subtraction and adaptive beamforming based on higher order statistics," IEEE Trans. Audio, Speech, Language Process., vol.19, no.6, pp.1457–1466, 2011.

Appendix: Validity of Time-Frequency-Varying Complex Poisson Distribution Model

As shown in Sect. 2.2, the low-rank approximation based on the time-frequency-varying complex Poisson distribution is equivalent to KL-NMF. In this appendix, we clarify what kind of real-world signal follows the statistical model.

We compared the accuracy for modeling various real-world signals with the time-frequency-varying complex Poisson distribution (KL-NMF) and the time-frequency-varying complex Gaussian distribution (IS-NMF). We used three kinds of signals for evaluation: 18 music signals, 18 speech signals [17], and 18 environmental sounds [18], [19]. We plotted the source-to-artifact ratio (SAR) of the signals approximated via both NMFs in Fig. A· 1. We can confirm that SAR of the signals approximated via KL-NMF is higher than that approximated via IS-NMF. This result indicates that the generative model based on the time-frequency-varying complex Poisson distribution can be valid for real-world signals, especially for music signals.

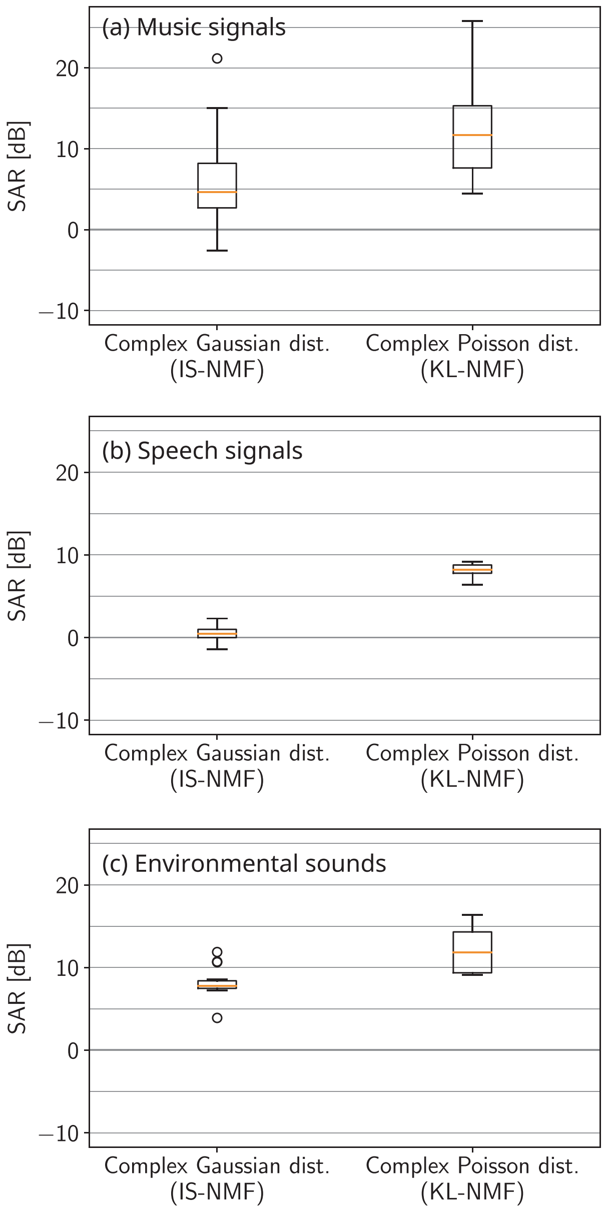


Fig. A· 1 Box-and-whisker plot on SAR of signals approximated via IS-NMF and KL-NMF: (a) music signals, (b) speech signals, and (c) environmental sounds.