EURASIP Journal on Advances
in Signal Processing

RESEARCH Open Access

Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source separation

Daichi Kitamura[1]*, Shinichi Mogami[2], Yoshiki Mitsui[2], Norihiro Takamune[2], Hiroshi Saruwatari[2], Nobutaka Ono[3], Yu Takahashi[4] and Kazunobu Kondo[4]

Abstract

In this paper, statistical-model generalizations of independent low-rank matrix analysis (ILRMA) are proposed for achieving high-quality blind source separation (BSS). BSS is a crucial problem in realizing many audio applications, where the audio sources must be separated using only the observed mixture signal. Many algorithms for solving BSS have been proposed, especially in the history of independent component analysis and nonnegative matrix factorization. In particular, ILRMA can achieve the highest separation performance for music or speech mixtures, where ILRMA assumes both independence between sources and the low-rankness of time-frequency structure in each source. In this paper, we propose two extensions of the source distribution assumed in ILRMA. We introduce a heavy-tailed property by replacing the conventional Gaussian source distribution with a generalized Gaussian or Student's *t* distribution. Convergence-guaranteed efficient algorithms are derived for the proposed methods, and the relationship between the generalized Gaussian and Student's *t* distributions in the source model estimation is revealed. By experimental evaluation, the validity of the heavy-tailed generalizations of ILRMA is confirmed.

Keywords: Blind audio source separation, Independent low-rank matrix analysis, Nonnegative matrix factorization, Student's *t* distribution, Generalized Gaussian distribution

1 Introduction

Blind source separation (BSS) is a technique for separating individual sources from an observed multichannel mixture without knowing the mixing system, such as the spatial locations of the sensors or sources, in advance. In particular, BSS for multichannel audio signals have been well studied so far. This problem can be divided into two situations: underdetermined (number of microphones $<$ number of sources) and (over-)determined (number of microphones $\geq$ number of sources) cases. In the underdetermined case, the mixing system of the sources has to be estimated using several assumptions. For example, sparseness-assumption-based methods are popular and reliable approaches [1–3]. In contrast, the determined BSS methods often estimate the inverse system of a mixing process, and high-quality separation can be achieved compared with the underdetermined BSS methods. In this paper, we only focus on the determined BSS problem.

The most popular and successful algorithm for solving determined BSS problem is independent component analysis (ICA) [4], which assumes statistical independence between the sources and estimates a demixing matrix (the inverse system of the mixing process). For a mixture of audio signals, because the sources are mixed by convolution owing to the room reverberation, ICA is often applied to the time-frequency signals (spectrograms) of the observed signal, which are obtained by a short-time Fourier transform (STFT). Frequency-domain ICA (FDICA) [5–8] independently applies ICA to the complex-valued time-series signals in each frequency bin and estimates a frequency-wise demixing matrix. Then, the estimated components in each frequency must be aligned over all frequency bins so that the components of the same source are grouped. This postprocessing of FDICA is the so-called permutation problem [6, 9–11],

*Correspondence: d-kitamura@ieee.org
[1] National Institute of Technology, Kagawa College, 355 Chokushi, Takamatsu, Kagawa 761-8058, Japan
Full list of author information is available at the end of the article

and several criteria have been used to solve this ambiguity of the signal permutation.

Independent vector analysis (IVA) [12–14] is a sophisticated algorithm that can simultaneously estimate the frequency-wise demixing matrix and solve the permutation problem using only one objective function. IVA assumes higher-order dependences (co-occurrence among the frequency bins) of each source by employing a spherical generative model of the source frequency vector, thus avoiding the permutation problem. The original IVA employs the spherical multivariate Laplace distribution as the source model (hereafter referred to as *Laplace IVA*). To improve the statistical model flexibility and source separation performance, Laplace IVA has been extended by replacing its source model with a spherical generalized Gaussian distribution [15] (GGD, also known as an exponential power distribution) in many papers [16–20] (hereafter referred to as *GGD-IVA*), or with a Gaussian distribution having a time-varying variance [21] (hereafter referred to as *time-varying Gaussian IVA*). Note that the GGD includes the Laplace and Gaussian distributions as special cases.

As another means of audio source modeling and separation, nonnegative matrix factorization (NMF) [22, 23] has been a very common approach during the last decade. NMF is a nonnegative-parts-based low-rank decomposition of an observed nonnegative data matrix that is typically a power or amplitude spectrogram. The decomposed nonnegative parts (bases and activations) can be used for source separation by clustering the parts into each source [24–28]. Also, NMF can be statistically interpreted as a parameter estimation based on a generative model of data, and the distribution of the model defines the objective function (divergence) in NMF. For example, it was revealed that NMF based on Itakura–Saito divergence (IS-NMF) assumes an isotropic complex Gaussian distribution independently defined in each time-frequency slot [29], where the variance of each Gaussian distribution can fluctuate depending on time and frequency. For multichannel audio signals, spatial modeling of the mixing system was introduced into the simple NMF, which is called multichannel NMF (MNMF) [30–32], to solve the BSS problem. MNMF estimates the spatial mixing system, whereas ICA-based BSS techniques optimize the demixing matrix, which yields a more stable and efficient algorithm than MNMF.

Motivated by this issue, a new BSS algorithm called *independent low-rank matrix analysis* (ILRMA) [33–35] has been proposed[1]. In this method, IS-NMF-based low-rank source modeling is introduced into the source model of IVA, namely, a low-rank time-frequency structure (co-occurrence among the time-frequency slots) is estimated for each source by NMF, and the frequency-wise demixing matrix is optimized taking the NMF source model into account without causing the permutation problem. Since the vector source model in time-varying Gaussian IVA can be interpreted as NMF with a single spectral basis, ILRMA is a natural extension of IVA, where ILRMA utilizes an arbitrary number of bases in the source model. Also, ILRMA can be considered as a dual problem of MNMF (mixing) because ILRMA estimates the demixing matrix, i.e., the inverse of the mixing system (MNMF model), using the low-rank source modeling with NMF.

In this paper, to increase the model flexibility and improve the source separation accuracy, we generalize the source model in ILRMA from the isotropic complex Gaussian distribution of IS-NMF to more heavy-tailed distributions. An important extension is to employ the isotropic complex GGD because it has been reported that GGD-IVA can achieve a better separation result in many papers [17, 19, 20]. As another possible generalization, the isotropic complex Student's t distribution can also be employed in ILRMA. Student's t distribution includes the Cauchy and Gaussian distributions as special cases and has been used to model audio sources [36, 37]. For use in NMF-based modeling, Cauchy NMF [38], Student's t NMF (t-NMF) [39], and its multichannel extension (t-MNMF) [40] have been proposed. The motivation of employing Student's t distribution is that the Cauchy and Gaussian distributions are a part of the α-stable distribution family [41], which has a stable property of a random variable, namely, a linear combination of two independent random variables generated from the same distribution family also has the same distribution up to location and scale parameters. This property is desirable for NMF-based audio source modeling because it justifies the nonnegative linear decomposition of complex-valued signals [42]. For instance, multichannel BSS based on an α-stable distribution was recently proposed [43] to benefit from this advantage. However, analytical maximum likelihood (ML) estimation with an α-stable distribution is still an open problem because its probability density function (p.d.f.) cannot be represented in a closed form except for several cases. Therefore, instead of employing the α-stable distribution family, we adopt Student's t distribution as the source model in ILRMA, which partly corresponds to the α-stable distribution and has the stable property. The relationship among the conventional methods and the proposed ILRMA is depicted in Fig. 1. As shown in this figure, the proposed ILRMA based on the GGD (GGD-ILRMA) and that based on Student's t distribution (t-ILRMA) can be interpreted as a new extension of conventional IVA or ILRMA as well as a computationally efficient solution to the dual problem of MNMF.

Note that this work extends our preliminary work on t-ILRMA in [44] by developing a new extension, GGD-ILRMA, and providing additional discussion that explains the theoretical relationship between GGD- and t-ILRMA.

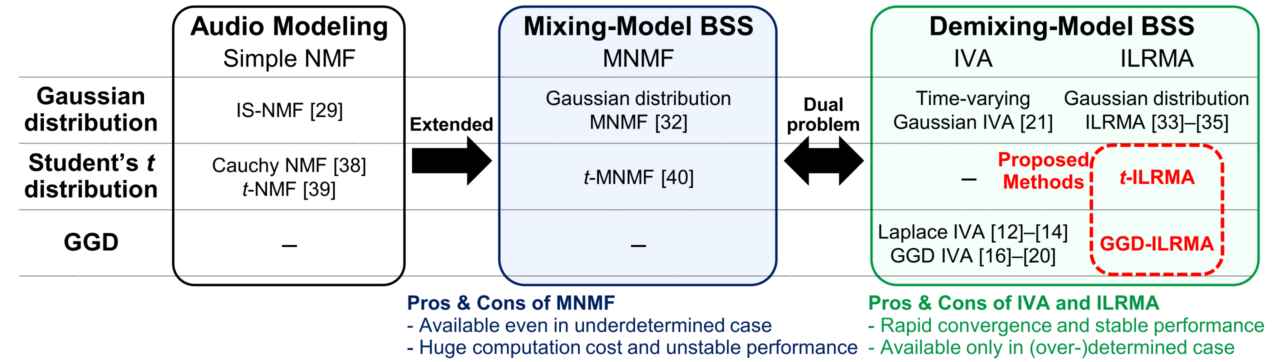
Audio Modeling
Simple NMF
Mixing-Model BSS
MNMF
Demixing-Model BSS
IVA
ILRMA
Gaussian distribution
IS-NMF [29]
Extended
Gaussian distribution MNMF [32]
Dual problem
Time-varying Gaussian IVA [21]
Gaussian distribution ILRMA [33]–[35]
Student's t distribution
Cauchy NMF [38]
t-NMF [39]
t-MNMF [40]
–
Proposed Methods
t-ILRMA
GGD
–
–
Laplace IVA [12]–[14]
GGD IVA [16]–[20]
GGD-ILRMA
Pros & Cons of MNMF
- Available even in underdetermined case
- Huge computation cost and unstable performance
Pros & Cons of IVA and ILRMA
- Rapid convergence and stable performance
- Available only in (over-)determined case

Fig. 1 Relationship among conventional methods based on Gaussian, Student's *t*, and generalized Gaussian distributions. NMF-based low-rank audio source modeling is extended to a multichannel signal, resulting in MNMF, which estimates the mixing system for solving the BSS problem. ILRMA can be considered as a dual problem of MNMF because it estimates the demixing system using the NMF source model

Also, the experimental results have been updated with new datasets and conditions for more difficult situations in BSS.

The rest of this paper is organized as follows. Section 2 describes the conventional algorithms including IVA and ILRMA, which are the basis for the proposed GGD- and *t*-ILRMA described in Section 3. Section 4 reports the validation of the proposed methods by conducting BSS experiments with music and speech sources. Finally, Section 5 concludes this paper.

2 Conventional method

2.1 Formulation

Let $\tilde{s}_n(\tau)$, $\tilde{x}_m(\tau)$, and $\tilde{y}_n(\tau)$ be the source, observed (mixture), and estimated (separated) time-domain signals, respectively, where $n = 1, \cdots, N$ and $m = 1, \cdots, M$ are the integral indexes of the sources and channels (microphones), respectively. Also, τ is the integral index of the discrete time. The source signal $\tilde{s}_n(\tau)$ is unknown, and only the observed signal $\tilde{x}_m(\tau)$ can be obtained by using the synchronized multiple microphones. The estimated signal $\tilde{y}_n(\tau)$ is the output data of BSS algorithm. In this paper, these time-domain signals are transformed into the time-frequency domain to treat the convolutive mixture with the room reverberation. The complex-valued time-frequency components of $\tilde{s}_n(\tau)$, $\tilde{x}_m(\tau)$, and $\tilde{y}_n(\tau)$ can be obtained via STFT and are respectively denoted as follows:

$$\boldsymbol{s}_{ij} = (s_{ij,1}, \cdots, s_{ij,n}, \cdots, s_{ij,N})^{\mathrm{T}} \in \mathbb{C}^{N\times 1}, \tag{1}$$

$$\boldsymbol{x}_{ij} = (x_{ij,1}, \cdots, x_{ij,m}, \cdots, x_{ij,M})^{\mathrm{T}} \in \mathbb{C}^{M\times 1}, \tag{2}$$

$$\boldsymbol{y}_{ij} = (y_{ij,1}, \cdots, y_{ij,n}, \cdots, y_{ij,N})^{\mathrm{T}} \in \mathbb{C}^{N\times 1}, \tag{3}$$

where $i = 1, \cdots, I$ and $j = 1, \cdots, J$ are the integral indexes of the frequency bins and time frames, respectively, and $^{\mathrm{T}}$ denotes a transpose. We also denote the spectrograms (time-frequency matrices) of the source, observed, and estimated signals as $\boldsymbol{S}_n \in \mathbb{C}^{I\times J}$, $\boldsymbol{X}_m \in \mathbb{C}^{I\times J}$, and $\boldsymbol{Y}_n \in \mathbb{C}^{I\times J}$, whose elements are $s_{ij,n}$, $x_{ij,m}$, and $y_{ij,n}$, respectively. In FDICA, IVA, and ILRMA, the following mixing system is assumed:

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

where $\boldsymbol{A}_i = (\boldsymbol{a}_{i,1} \cdots \boldsymbol{a}_{i,n} \cdots \boldsymbol{a}_{i,N}) \in \mathbb{C}^{M\times N}$ is a frequency-wise mixing matrix and $\boldsymbol{a}_{i,n} = (a_{i,n1}, \cdots, a_{i,nm}, \cdots, a_{i,nM})^{\mathrm{T}}$is the steering vector for the nth source, which represents the acoustic transfer functions from the nth source to each of the microphones ($m = 1, \cdots, M$). The assumed mixing system (4) is called a linear time-invariant mixture or rank-1 spatial model [45] because the spatial covariance of each source image (multichannel observation of each source signal) is restricted to a rank-1 matrix in this system [34]. If the mixing system is determined, namely, $M = N$, and $\boldsymbol{A}_i$ is a non-singular matrix for all i, we can define the frequency-wise demixing matrix $\boldsymbol{W}_i = (\boldsymbol{w}_{i,1} \cdots \boldsymbol{w}_{i,n} \cdots \boldsymbol{w}_{i,N})^{\mathrm{H}} = \boldsymbol{A}_i^{-1}$ that recovers the source signal, and the estimated signal $\boldsymbol{y}_{ij}$ is obtained as

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

where $\boldsymbol{w}_{i,n}$ is the demixing filter for the nth source and $^{\mathrm{H}}$ denotes a Hermitian transpose. The goal of BSS based on FDICA, IVA, or ILRMA is to estimate $\boldsymbol{W}_i$ and obtain $\boldsymbol{y}_{ij}$ from only the observations $\boldsymbol{x}_{ij}$ by assuming statistical independence between $s_{ij,n}$ and $s_{ij,n'}$, where $n' \neq n$. In this paper, we only focus on BSS with the determined situation $M = N$. For the overdetermined situation $M > N$, principal component analysis is often applied to $\boldsymbol{x}_{ij}$ for dimensionality reduction so that $M = N$ [46].

2.2 IVA

IVA [12–14] is an elegant solution of the permutation problem [6, 9–11], which considers not the frequency-wise component $x_{ij,m}$ but the vector of all frequency components, $\bar{\boldsymbol{x}}_{j,m} = (x_{1j,m}, \ \cdots \ x_{Ij,m})^{\mathrm{T}} \in \mathbb{C}^{I \times 1}$, as an independent variable as shown in Fig. 2. Thus, in IVA, ICA is applied to the time-series vectors $\bar{\boldsymbol{x}}_{1,m} \ \cdots \ \bar{\boldsymbol{x}}_{J,m}$ while assuming the spherical I-dimensional non-Gaussian distribution $p(\bar{\boldsymbol{s}}) \approx p(\bar{\boldsymbol{y}})$. For example, the generative model in GGD-IVA [16–18, 20] is represented as

$$p(\bar{\boldsymbol{y}}_{j,n}) \propto \exp\left(-\|\bar{\boldsymbol{y}}_{j,n}\|_2^{\beta}\right), \tag{6}$$

where $\|\cdot\|_2$ denotes the L_2 norm and $\beta > 0$ is the shape parameter of GGD. Laplace IVA [12–14] corresponds to $\beta = 1$. Since the probability of (6) only depends on the norm of $\bar{\boldsymbol{y}}_{j,n}$ (spherical property), the components in the vector $\bar{\boldsymbol{y}}_{j,n}$ have higher-order dependence. Therefore, frequency components that have similar activations, such as a fundamental frequency and its harmonic components, will be merged as one source avoiding the permutation problem.

By assuming the independence between the source vectors, the objective function (negative log-likelihood function of the observed signal) in IVA can be obtained as

$$\mathcal{L}_{\mathrm{IVA}} = -2J\sum_i \log|\det \boldsymbol{W}_i| + \sum_{j,n} G(\bar{\boldsymbol{y}}_{j,n}), \tag{7}$$

where $G(\bar{\boldsymbol{y}}_{j,n}) = -\log p(\bar{\boldsymbol{y}}_{j,n})$ is called a contrast function and $\det \boldsymbol{W}_i$ denotes the determinant of a matrix $\boldsymbol{W}_i$. Note that the separated signal $y_{ij,n}$ in $\bar{\boldsymbol{y}}_{j,n}$ includes the variable $\boldsymbol{W}_i$ as $y_{ij,n} = \boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{x}_{ij}$.

As another generative model of source signals, an isotropic complex Gaussian distribution with time-varying variance can be utilized in IVA [21], which is represented as

$$\begin{aligned} p(\bar{\boldsymbol{y}}_{1,n}, \ \cdots, \bar{\boldsymbol{y}}_{J,n}) &= \prod_j p(\bar{\boldsymbol{y}}_{j,n}) \\ &= \prod_j \frac{1}{\pi r_{j,n}} \exp\left(-\frac{\|\bar{\boldsymbol{y}}_{j,n}\|_2^2}{r_{j,n}}\right), \end{aligned} \tag{8}$$

where $r_{j,n}$ is the time-varying variance shared over all frequency bins. Similar to (6), (8) also has a spherical property. Note that even though (8) consists of Gaussian distributions, its marginal distribution over j becomes a super-Gaussian distribution because the variance can fluctuate depending on j [17].

Regarding the optimization of $\boldsymbol{W}_i$, a fast and stable optimization algorithm called iterative projection (IP), which is based on a majorization-minimization (MM) algorithm [47], has been derived for ICA [48], Laplace IVA [49], GGD-IVA [17], and time-varying Gaussian IVA [21]. IP can achieve better convergence than classical gradient-based algorithms.

2.3 ILRMA based on Gaussian distribution

2.3.1 Generative model

ILRMA [33–35] is a method unifying IVA and IS-NMF, namely, we assume both statistical independence between sources and the low-rankness of the time-frequency structure in each source. Similar to ICA or IVA, we must assume a non-Gaussian distribution as the generative model of source signals to solve the BSS problem. In ILRMA, the following distribution is assumed for the spectrogram of each source:

$$\begin{aligned} p(\boldsymbol{Y}_n) &= \prod_{i,j} p(y_{ij,n}) \\ &= \prod_{i,j} \frac{1}{\pi r_{ij,n}} \exp\left(-\frac{|y_{ij,n}|^2}{r_{ij,n}}\right), \end{aligned} \tag{9}$$

$$r_{ij,n} = \sum_k t_{ik,n} v_{kj,n}, \tag{10}$$

where $t_{ik,n} \geq 0$ and $v_{kj,n} \geq 0$ are the nonnegative basis and activation elements (NMF variables) of $\boldsymbol{T}_n \in \mathbb{R}_{\geq 0}^{I \times K}$ (basis matrix) and $\boldsymbol{V}_n \in \mathbb{R}_{\geq 0}^{K \times J}$ (activation matrix), respectively, $k = 1, \cdots, K$ is the integral index of the basis, and K is the number of NMF bases (spectral patterns). Also, $r_{ij,n} \geq 0$ is a sourcewise time-frequency-varying

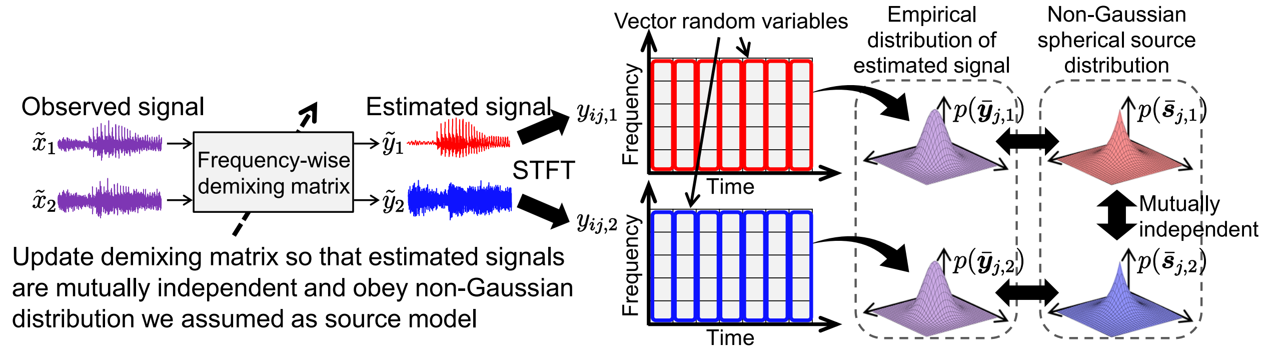
Observed signal
$\tilde{x}_1$
$\tilde{x}_2$
Frequency-wise demixing matrix
Estimated signal
$\tilde{y}_1$
$\tilde{y}_2$
STFT
$y_{ij,1}$
$y_{ij,2}$
Vector random variables
Frequency
Time
Frequency
Time
Empirical distribution of estimated signal
$p(\bar{\boldsymbol{y}}_{j,1})$
$p(\bar{\boldsymbol{y}}_{j,2})$
Non-Gaussian spherical source distribution
$p(\bar{\boldsymbol{s}}_{j,1})$
Mutually independent
$p(\bar{\boldsymbol{s}}_{j,2})$
Update demixing matrix so that estimated signals are mutually independent and obey non-Gaussian distribution we assumed as source model

Fig. 2 Principle of source separation based on IVA. IVA assumes both statistical independence between sources and non-Gaussian spherical distribution for vector variables

variance that corresponds to the low-rank source model. Therefore, the nonnegative matrix $\boldsymbol{T}_n\boldsymbol{V}_n$ represents the rank-K model spectrogram of the nth source as $|\boldsymbol{Y}_n|^{\cdot 2} \approx \boldsymbol{T}_n\boldsymbol{V}_n$, where $|\cdot|^{\cdot q}$ for matrices denotes the element-wise absolute and qth-power operations. Because of the fluctuation of the variance $r_{ij,n}$ of the time and frequency, the marginal distribution of the generative model (9) over j becomes a super-Gaussian distribution, which can be used for independence-based BSS.

The local distribution $p(y_{ij,n})$ is circularly symmetric in the complex plane, and the probability only depends on the power $|y_{ij,n}|^2$. For this reason, the variance $r_{ij,n}$ corresponds to the expectation value of the power spectrum $|y_{ij,n}|^2$, namely, $r_{ij,n} = \mathrm{E}\left[|y_{ij,n}|^2\right]$. In addition, if we assume that the source spectrogram $y_{ij,n}$ consists of K components $c_{ij,nk}$, namely, $y_{ij,n} = \sum_k c_{ij,nk}$, the generative model of $c_{ij,nk}$ also becomes the complex Gaussian distribution because of the stable property as follows:

$$p(c_{ij,nk}) = \frac{1}{\pi\, t_{ik,n}v_{kj,n}} \exp\left(-\frac{|c_{ij,nk}|^2}{t_{ik,n}v_{kj,n}}\right). \tag{11}$$

Note that the variances in $p(y_{ij,n})$ and $p(c_{ij,nk})$ are $r_{ij,n} = \sum_k t_{ik,n}v_{kj,n}$ and $t_{ik,n}v_{kj,n}$, respectively, and they correspond to the expectation values of $|y_{ij,n}|^2$ and $|c_{ij,nk}|^2$ as $r_{ij,n} = \mathrm{E}\left[|y_{ij,k}|^2\right]$ and $t_{ik,n}v_{kj,n} = \mathrm{E}\left[|c_{ij,nk}|^2\right]$, respectively. Even if $y_{ij,n} = \sum_k c_{ij,nk}$, the additivity of the power spectra does not hold ($|y_{ij,n}|^2 \neq \sum_k |c_{ij,nk}|^2$) because of the phase cancelation. However, (9) and (11) mean that the additivity of expectations $t_{ik,n}v_{kj,n} = \mathrm{E}\left[|c_{ij,nk}|^2\right]$ is satisfied as $r_{ij,n} = \sum_k t_{ik,n}v_{kj,n}$ because of the stable property in Gaussian distribution. Therefore, the generative model (9) theoretically justifies to linearly decompose the power spectrogram $|y_{ij,n}|^2$ into K nonnegative parts $t_{ik,n}v_{kj,n}$. This advantage was extended to a more general domain in [42] using an α-stable distribution, which is a distribution family ensuring the stable property. When $\alpha = 2$, α-stable distribution is equal to Gaussian distribution (9) and the additivity of power spectra holds in the expectation sense. When $\alpha = 1$, α-stable distribution converges to Cauchy distribution, which ensures the additivity of amplitude spectra in the expectation sense [38].

2.3.2 Objective function and update rules

The objective function of ILRMA is the negative log-likelihood function of the observed signal $\boldsymbol{x}_{ij}$ and can be obtained from (9) by assuming independence between all sources as

$$\begin{aligned}
\mathcal{L} &\equiv -\log p(\mathsf{X}) \\
&= -2J\sum_i \log|\det \boldsymbol{W}_i| - \log p(\mathsf{Y}) \qquad (12)\\
&= -2J\sum_i \log|\det \boldsymbol{W}_i| \\
&\quad + \sum_{i,j,n}\left(\log\sum_k t_{ik,n}v_{kj,n} + \frac{|y_{ij,n}|^2}{\sum_k t_{ik,n}v_{kj,n}}\right) \qquad (13)\\
&\quad + IJN\log\pi \\
&= -2J\sum_i \log|\det \boldsymbol{W}_i| \\
&\quad + \sum_{i,j,n}\log\sum_k t_{ik,n}v_{kj,n} + J\sum_{i,n}\boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{U}_{i,n}\boldsymbol{w}_{i,n} \qquad (14)\\
&\quad + IJN\log\pi,
\end{aligned}$$

$$\boldsymbol{U}_{i,n} = \frac{1}{J}\sum_j \frac{1}{\sum_k t_{ik,n}v_{kj,n}}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}}, \tag{15}$$

where $\mathsf{X} = \{\boldsymbol{X}_1, \cdots, \boldsymbol{X}_M\}$ and $\mathsf{Y} = \{\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_N\}$ are the set of the observed and estimated signals and the independence between sources, $p(\mathsf{Y}) = \prod_n p(\boldsymbol{Y}_n)$, is assumed. The first and third terms in (13) correspond to the objective function in time-varying Gaussian IVA [21], and the second and third terms correspond to the objective function in IS-NMF [29]. The task of the ILRMA algorithm is to minimize the objective function $\mathcal{L}$ w.r.t. $\boldsymbol{T}_n$, $\boldsymbol{V}_n$, and $\boldsymbol{W}_i$.

For the optimization of the demixing matrix $\boldsymbol{W}_i$, similar to IVA, IP can be used for minimizing $\mathcal{L}$. The update rules based on IP are expressed as follows:

$$\boldsymbol{U}_{i,n} \leftarrow \frac{1}{J}\sum_j \frac{1}{r_{ij,n}}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}}, \tag{16}$$

$$\boldsymbol{w}_{i,n} \leftarrow \left(\boldsymbol{W}_i\boldsymbol{U}_{i,n}\right)^{-1}\boldsymbol{e}_n, \tag{17}$$

$$\boldsymbol{w}_{i,n} \leftarrow \boldsymbol{w}_{i,n}\left(\boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{U}_{i,n}\boldsymbol{w}_{i,n}\right)^{-\frac{1}{2}}, \tag{18}$$

$$y_{ij,n} \leftarrow \boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{x}_{ij}, \tag{19}$$

where $\boldsymbol{e}_n$ denotes the $N \times 1$ unit vector with the nth element equal to unity. By iterating these algorithms, the demixing matrix $\boldsymbol{W}_i$ is updated so that the objective function (14) decreases. Note that IP does not include any step-size parameter in its update rules. Regarding the NMF variables $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$, the following convergence-guaranteed update rules based on the MM algorithm have been derived [50]:

$$t_{ik,n} \leftarrow t_{ik,n}\left[\frac{\sum_j \frac{|y_{ij,n}|^2}{\left(\sum_k t_{ik,n}v_{kj,n}\right)^2}v_{kj,n}}{\sum_j \frac{1}{\sum_k t_{ik,n}v_{kj,n}}v_{kj,n}}\right]^{\frac{1}{2}}, \tag{20}$$

$$v_{kj,n} \leftarrow v_{kj,n}\left[\frac{\sum_i \frac{|y_{ij,n}|^2}{\left(\sum_k t_{ik,n}v_{kj,n}\right)^2}t_{ik,n}}{\sum_i \frac{1}{\sum_k t_{ik,n}v_{kj,n}}t_{ik,n}}\right]^{\frac{1}{2}}, \tag{21}$$

$$r_{ij,n} \leftarrow \sum_k t_{ik,n}v_{kj,n}. \tag{22}$$

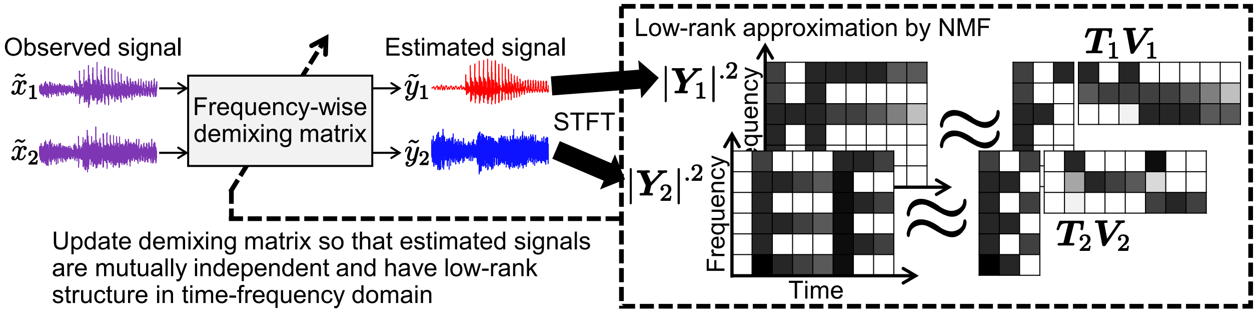
Observed signal
$\tilde{x}_1$
$\tilde{x}_2$
Frequency-wise demixing matrix
Estimated signal
$\tilde{y}_1$
$\tilde{y}_2$
STFT
Low-rank approximation by NMF
$|Y_1|^{.2}$
$|Y_2|^{.2}$
T_1V_1
T_2V_2
Frequency
Time
Update demixing matrix so that estimated signals are mutually independent and have low-rank structure in time-frequency domain

Fig. 3 Principle of source separation based on ILRMA. ILRMA assumes both statistical independence between sources and the low-rankness of time-frequency structures in each source

From the above, the objective function can be efficiently optimized by iterating the update rules (16)–(22). However, a scale ambiguity exists between $\boldsymbol{W}_i$ and $r_{ij,n}$ because both of them can determine the scale of the separated signal $y_{ij,n}$. Therefore, $\boldsymbol{W}_i$ or $r_{ij,n}$ has a risk of diverging during the optimization. To avoid this problem, the following normalization should be applied at each iteration:

$$\boldsymbol{w}_{i,n} \leftarrow \boldsymbol{w}_{i,n}\lambda_n^{-1}, \tag{23}$$

$$y_{ij,n} \leftarrow y_{ij,n}\lambda_n^{-1}, \tag{24}$$

$$r_{ij,n} \leftarrow r_{ij,n}\lambda_n^{-2}, \tag{25}$$

$$t_{ik,n} \leftarrow t_{ik,n}\lambda_n^{-2}, \tag{26}$$

where λ_n is an arbitrary sourcewise normalization coefficient such as the sourcewise average power $\lambda_n = \left[(IJ)^{-1}\sum_{i,j}|y_{ij,n}|^2\right]^{\frac{1}{2}}$. These normalizations do not change the value of the objective function (13). The scale of the separated signal $y_{ij,n}$ can be restored by applying the following back-projection technique [51] after the optimization:

$$\hat{\boldsymbol{y}}_{ij,n} = \boldsymbol{W}_i^{-1}\left(\boldsymbol{e}_n \circ \boldsymbol{y}_{ij}\right), \tag{27}$$

where $\hat{\boldsymbol{y}}_{ij,n} = (\hat{y}_{ij,n1} \ \cdots \ \hat{y}_{ij,nM})^{\mathrm{T}}$ is a separated source image whose scale is fitted to the observed signals at each

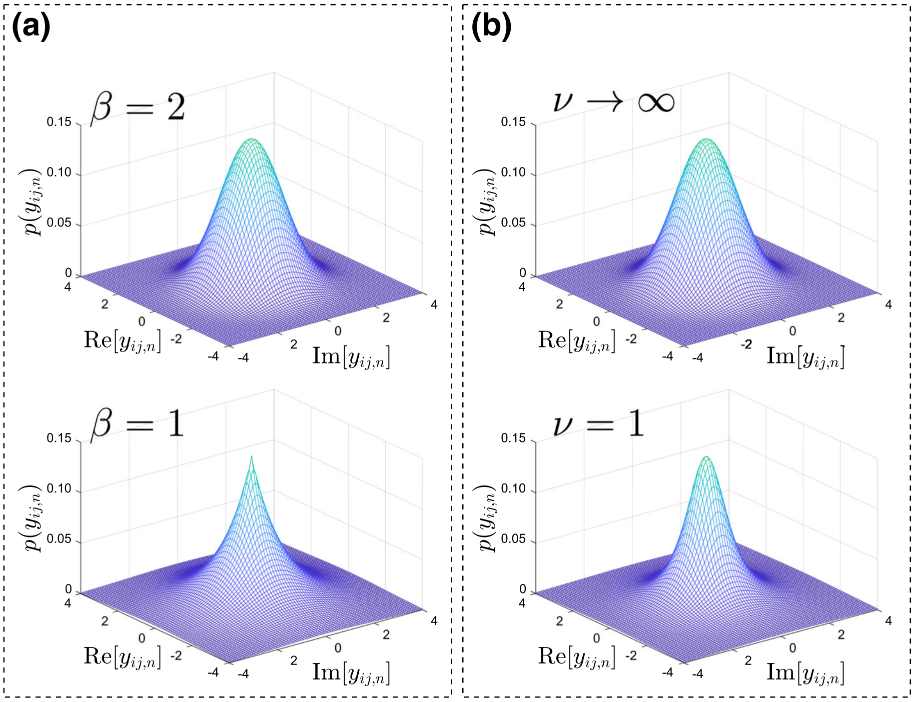

Fig. 4 Source models assumed in proposed methods: **a** isotropic complex GGD and **b** isotropic complex Student's *t* distribution. The GGD includes Gaussian and Laplace distributions as special cases, and the Student's *t* distribution includes Gaussian and Cauchy distributions as special cases

Table 1 Summary of parameterized properties in GGD- and *t*-ILRMA

	Shape parameter (β, ν)	Domain parameter (p)
GGD-ILRMA	$\beta \to 0$	$p \to 0$
	• Low-rankness injection via geometric mean	• Low-rankness mitigation
	• Faster NMF update	• Slower NMF update
	Special cases	
	• $\beta = 2$: Gaussian dist.	
	• $\beta = 1$: Laplace dist.	
t-ILRMA	$\nu \to 1$	$p \to 0$
	• Low-rankness injection	• Low-rankness mitigation
	via harmonic mean	• Slower NMF update
	Special cases	
	• $\nu \to \infty$: Gaussian dist.	
	• $\nu = 1$: Cauchy dist.	

microphone and $\circ$ denotes the Hadamard product (entry-wise multiplication). The detailed implementation can be found in [52].

Figure 3 shows the separation principle of ILRMA. When the original sources have a low-rank spectrogram $|\boldsymbol{S}_n|^{\cdot 2}$, the spectrogram of their mixture, $|\boldsymbol{X}_m|^{\cdot 2}$, should be more complicated, where the rank of $|\boldsymbol{X}_m|^{\cdot 2}$ should be greater than that of $|\boldsymbol{S}_n|^{\cdot 2}$. On the basis of this assumption, in ILRMA, the low-rank constraint for each estimated spectrogram $|\boldsymbol{Y}_n|^{\cdot 2}$ is introduced by employing NMF. The demixing matrix $\boldsymbol{W}_i$ is estimated so that the spectrogram of the estimated signal $|\boldsymbol{Y}_n|^{\cdot 2}$ becomes a low-rank matrix modeled by $\boldsymbol{T}_n\boldsymbol{V}_n$, whose rank is at most K. The estimation of $\boldsymbol{W}_i$, $\boldsymbol{T}_n$, and $\boldsymbol{V}_n$ can consistently be carried out by minimizing (13) in a fully blind manner. ILRMA is theoretically equivalent to conventional MNMF only when the rank-1 spatial model (4) is assumed, which yields a stable and computationally efficient algorithm for ILRMA. This issue has been well discussed in [34, 35].

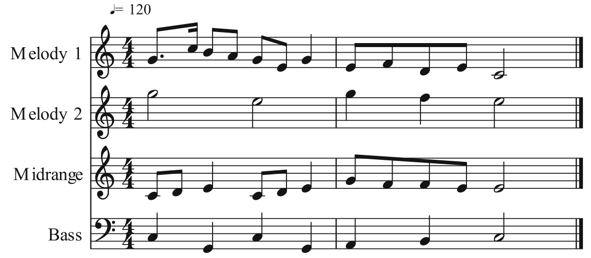

Fig. 5 Scores of each part. The signal length is approximately 5 s

Table 2 Musical instruments used in the music dataset

Part	Instruments
Melody 1	Oboe, trumpet, and horn
Melody 2	Flute, violin, and clarinet
Midrange	Piano and harpsichord
Bass	Trombone, bassoon, and cello

3 Proposed generalization of ILRMA

3.1 Motivation and strategy

The conventional ILRMA described in Section 2.3 is based on the isotropic complex Gaussian distribution (9) with a time-frequency-varying variance $r_{ij,n}$. For independence-based BSS, non-Gaussianity of the source signals is required for the separation, and the model (9) relies on only the fluctuation of the variance $r_{ij,n}$. If the variance $r_{ij,n}$ is a constant value for all i and j, the model (9) becomes completely Gaussian and the independence-based BSS collapses because the ICA algorithm cannot distinguish multiple Gaussian sources. Therefore, it is worth generalizing the distribution in ILRMA to a more flexible non-Gaussian source model. In fact, several approaches based on a non-Gaussian distribution with a time-frequency-varying parameter, such as t-NMF, have been proposed, and it has been reported that NMF audio source modeling based on a non-Gaussian distribution provides better separation performance [39]. From the IVA side, the source distribution has also been generalized by employing the GGD in many studies [16–20], which gave more accurate BSS results.

For the reasons mentioned above, in this section, we propose two generalizations of the source distribution (generative model) in ILRMA using heavy-tailed distributions: the isotropic complex GGD and the isotropic complex Student's t distribution. The former is a natural extension of the conventional generative model (9) and has often been used for the generalization of Laplace IVA or time-varying Gaussian IVA as GGD-IVA. The GGD

Table 3 Dry sources used in two-source case

Signal	Data name	Sources (1/2)	Signal length [s]
Music 1	Melody 2/Midrange	Flute/piano	5.0
Music 2	Melody 1/Melody 2	Oboe/flute	5.0
Music 3	Melody 1/Bass	Trumpet/bassoon	5.0
Music 4	Melody 2/Midrange	Violin/harpsichord	5.0
Music 5	Melody 1/Melody 2	Horn/blarinet	5.0
Music 6	Midrange/Bass	Piano/cello	5.0
Speech 1	dev1_female4	src_1/src_2	10.0
Speech 2	dev1_female4	src_3/src_4	10.0
Speech 3	dev1_male4	src_1/src_2	10.0
Speech 4	dev1_male4	src_3/src_4	10.0

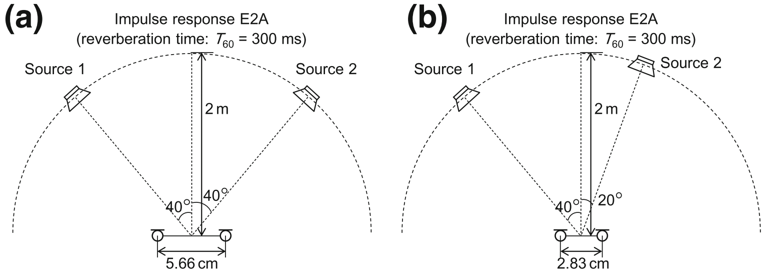

Fig. 6 Spatial arrangements of impulse responses used in two-source case: **a** `E2A_1` and **b** `E2A_2`. Since the microphone spacing and the angle between the two sources in `E2A_2` are smaller than those in `E2A_1`, BSS is more difficult for `E2A_2`

has a shape parameter that controls the super- or sub-Gaussianity. In particular, the GGD includes Laplace and Gaussian distributions as special cases. Since most audio sources follow super-Gaussian distributions, in this paper, we only focus on GGD-ILRMA with a super-Gaussian region.

The latter generalization was inspired by a recently developed framework [42] that ensures the stable property of complex-valued random variables, i.e., audio modeling based on an α-stable distribution. In this model, similar to IS-NMF (11), the decomposition of a complex-valued spectrogram into several nonnegative parts is theoretically justified by the stable property of this distribution family. Student's t distribution has a degree-of-freedom parameter that determines the shape of the distribution and its super-Gaussianity. Similar to the GGD, Student's t distribution includes Cauchy and Gaussian distributions as special cases, which are also special cases of the α-stable distribution. Therefore, NMF source modeling (decomposition of complex-valued spectrogram $\boldsymbol{Y}_n$) in t-ILRMA is partially justified when the Gaussian or Cauchy distribution is assumed, which is theoretically preferable for audio signal processing.

In addition, we introduce a new domain parameter for NMF modeling in GGD- and t-ILRMA because the generative model of a spectrogram strongly depends on the data domain, such as the selection of the amplitude- or power-domain spectrogram to be used. By controlling both the generative model and the modeling domain of data, we can find a suitable statistical assumption for the audio BSS problem.

Table 4 Experimental conditions

Sampling frequency	16 kHz
Window function in STFT	Hamming window
Window length in STFT	4096 points (256 ms)
Shift length in STFT	2048 points (128 ms)
Number of NMF bases K	Four for music case and two for speech case
Number of iterations of update rules	200
Initial values of $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$	Uniform random values in the range (0,1)
Initial values of $\boldsymbol{W}_i$	Identity matrix

3.2 ILRMA based on GGD

3.2.1 *Generative model and objective function in GGD-ILRMA*

In GGD-ILRMA, we assume the isotropic complex GGD as the source generative model, which is independently defined in each time-frequency slot as follows:

$$\begin{aligned} p(\boldsymbol{Y}_n) &= \prod_{i,j} p(y_{ij,n}) \\ &= \prod_{i,j} \frac{\beta}{2\pi\sigma_{ij,n}^2 \Gamma\left(\frac{2}{\beta}\right)} \exp\left[-\left(\frac{|y_{ij,n}|}{\sigma_{ij,n}}\right)^{\beta}\right], \qquad (28) \\ \sigma_{ij,n}^{p} &= \sum_k t_{ik,n} v_{kj,n}, \qquad (29) \end{aligned}$$

where $\sigma_{ij,n}$ is the time-frequency-varying scale parameter, $\Gamma(\cdot)$ is a gamma function, and p is the domain parameter in the NMF modeling. The distribution (28) is depicted in Fig. 4a. The p.d.f. becomes identical to (9) when $\beta = 2$. For $\beta = 1$, (28) corresponds to the complex Laplace distribution. Similar to (9), the probability of (28) only depends on $|y_{ij,n}|$, and the phase of $y_{ij,n}$ is uniformly distributed. From (28), the objective function in GGD-ILRMA can be obtained as follows by assuming independence between sources:

$$\begin{aligned} \mathcal{L}_{\mathrm{GGD}} =& -2J \sum_i \log|\det \boldsymbol{W}_i| \\ &+ \sum_{i,j,n} \left[\frac{|y_{ij,n}|^{\beta}}{\left(\sum_k t_{ik,n} v_{kj,n}\right)^{\frac{\beta}{p}}} + \frac{2}{p} \log\left(\sum_k t_{ik,n} v_{kj,n}\right) \right] \\ &+ IJN \log \frac{2\pi\Gamma\left(\frac{2}{\beta}\right)}{\beta}. \qquad (30) \end{aligned}$$

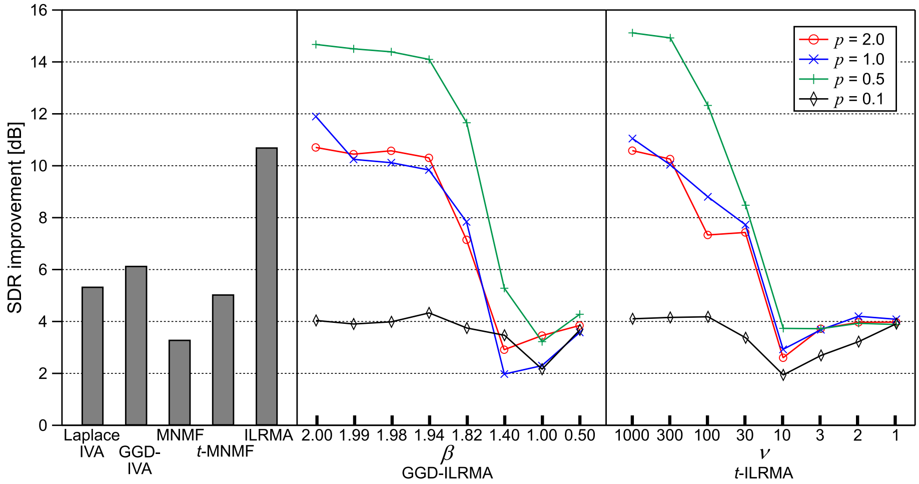

Fig. 7 Average SDR improvements of Music 1 signal with `E2A_1` using fixed parameters. The results for GGD-IVA and *t*-MNMF using $\beta = 0.5$ and $\nu = 1$ are shown, which are their best parameter settings

It is obvious that GGD-ILRMA (30) coincides with the conventional ILRMA (13) when $\beta = p = 2$.

3.2.2 Derivation of update rules for GGD-ILRMA

First, we derive the iterative update rules for obtaining $\boldsymbol{W}_i$ that optimizes (30). Since it is difficult to directly calculate the partial derivative of (30) w.r.t. $\boldsymbol{w}_{i,n}$, we use an MM algorithm, i.e., we minimize the majorization function (upper-bound function) instead of the original objective function. This approach can indirectly minimize the original function (30). Unlike the conventional ILRMA (13), GGD-ILRMA (30) includes the term $|y_{ij,n}|^{\beta} = \left|\boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{x}_{ij}\right|^{\beta}$. If we bound this term by $|y_{ij,n}|^2$, the MM-algorithm-based efficient optimization, IP, can be used for GGD-ILRMA because the objective function becomes identical to the conventional ILRMA (13) w.r.t. $\boldsymbol{w}_{i,n}$. To achieve this, we use the following inequality:

$$|y_{ij,n}|^{\beta} \leq \frac{\beta}{2\gamma_{ij,n}^{2-\beta}}|y_{ij,n}|^2 + \left(1 - \frac{\beta}{2}\right)\gamma_{ij,n}^{\beta} \qquad (31)$$

to design a majorization function of (30), where $\gamma_{ij,n} > 0$ is an auxiliary variable and the equality of (31) holds if and only if

$$\gamma_{ij,n} = |y_{ij,n}|. \qquad (32)$$

Note that the inequality (31) holds only for $0 < \beta < 2$, and the other values of β are beyond the scope of this

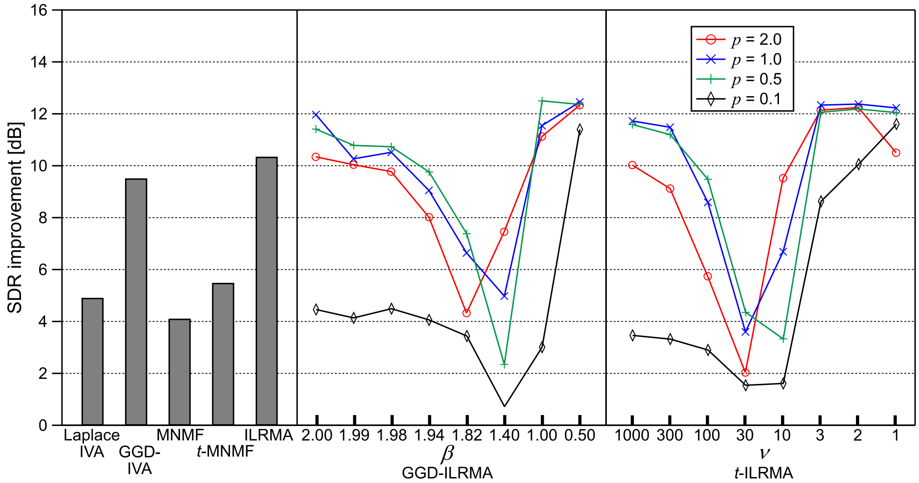

Fig. 8 Average SDR improvements of Music 4 with `E2A_1` signal using fixed parameters. The results for GGD-IVA and *t*-MNMF using $\beta = 0.5$ and $\nu = 30$ are shown, which are their best parameter settings

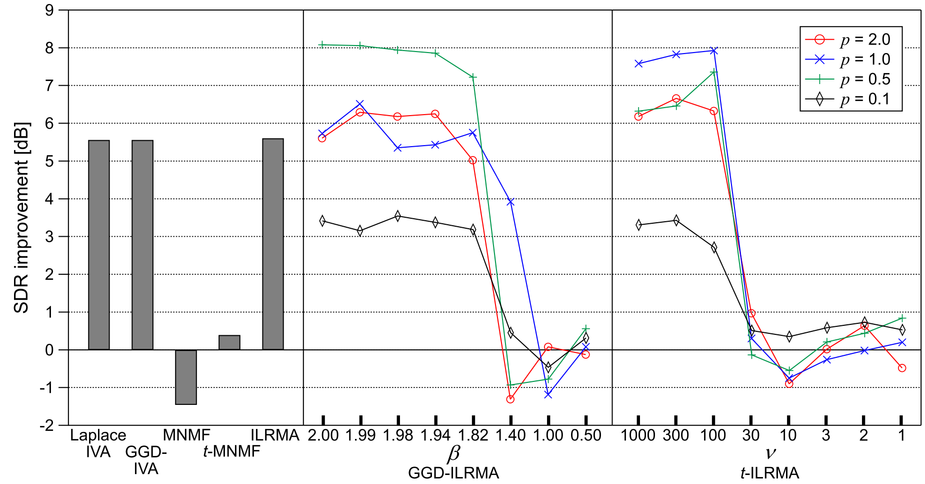

Fig. 9 Average SDR improvements of Speech 2 with `E2A_1` signal using fixed parameters. The results for GGD-IVA and *t*-MNMF using $\beta = 1.0$ and $\nu = 1$ are shown, which are their best parameter settings

paper. By applying (31) to (30), the majorization function of (30) can be designed as

$$\begin{aligned}\mathcal{L}_{\mathrm{GGD}} \le & -2J\sum_i \log|\det \boldsymbol{W}_i| \\ & + \sum_{i,j,n}\left[\frac{\beta|y_{ij,n}|^2}{2\gamma_{ij,n}^{2-\beta}\left(\sum_k t_{ik,n}v_{kj,n}\right)^{\frac{\beta}{p}}} + \frac{(2-\beta)\,\gamma_{ij,n}^{\beta}}{2\left(\sum_k t_{ik,n}v_{kj,n}\right)^{\frac{\beta}{p}}} \right. \\ & \left. + \frac{2}{p}\log\left(\sum_k t_{ik,n}v_{kj,n}\right)\right] + IJN\log\frac{2\pi\Gamma\left(\frac{2}{\beta}\right)}{\beta} \\ = & -2J\sum_i \log|\det \boldsymbol{W}_i| + J\sum_{i,n}\boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{G}_{i,n}\boldsymbol{w}_{i,n} + \mathcal{C}_1,\end{aligned} \tag{33}$$

$$\boldsymbol{G}_{i,n} = \frac{\beta}{2J}\sum_j \frac{1}{\gamma_{ij,n}^{2-\beta}\left(\sum_k t_{ik,n}v_{kj,n}\right)^{\frac{\beta}{p}}}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}}, \tag{34}$$

where $\mathcal{C}_1$ includes the constant terms that do not depend on $\boldsymbol{w}_{i,n}$. Since (33) has the same form as the conventional ILRMA (14) w.r.t. $\boldsymbol{w}_{i,n}$, we can apply IP to the majorization function (33). The update rules for $\boldsymbol{w}_{i,n}$ are derived as (34) with (32) and (17)–(19), where (34) coincides with (16) when $\beta = p = 2$.

Next, we derive the update rules for $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$. They can also be derived by designing a majorization function and applying the MM algorithm. Since the term $\left(\sum_k t_{ik,n}v_{kj,n}\right)^{-\frac{\beta}{p}}$ in (30) is always convex for all values of

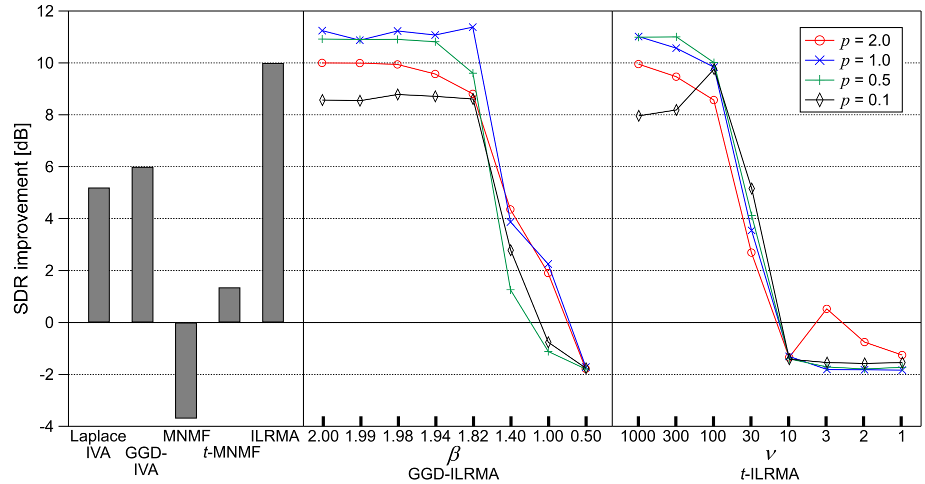

Fig. 10 Average SDR improvements of Speech 4 with `E2A_1` signal using fixed parameters. The results for GGD-IVA and *t*-MNMF using $\beta = 0.5$ and $\nu = 1$ are shown, which are their best parameter settings

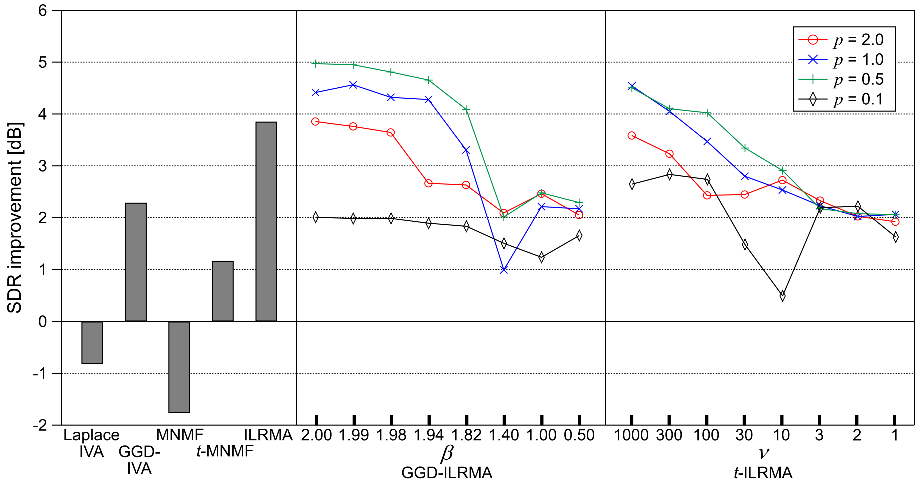

Fig. 11 Average SDR improvements of Music 2 signal with `E2A_2` using fixed parameters. The results for GGD-IVA and *t*-MNMF using $\beta = 0.5$ and $\nu = 1$ are shown, which are their best parameter settings

$\beta > 0$ and $p > 0$, we can bound this term using Jensen's inequality as

$$\left(\sum_k t_{ik,n} v_{kj,n}\right)^{-\frac{\beta}{p}} = \sum_k \left(\delta_{ij,nk} \frac{t_{ik,n} v_{kj,n}}{\delta_{ij,nk}}\right)^{-\frac{\beta}{p}}$$
$$\leq \sum_k \delta_{ij,nk} \left(\frac{t_{ik,n} v_{kj,n}}{\delta_{ij,nk}}\right)^{-\frac{\beta}{p}}, \qquad (35)$$

where $\delta_{ij,nk} > 0$ is an auxiliary variable that satisfies $\sum_k \delta_{ij,nk} = 1$. Also, the term $\log \sum_k t_{ik,n} v_{kj,n}$ in (30) can be bounded by the tangent-line inequality as

$$\log \sum_k t_{ik,n} v_{kj,n} \leq \frac{1}{\epsilon_{ij,n}} \left(\sum_k t_{ik,n} v_{kj,n} - 1\right)$$
$$+ \log \epsilon_{ij,n}, \qquad (36)$$

where $\epsilon_{ij,n} > 0$ is an auxiliary variable. The equalities of (35) and (36) hold if and only if

$$\delta_{ij,nk} = \frac{t_{ik,n} v_{kj,n}}{\sum_{k'} t_{ik',n} v_{k'j,n}}, \qquad (37)$$

$$\epsilon_{ij,n} = \sum_k t_{ik,n} v_{kj,n}, \qquad (38)$$

respectively. By applying (35) and (36) to (30), the majorization function of (30) can be designed as

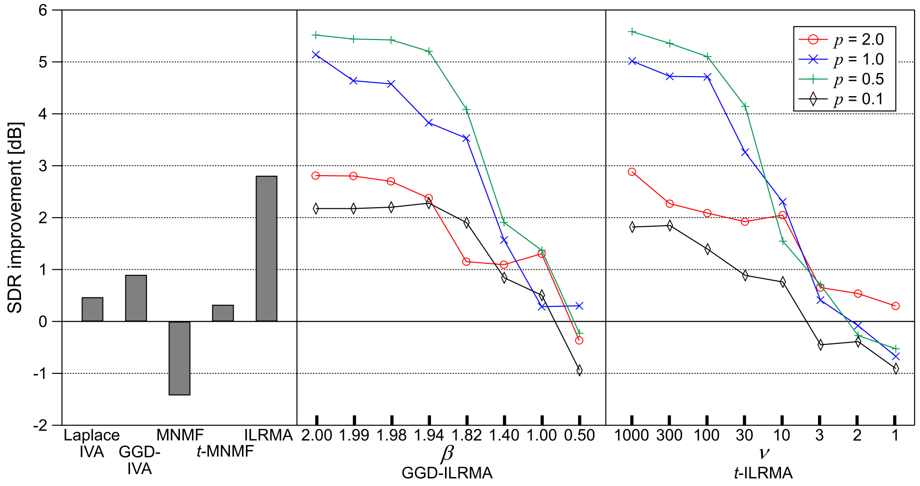

Fig. 12 Average SDR improvements of Music 3 with `E2A_2` signal using fixed parameters. The results for GGD-IVA and *t*-MNMF using $\beta = 0.5$ and $\nu = 30$ are shown, which are their best parameter settings

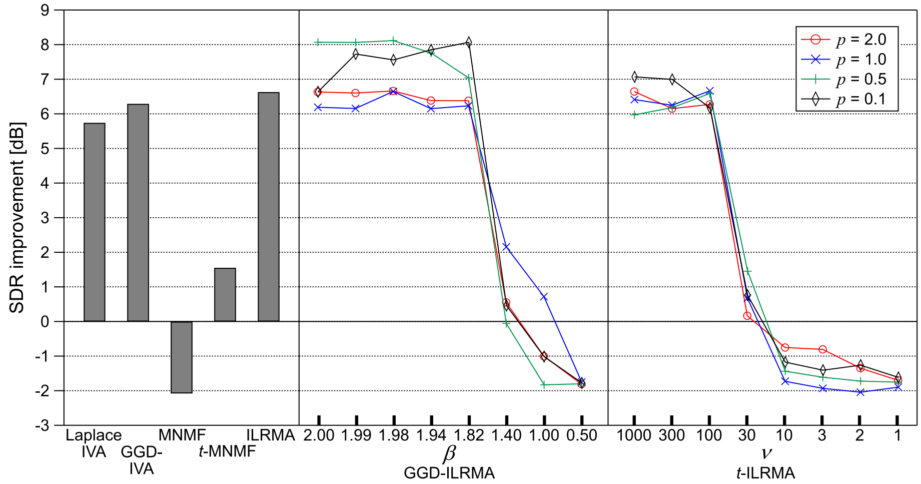

Fig. 13 Average SDR improvements of Speech 1 with `E2A_2` signal using fixed parameters. The results for GGD-IVA and *t*-MNMF using $\beta = 1.0$ and $\nu = 1$ are shown, which are their best parameter settings

$$\begin{aligned}\mathcal{L}_{\mathrm{GGD}} \leq & -2J\sum_i \log|\det \boldsymbol{W}_i| \\ & + \sum_{i,j,n}\left[\sum_k \frac{\delta_{ij,nk}^{\frac{\beta}{p}+1}|y_{ij,n}|^{\beta}}{\left(zt_{ik,n}v_{kj,n}\right)^{\frac{\beta}{p}}} + \frac{2}{p\epsilon_{ij,n}}\left(\sum_k t_{ik,n}v_{kj,n}-1\right)+\frac{2}{p}\log\epsilon_{ij,n}\right] \\ & + IJN\log\frac{2\pi\Gamma\left(\frac{2}{\beta}\right)}{\beta} \\ = & \sum_{i,j,n}\left[\sum_k \frac{\delta_{ij,nk}^{\frac{\beta}{p}+1}|y_{ij,n}|^{\beta}}{\left(t_{ik,n}v_{kj,n}\right)^{\frac{\beta}{p}}} + \frac{2}{p\epsilon_{ij,n}}\sum_k t_{ik,n}v_{kj,n}\right] + \mathcal{C}_2, \end{aligned} \tag{39}$$

where $\mathcal{C}_2$ includes the constant terms that do not depend on $t_{ik,n}$ or $v_{kj,n}$. By setting the partial derivative of (39) w.r.t. $t_{ik,n}$ to zero, we have

$$\sum_j\left[-\frac{\beta}{p}\frac{\delta_{ij,nk}^{\frac{\beta}{p}+1}|y_{ij,n}|^{\beta}}{\left(t_{ik,n}v_{kj,n}\right)^{\frac{\beta}{p}+1}}v_{kj,n}+\frac{2}{p\epsilon_{ij,n}}v_{kj,n}\right]=0.$$

The solution of this equation is

$$t_{ik,n}=\left(\frac{\beta\sum_j \frac{\delta_{ij,nk}^{\frac{\beta}{p}+1}|y_{ij,n}|^{\beta}}{v_{kj,n}^{\frac{\beta}{p}+1}}v_{kj,n}}{2\sum_j\frac{1}{\epsilon_{ij,n}}v_{kj,n}}\right)^{\frac{p}{\beta+p}}. \tag{40}$$

Then, we can obtain the following update rule for $t_{ik,n}$ by substituting (37) and (38) into (40):

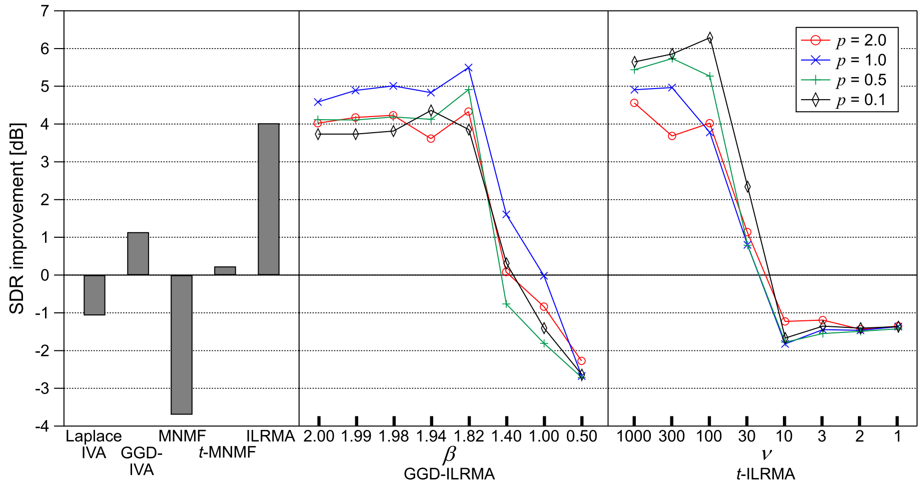

Fig. 14 Average SDR improvements of Speech 3 with `E2A_2` signal using fixed parameters. The results for GGD-IVA and *t*-MNMF using $\beta = 0.5$ and $\nu = 1$ are shown, which are their best parameter settings

Table 5 Overall average SDR improvements (dB) in two-source case for the best parameter settings

Source and impulse response	Laplace IVA	GGD-IVA	MNMF	*t*-MNMF	ILRMA	GGD-ILRMA	*t*-ILRMA
Music and `E2A_1`	2.41	3.11 ($\beta=0.5$)	2.42	3.30 ($\nu=1$)	6.24	7.52 ($\beta=1.94, p=0.5$)	7.61 ($\nu=1000, p=0.5$)
Speech and `E2A_1`	3.94	4.89 ($\beta=0.5$)	- 2.04	0.94 ($\nu=15$)	7.73	8.70 ($\beta=1.94, p=0.5$)	8.73 ($\nu=1000, p=1.0$)
Music and `E2A_2`	1.97	2.19 ($\beta=0.5$)	- 2.25	- 0.03 ($\nu=1$)	4.97	6.30 ($\beta=1.98, p=0.5$)	6.39 ($\nu=1000, p=0.5$)
Speech and `E2A_2`	3.76	4.63 ($\beta=0.5$)	- 3.41	0.79 ($\nu=15$)	5.76	6.36 ($\beta=1.94, p=0.5$)	6.17 ($\nu=1000, p=1.0$)

$$t_{ik,n} \leftarrow t_{ik,n} \left[\frac{\beta \sum_j \frac{|y_{ij,n}|^\beta}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^{\frac{\beta}{p}+1}} v_{kj,n}}{2 \sum_j \frac{1}{\sum_{k'} t_{ik',n} v_{k'j,n}} v_{kj,n}} \right]^{\frac{p}{\beta+p}}. \qquad (41)$$

Similar to (41), we can obtain the update rules for $v_{kj,n}$ as

$$v_{kj,n} \leftarrow v_{kj,n} \left[\frac{\beta \sum_i \frac{|y_{ij,n}|^\beta}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^{\frac{\beta}{p}+1}} t_{ik,n}}{2 \sum_i \frac{1}{\sum_{k'} t_{ik',n} v_{k'j,n}} t_{ik,n}} \right]^{\frac{p}{\beta+p}}. \qquad (42)$$

These algorithms can be interpreted as NMF based on the GGD (hereafter called GGD-NMF). Since the derivations of the update rules are based on the MM algorithm, they ensure the monotonic decrease in the objective function in each iteration.

3.3 ILRMA based on Student's *t* distribution

3.3.1 Generative model and objective function in t-ILRMA

In *t*-ILRMA, the isotropic complex Student's *t* distribution is independently assumed in each time-frequency slot as the following source generative model:

$$\begin{aligned} p(\boldsymbol{Y}_n) &= \prod_{i,j} p(y_{ij,n}) \\ &= \prod_{i,j} \frac{1}{\pi \sigma_{ij,n}^2} \left(1 + \frac{2}{\nu} \frac{|y_{ijn}|^2}{\sigma_{ij,n}^2} \right)^{-\frac{2+\nu}{2}}, \end{aligned} \qquad (43)$$

where $\nu > 0$ is the degree-of-freedom parameter that controls the super-Gaussianity of Student's t distribution and $\sigma_{ij,n}$ is defined as (29). The distribution (43) is depicted in Fig. 4b. Similar to (28), this p.d.f. also becomes identical to (9) when $\nu \rightarrow \infty$, and the probability of (28) does not depend on the phase of $y_{ij,n}$. For $\nu = 1$, (43) corresponds to the complex Cauchy distribution. The objective function of t-ILRMA can be obtained from (43) as

$$\begin{aligned} \mathcal{L}_t = & -2J \sum_i \log |\det \boldsymbol{W}_i| \\ & + \sum_{i,j,n} \left\{ \left(1 + \frac{\nu}{2}\right) \log \left[1 + \frac{2}{\nu} \frac{|y_{ij,n}|^2}{\left(\sum_k t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}}} \right] \right. \\ & \left. + \frac{2}{p} \log \left(\sum_k t_{ik,n} v_{kj,n} \right) \right\} + IJN \log \pi. \end{aligned} \qquad (44)$$

When $\nu \rightarrow \infty$ and $p=2$, (44) coincides with (13).

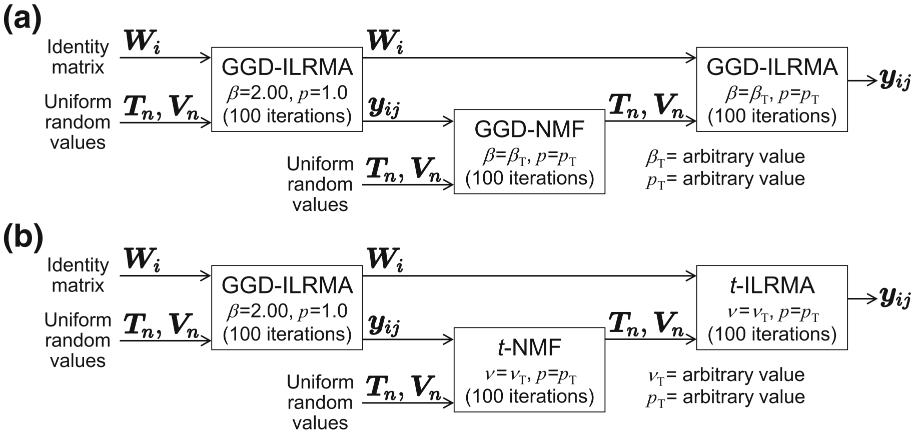

Fig. 15 Tempering approach for **a** GGD- and **b** *t*-ILRMA. The first half of the optimization is based on GGD-ILRMA with $\beta = 2$ and $p = 1$, and the second half is the proposed method with arbitrary parameters, where the NMF source model is retrained using the temporary estimated signal after the first-half optimization

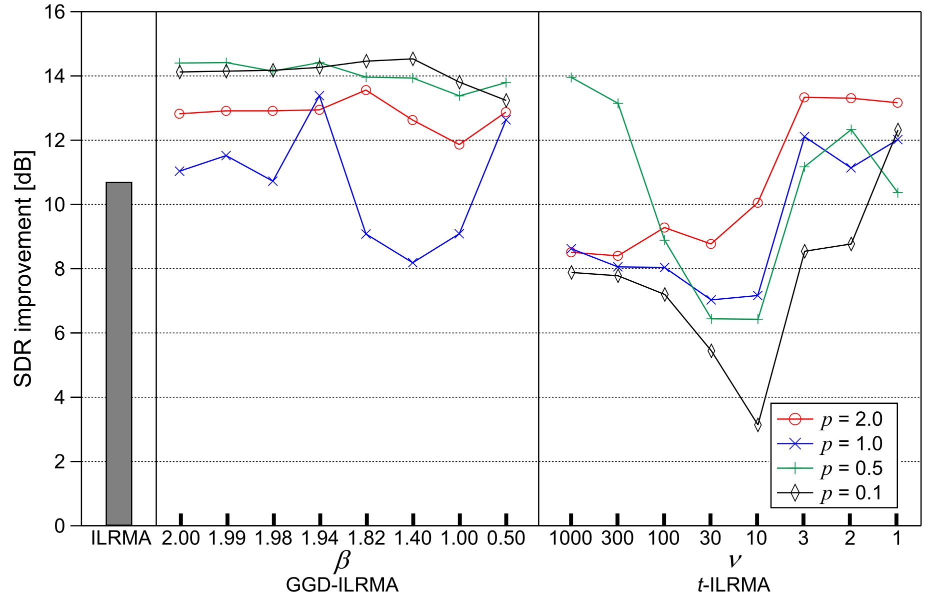

Fig. 16 Average SDR improvements of Music 1 with `E2A_1` signal using parameter tempering

3.3.2 Derivation of update rules for t-ILRMA

Similarly in Section 3.2.2, we first derive the iterative update rules for $\boldsymbol{W}_i$ that optimizes (44) using the MM algorithm. In the case of t-ILRMA, the objective function (44) includes the term $|y_{ij,n}|^2 = \left|\boldsymbol{w}_{i,n}^{\mathrm{H}}\boldsymbol{x}_{ij}\right|^2$ inside of the logarithm function. Therefore, we bound this term by a linear function of $|y_{ij,n}|^2$ using the following tangent-line inequality:

$$\log\left(1+\frac{2}{\nu}\frac{|y_{ij,n}|^2}{\left(\sum_k t_{ik,n}v_{kj,n}\right)^{\frac{2}{p}}}\right) \le \frac{1}{\zeta_{ij,n}}\left[1+\frac{2}{\nu}\frac{|y_{ij,n}|^2}{\left(\sum_k t_{ik,n}v_{kj,n}\right)^{\frac{2}{p}}}-\zeta_{ij,n}\right] + \log\zeta_{ij,n}, \tag{45}$$

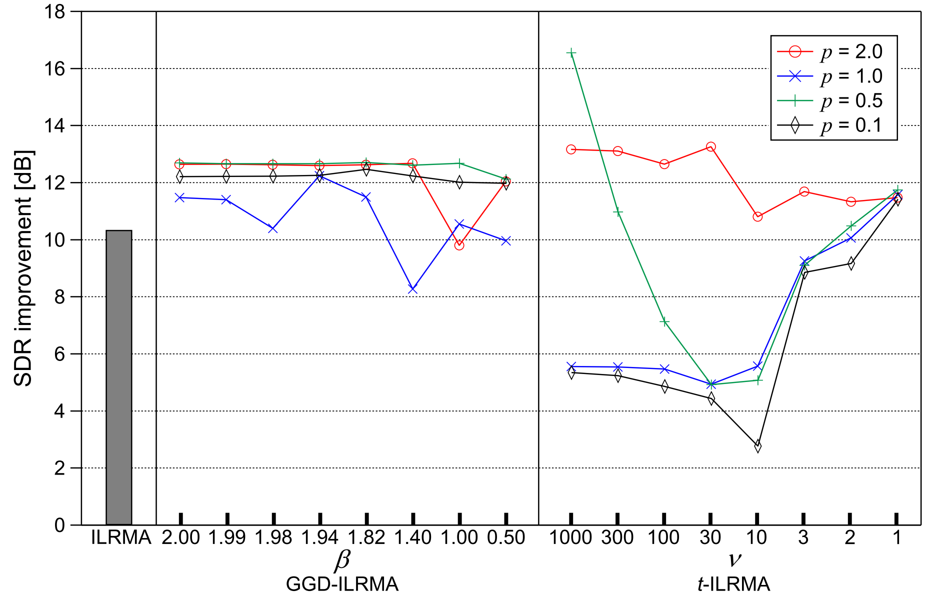

Fig. 17 Average SDR improvements of Music 4 with `E2A_1` signal using parameter tempering

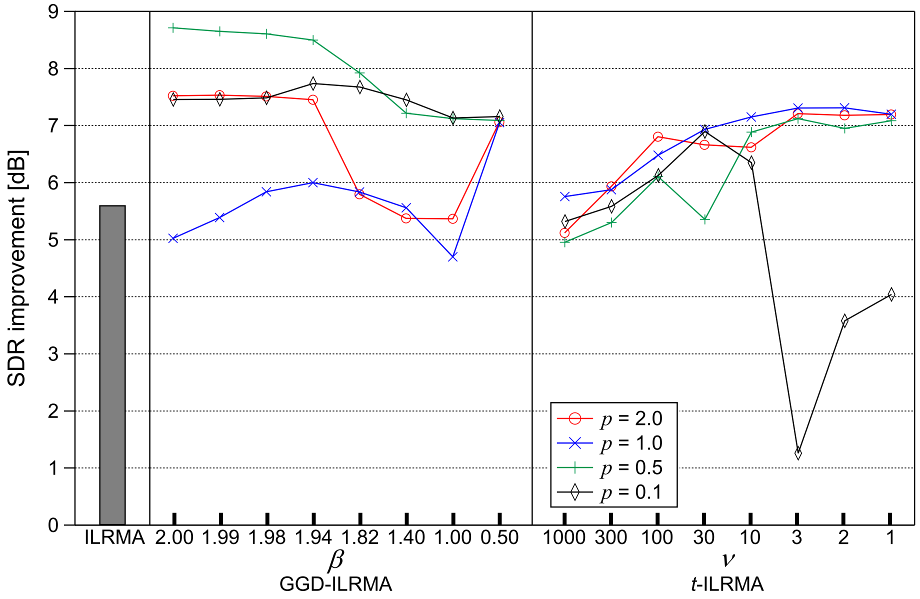

Fig. 18 Average SDR improvements of Speech 2 with `E2A_1` signal using parameter tempering

where $\zeta_{ij,n} > 0$ is an auxiliary variable and the equality of (45) holds if and only if

$$\zeta_{ij,n} = 1 + \frac{2}{\nu} \frac{|y_{ij,n}|^2}{\left(\sum_k t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}}}. \tag{46}$$

By applying (45) to (44), the majorization function of (44) can be designed as

$$\begin{aligned}\mathcal{L}_t \le & -2J \sum_i \log|\det \boldsymbol{W}_i| \\ & + \sum_{i,j,n} \left\{ \left(1+\frac{\nu}{2}\right) \frac{1}{\zeta_{ij,n}} \left[1 + \frac{2}{\nu} \frac{|y_{ij,n}|^2}{\left(\sum_k t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}}} - \zeta_{ij,n} \right] \right. \\ & \left. + \left(1+\frac{\nu}{2}\right) \log \zeta_{ij,n} + \frac{2}{p} \log \left(\sum_k t_{ik,n} v_{kj,n} \right) \right\} \\ & + IJN \log \pi \end{aligned} \tag{47}$$

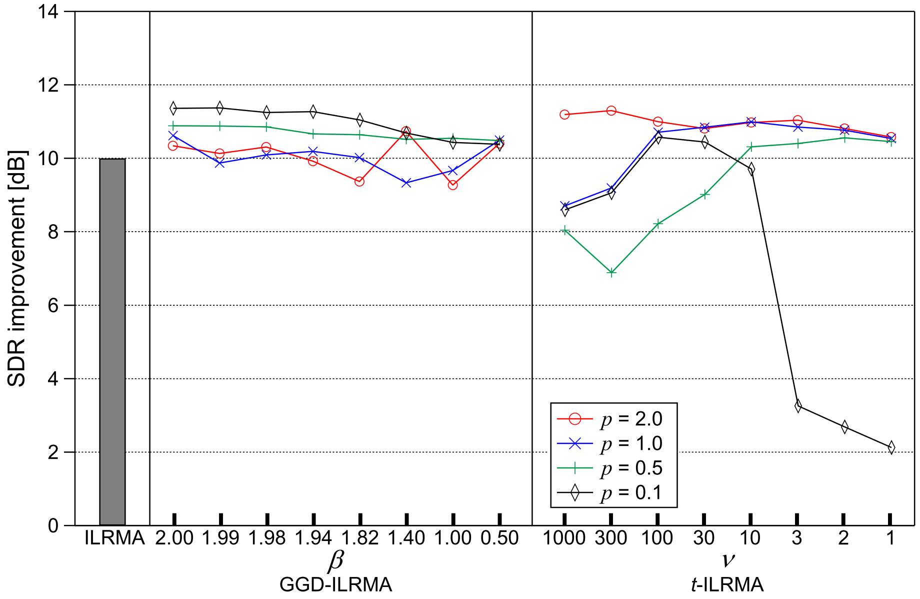

Fig. 19 Average SDR improvements of Speech 4 with `E2A_1` signal using parameter tempering

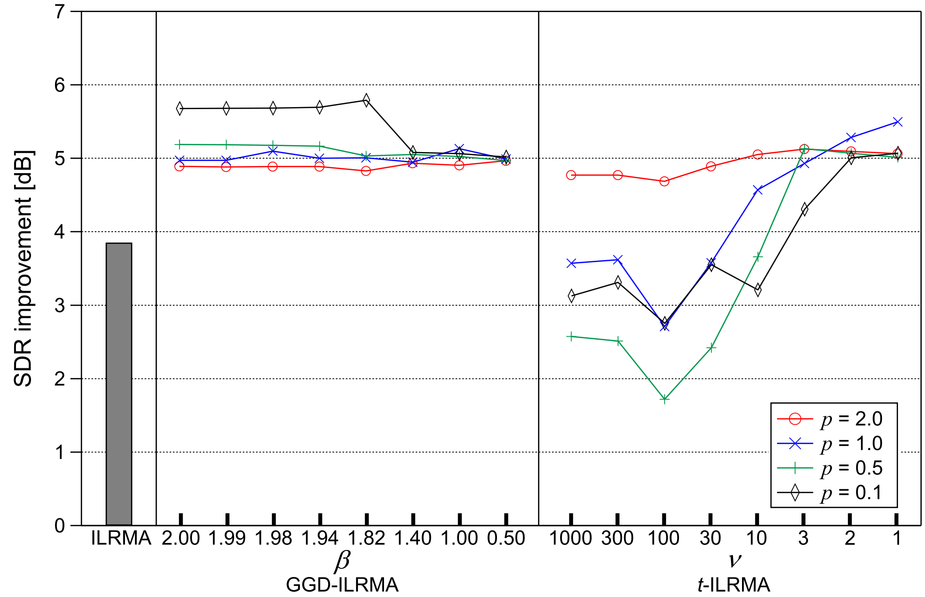

Fig. 20 Average SDR improvements of Music 2 with `E2A_2` signal using parameter tempering

$$
\begin{aligned}
&= -2J\sum_i \log|\det \boldsymbol{W}_i| \\
&\quad + J\sum_{i,n} \boldsymbol{w}_{i,n}^{\mathrm{H}} \boldsymbol{H}_{i,n} \boldsymbol{w}_{i,n} + \mathcal{C}_3 \\
&\equiv \mathcal{L}_t^{+},
\end{aligned}
\tag{48}
$$

$$
\boldsymbol{H}_{i,n} = \frac{1}{J}\left(\frac{2}{\nu}+1\right)\sum_j \frac{1}{\zeta_{ij,n}\left(\sum_k t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}}} \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}},
\tag{49}
$$

where $\mathcal{C}_3$ includes the constant terms that do not depend on $\boldsymbol{w}_{i,n}$. Since (48) has the same form as the conventional ILRMA (14) w.r.t. $\boldsymbol{w}_{i,n}$, we can apply IP to the majorization function (48). The update rules for $\boldsymbol{w}_{i,n}$ are derived as (49) with (46) and (17)–(19), where (49) coincides with (16) when $\nu \rightarrow \infty$ and $p = 2$.

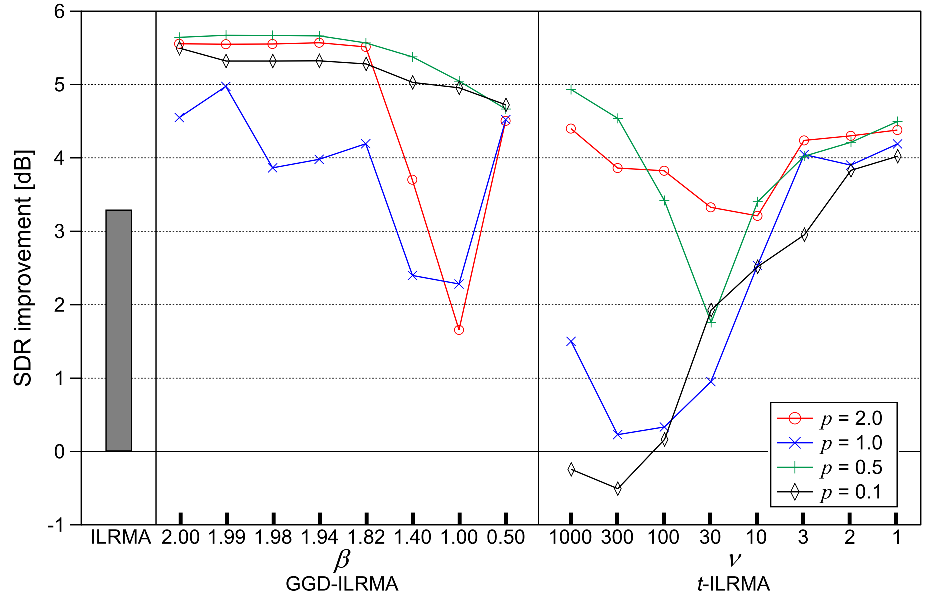

Fig. 21 Average SDR improvements of Music 3 with `E2A_2` signal using parameter tempering

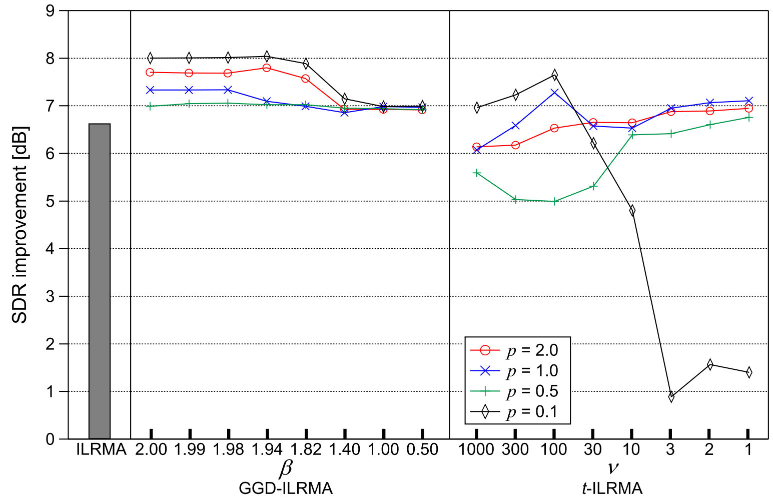

Fig. 22 Average SDR improvements of Speech 1 with `E2A_2` signal using parameter tempering

Next, we derive the update rules for $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$ in the same manner as for GGD-NMF. Since the term $\left(\sum_k t_{ik,n}v_{kj,n}\right)^{-\frac{2}{p}}$ in (47) is always convex for any value of p, we can bound this term in the same manner as (35), i.e.,

$$\left(\sum_k t_{ik,n}v_{kj,n}\right)^{-\frac{2}{p}} \leq \sum_k \eta_{ij,nk}\left(\frac{t_{ik,n}v_{kj,n}}{\eta_{ij,nk}}\right)^{-\frac{2}{p}}, \tag{50}$$

where $\eta_{ij,nk} > 0$ is an auxiliary variable that satisfies $\sum_k \eta_{ij,nk} = 1$. The equality of (50) holds if and only if

$$\eta_{ij,nk} = \frac{t_{ik,n}v_{kj,n}}{\sum_{k'} t_{ik',n}v_{k'j,n}}. \tag{51}$$

Also, the term $\log \sum_k t_{ik,n}v_{kj,n}$ in (47) can be bounded by (36). By applying (50) and (36) to (47), we can design a further majorization function of (47) as follows:

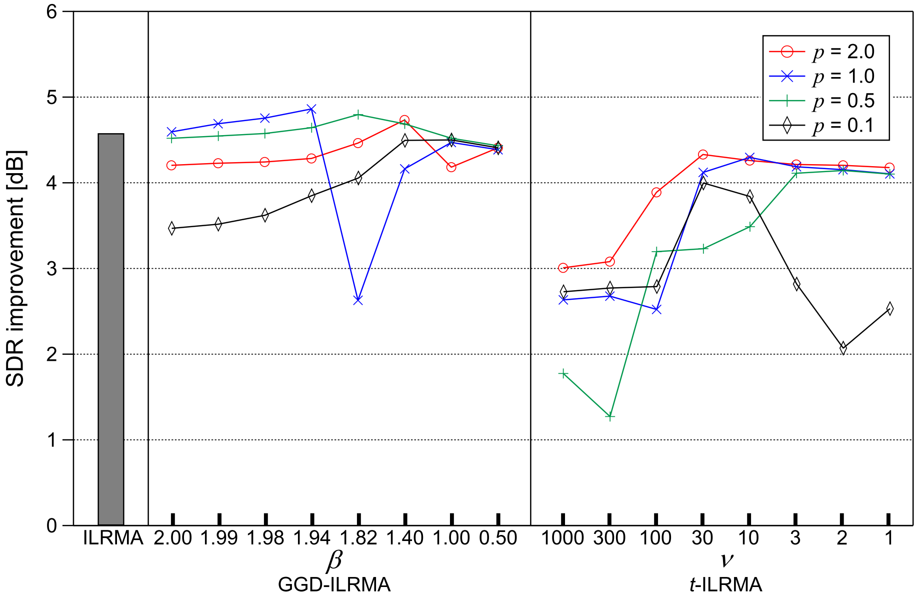

Fig. 23 Average SDR improvements of Speech 3 with `E2A_2` signal using parameter tempering

Table 6 Overall average SDR improvements (dB) in two-source case employing parameter tempering for the best parameter settings

Source and impulse response	ILRMA	GGD-ILRMA	*t*-ILRMA
Music and `E2A_1`	6.24	7.66 ($\beta=1.99, p=0.5$)	7.47 ($\nu=1000, p=0.5$)
Speech and `E2A_1`	7.73	9.09 ($\beta=1.94, p=0.5$)	8.61 ($\nu=3, p=1.0$)
Music and `E2A_2`	5.22	6.87 ($\beta=1.94, p=0.5$)	6.81 ($\nu=1000, p=0.5$)
Speech and `E2A_2`	6.09	6.45 ($\beta=1.98, p=0.5$)	6.05 ($\nu=30, p=2.0$)

$$\begin{aligned}\mathcal{L}_t^+ \le& -2J\sum_i \log|\det \boldsymbol{W}_i|\\ &+\sum_{i,j,n}\Biggl\{\left(1+\frac{\nu}{2}\right)\frac{1}{\zeta_{ij,n}}\left[1+\frac{2}{\nu}\sum_k \frac{\eta_{ij,nk}^{\frac{2}{p}+1}|y_{ij,n}|^2}{\left(t_{ik,n}v_{kj,n}\right)^{\frac{2}{p}}}-\zeta_{ij,n}\right]\\ &+\left(1+\frac{\nu}{2}\right)\log\zeta_{ij,n}+\frac{2}{p\epsilon_{ij,n}}\left(\sum_k t_{ik,n}v_{kj,n}-\epsilon_{ij,n}\right)\\ &+\frac{2}{p}\log\epsilon_{ij,n}\Biggr\}+IJN\log\pi\\ r=&+\sum_{i,j,n}\left[\left(\frac{2}{\nu}+1\right)\sum_k\frac{\eta_{ij,nk}^{\frac{2}{p}+1}|y_{ij,n}|^2}{\zeta_{ij,n}\left(t_{ik,n}v_{kj,n}\right)^{\frac{2}{p}}}\right.\\ &\left.+\frac{2}{p\epsilon_{ij,n}}\sum_k t_{ik,n}v_{kj,n}\right]+\mathcal{C}_4,\end{aligned} \tag{52}$$

where $\mathcal{C}_4$ includes the constant terms that do not depend on $t_{ik,n}$ or $v_{kj,n}$. By setting the partial derivative of (52) w.r.t. $t_{ik,n}$ to zero, we have

$$\sum_j\left[-\frac{2}{p}\left(\frac{2}{\nu}+1\right)\frac{\eta_{ij,nk}^{\frac{2}{p}+1}|y_{ij,n}|^2}{\zeta_{ij,n}\left(t_{ik,n}v_{kj,n}\right)^{\frac{2}{p}+1}}v_{kj,n}+\frac{2}{p\epsilon_{ij,n}}v_{kj,n}\right]=0. \tag{53}$$

The solution of this equation is obtained as

$$t_{ik,n}=\left[\frac{\left(\frac{2}{\nu}+1\right)\sum_j\frac{\eta_{ij,nk}^{\frac{2}{p}+1}|y_{ij,n}|^2}{\zeta_{ij,n}v_{kj,n}^{\frac{2}{p}+1}}v_{kj,n}}{\sum_j\frac{1}{\epsilon_{ij,n}}v_{kj,n}}\right]^{\frac{p}{p+2}}. \tag{54}$$

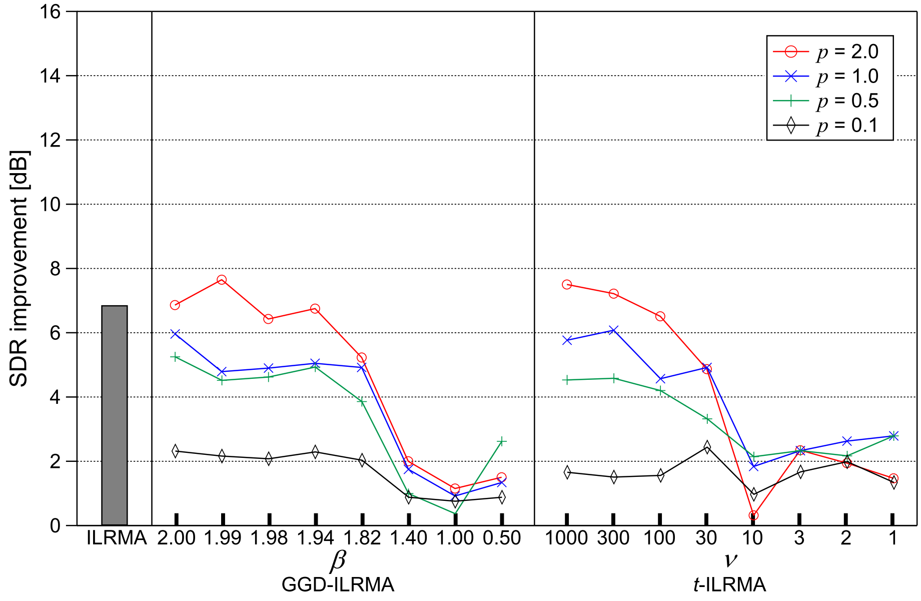

Fig. 24 Average SDR improvements of shorter Music 1 signal (2.5 s) with `E2A_1` using fixed parameters

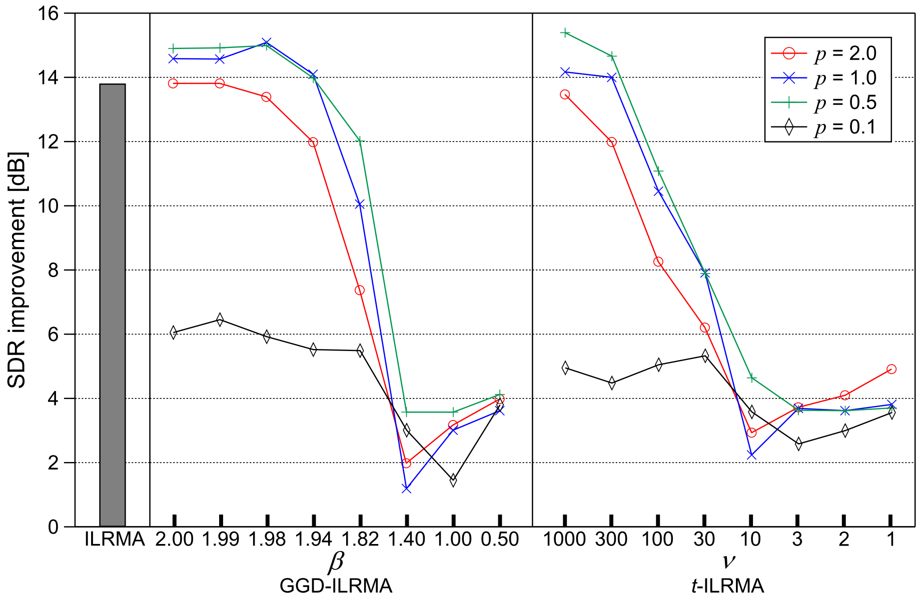

Fig. 25 Average SDR improvements of longer Music 1 signal (10.0 s) with `E2A_1` using fixed parameters

Then, we can obtain the following update rule for $t_{ik,n}$ by substituting (51) and (38) into (54):

$$t_{ik,n} \leftarrow t_{ik,n} \left(\frac{\sum_j \frac{|y_{ij,n}|^2}{b_{ij,n} \sum_{k'} t_{ik',n} v_{k'j,n}} v_{kj,n}}{\sum_j \frac{1}{\sum_{k'} t_{ik',n} v_{k'j,n}} v_{kj,n}} \right)^{\frac{p}{p+2}}, \tag{55}$$

where

$$b_{ij,n} = \frac{\nu}{\nu+2} \left(\sum_{k'} t_{ik',n} v_{k'j,n} \right)^{\frac{2}{p}} + \frac{2}{\nu+2} |y_{ij,n}|^2. \tag{56}$$

Similar to (55), we can obtain the update rules for $v_{kj,n}$ as

$$v_{kj,n} \leftarrow v_{kj,n} \left(\frac{\sum_i \frac{|y_{ij,n}|^2}{b_{ij,n} \sum_{k'} t_{ik',n} v_{k'j,n}} t_{ik,n}}{\sum_i \frac{1}{\sum_{k'} t_{ik',n} v_{k'j,n}} t_{ik,n}} \right)^{\frac{p}{p+2}}. \tag{57}$$

These update rules are similar to those in *t*-NMF [39], but they include the new domain parameter *p*. Similar to GGD-ILRMA, all the derivations of the update rules are based on the MM algorithm, thus ensuring their theoretical convergence.

3.4 Relationship between GGD- and *t*-ILRMA

The update rules for $t_{ik,n}$ and $v_{kj,n}$ (the GGD- and *t*-NMF parts in GGD- and *t*-ILRMA, respectively) have an interesting relationship. To clarify this issue, we here interpret these two NMF models in relation to the IS-NMF used in the original ILRMA. In GGD- and *t*-NMF, we introduced a new parameter *p* that determines the signal domain of the low-rank modeling, whereas IS-NMF is typically applied to the observed power spectrogram ($p = 2$) [29]. To fill the gap in the formulation between IS-NMF and GGD- or *t*-NMF, we use the following generalized version of the update rules for IS-NMF:

Table 7 Overall average SDR improvements (dB) in two-source case with various signal lengths for the best parameter settings

Source and signal length	ILRMA	GGD-ILRMA	*t*-ILRMA
Music (2.5 s, short)	3.38	3.43 ($\beta = 1.99, p = 2.0$)	3.51 ($\nu = 2, p = 1.0$)
Music (5.0 s, original)	6.24	7.52 ($\beta = 1.94, p = 0.5$)	7.61 ($\nu = 1000, p = 0.5$)
Music (10.0 s, long)	7.29	8.83 ($\beta = 2.00, p = 1.0$)	8.92 ($\nu = 1000, p = 0.5$)
Speech (5.0 s, short)	7.26	7.69 ($\beta = 1.98, p = 0.5$)	8.33 ($\nu = 1000, p = 1.0$)
Speech (10.0 s, original)	7.73	8.70 ($\beta = 1.94, p = 0.5$)	8.73 ($\nu = 1000, p = 1.0$)
Speech (20.0 s, long)	8.05	8.41 ($\beta = 1.94, p = 1.0$)	8.29 ($\nu = 300, p = 1.0$)

Table 8 Dry sources used in three-source case

Signal	Data name	Sources (1/2/3)	Signal lengths [s]
Music 1	Melody 2/Midrange/Bass	Clarinet/Piano/Cello	5.0
Music 2	Melody 1/Melody 2/Bass	Horn/Clarinet/Bassoon	5.0
Music 3	Melody 1/Midrange/Bass	Trumpet/Piano/Bassoon	5.0
Music 4	Melody 2/Midrange/Bass	Violin/Harpsichord/Bassoon	5.0
Speech 1	dev1_female4	src_1/src_2/src_3	10.0
Speech 2	dev1_female4	src_2/src_3/src_4	10.0
Speech 3	dev1_male4	src_1/src_2/src_3	10.0
Speech 4	dev1_male4	src_2/src_3/src_4	10.0

$$t_{ik,n} \leftarrow t_{ik,n}\left[\frac{\sum_j \frac{|y_{ij,n}|^2}{\left(\sum_k t_{ik,n}v_{kj,n}\right)^2} v_{kj,n}}{\sum_j \frac{1}{\sum_k t_{ik,n}v_{kj,n}} v_{kj,n}}\right]^b, \tag{58}$$

$$v_{kj,n} \leftarrow v_{kj,n}\left[\frac{\sum_i \frac{|y_{ij,n}|^2}{\left(\sum_k t_{ik,n}v_{kj,n}\right)^2} t_{ik,n}}{\sum_i \frac{1}{\sum_k t_{ik,n}v_{kj,n}} t_{ik,n}}\right]^b, \tag{59}$$

where b is a new exponent parameter. Note that (58) and (59) with $b = 0.5$ were originally derived on the basis of the MM algorithm [50], then the update rules with $b = 1$ were derived using the majorization-equalization (ME) algorithm [53]. Recently, we have proven that (58) and (59) with any value of b in the range $(0, 1]$ can be interpreted as valid update rules of IS-NMF, which are obtained by applying the parametric ME algorithm to the objective function in IS-NMF, and can be used for IS-NMF or ILRMA without losing the theoretical convergence [54]. This parameter b controls the optimization speed of the NMF variables $t_{ik,n}$ and $v_{kj,n}$, and $b = 1$ provides the fastest convergence in IS-NMF.

For GGD-NMF, (41) can be reformulated as

$$t_{ik,n} \leftarrow t_{ik,n}\left[\frac{\sum_j \frac{z_{ij,n}}{\left(\sum_{k'} t_{ik',n}v_{k'j,n}\right)^2} v_{kj,n}}{\sum_j \frac{1}{\sum_{k'} t_{ik',n}v_{k'j,n}} v_{kj,n}}\right]^{\frac{p}{\beta+p}}, \tag{60}$$

where

$$\begin{aligned} z_{ij,n} &= \frac{\beta}{2}|y_{ij,n}|^\beta \left(\sum_{k'} t_{ik',n}v_{k'j,n}\right)^{1-\frac{\beta}{p}} \\ &= \frac{\beta}{2}\left(|y_{ij,n}|^{\frac{\beta}{p}} \sigma_{ij,n}^{1-\frac{\beta}{p}}\right)^p. \end{aligned} \tag{61}$$

The update rule of GGD-NMF (60) corresponds to that of IS-NMF (58) by assuming the observed signal as (61), which is the "geometric mean" of the data $|y_{ij,n}|$ and the low-rank model $\sigma_{ij,n}$ with a ratio of β/p to $1 - (\beta/p)$. In contrast, for t-NMF, (55) can also be rewritten as

$$t_{ik,n} \leftarrow t_{ik,n}\left[\frac{\sum_j \frac{z_{ij,n}}{\left(\sum_{k'} t_{ik',n}v_{k'j,n}\right)^2} v_{kj,n}}{\sum_j \frac{1}{\sum_{k'} t_{ik',n}v_{k'j,n}} v_{kj,n}}\right]^{\frac{p}{p+2}}, \tag{62}$$

$$\begin{aligned} z_{ij,n} &= \left(\sum_{k'} t_{ik',n}v_{k'j,n}\right)^{1-\frac{2}{p}} \left[\frac{\nu}{\nu+2}|y_{ij,n}|^{-2} \right. \\ &\quad \left. + \frac{2}{\nu+2}\left(\sum_{k'} t_{ik',n}v_{k'j,n}\right)^{-\frac{2}{p}}\right]^{-1} \\ &= \sigma_{ij,n}^{p-2}\left(\frac{\nu}{\nu+2}|y_{ij,n}|^{-2} + \frac{2}{\nu+2}\sigma_{ij,n}^{-2}\right)^{-1}. \end{aligned} \tag{63}$$

As mentioned in [39], the update rule of t-NMF (62) corresponds to that of IS-NMF (58) by assuming the observed signal to be (63), which is the "harmonic mean" of $|y_{ij,n}|^2$ and $\sigma_{ij,n}^2$ with a ratio of ν to two. The same reformulation can be found for the variable $v_{kj,n}$.

These facts mean that both NMF algorithms approximate the virtual observation $z_{ij,n}$ by the low-rank model $\sigma_{ij,n}$ in the ISNMF sense. Since $z_{ij,n}$ consists of the

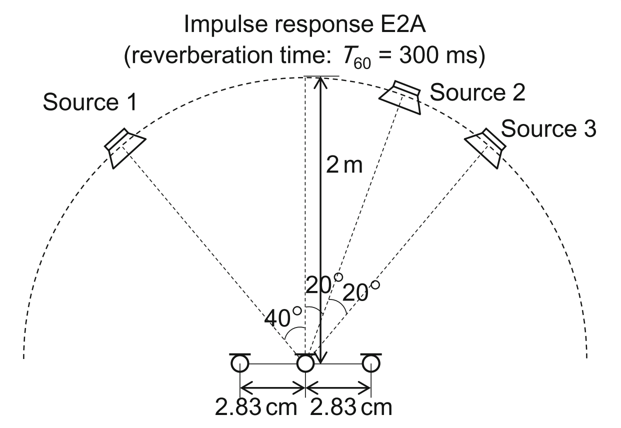

Fig. 26 Spatial arrangement of impulse responses used in three-source case

geometric or harmonic mean of the real observation $|y_{ij,n}|$ and the current low-rank model $\sigma_{ij,n}$, low-rankness of the estimated (updated) model $\sigma_{ij,n}$ tends to be more emphasized compared with the ISNMF decomposition using only the observation $|y_{ij,n}|$. In other words, the geometric or harmonic mean in $z_{ij,n}$ prevents $\sigma_{ij,n}$ from an overfitting to $|y_{ij,n}|$ by ignoring sparse outliers in $|y_{ij,n}|$, which enhances the low-rank decomposition. In (61) or (63), the shape parameter β or ν controls the intensity of such low-rank enhancement in NMF decomposition. However, intriguingly, the domain parameter p also affects the estimation of the low-rank model $\sigma_{ij,n}$. In GGD-NMF (61), by setting $p < \beta$, the geometric mean corresponds to the point externally dividing $|y_{ij,n}|$ and $\sigma_{ij,n}$, which mitigates the intensity of the low-rank enhancement mentioned above. Also, in t-NMF, $p < 2$ causes the same behavior because the term $\sigma_{ij,n}^{p-2}$ exists in (63), where the inverse of $\sigma_{ij,n}^{2-p}$ $(2 - p > 0)$ mitigates the low-rankness.

In summary, as shown in Table 1, smaller β and ν, which correspond to the sparse signal model, can inject the low-rank nature in GGD- and t-ILRMA, whereas a smaller p mitigates the property; the optimal balance among them will be discussed later on the basis of experimental evaluations. For ILRMA-based BSS, we can expect that such low-rank enhancement in NMF leads to the more accurate estimation of $\boldsymbol{W}_i$. This is because the estimation of the low-rank model $\sigma_{ij,n}$ becomes robust against outliers in the separated signal $|y_{ij,n}|$, and we can correctly capture the inherent spectral parts in the time-frequency structure of each source.

In addition, it is worth mentioning that the exponent value of the NMF update rules, b, is also important for ILRMA. It has been experimentally revealed that a smaller value of b is preferable for achieving better separation performance, although the optimized speed of $r_{ij,n}$ becomes slow. This may be to avoid trapping at a poor local minimum in the early and middle stages of the iteration in ILRMA because the optimization balance between $\boldsymbol{W}_i$ and $r_{ij,n}$ is significant for converging toward a better solution. In GGD- or t-ILRMA, the exponent value in (60) or (62) is defined as $p/(\beta + p)$ or $p/(p + 2)$, respectively. These values become small when p is small and β is large, which may result in a better separation result.

4 Results and discussion

To evaluate our proposed algorithms, we conducted some BSS experiments using music and speech mixtures. We first compared various conventional methods using observed signals in the case of two sources and two microphones. Then, we compared the conventional and proposed ILRMA in a more difficult situation with three sources and three microphones.

4.1 Dataset

We artificially produced monaural dry music sources of the four melody parts depicted in Fig. 5 using a YAMAHA MU-1000 PCM-based MIDI tone synthesizer, where several musical instruments were chosen to play these melody parts as shown in Table 2 [55]. The sources were selected to construct typical combinations of instruments with different melody parts (because the sources that simultaneously play the same melody are rare), where only the six combinations, Music 1–Music 6, were adopted for the sake of avoiding combinatorial explosion. For the speech signals, we used the monaural dry speech sources from the source separation task in SiSEC2011 [56] whose data names are `dev1_female4` and `dev1_male4` [57]. The detailed conditions of these speech signals are described in [56, 57].

4.2 BSS experiment with two sources

4.2.1 Conditions

In this experiment, we compared the seven methods shown in Fig. 1, namely, Laplace IVA (optimized by IP) [49], GGD-IVA (optimized by IP) [17], MNMF (based on a multivariate complex Gaussian distribution) [32], t-MNMF [40], ILRMA (based on a complex Gaussian distribution with a time-frequency-varying variance) [34], GGD-ILRMA, and t-ILRMA. The dry sources used in this experiment are shown in Table 3. To simulate a reverberant mixture, the mixture signals were produced by convoluting the impulse response E2A, which was obtained from the RWCP database [58], with two spatial arrangements, `E2A_1` and `E2A_2`. The recording conditions of the impulse responses in `E2A_1` and `E2A_2` are depicted in Fig. 6. The other conditions are shown in Table 4. As the evaluation score, we used the improvement of the signal-to-distortion ratio (SDR) [59], which indicates the overall separation quality.

Table 9 Overall average SDR improvements (dB) in three-source case for the best parameter settings

Source	ILRMA	GGD-ILRMA	t-ILRMA	GGD-ILRMA w/ tempering	t-ILRMA w/ tempering
Music	1.76	3.24 ($\beta = 1.94, p = 0.5$)	3.19 ($\nu = 300, p = 0.5$)	3.36 ($\beta = 1.82, p = 1.0$)	3.29 ($\nu = 1, p = 1.0$)
Speech	2.79	3.14 ($\beta = 1.94, p = 1.0$)	2.94 ($\nu = 1000, p = 1.0$)	3.32 ($\beta = 1.40, p = 0.5$)	3.22 ($\nu = 10, p = 2.0$)

4.2.2 Results using fixed parameters

Figures 7, 8, 9, and 10 show examples of the average SDR improvements for Music 1, Music 4, Speech 2, and Speech 4, respectively, with the `E2A_1` spatial arrangement. Ten trials with different random seeds were performed for all the methods. Note that conventional ILRMA and GGD-ILRMA with $\beta = p = 2$ are the same method. Also, for GGD-IVA and t-MNMF, the results are shown for the best parameter settings β and ν, as described in the caption of each figure. Similar to the `E2A_1` results, we show examples of results for Music 2, Music 3, Speech 1, and Speech 3 with the `E2A_2` spatial arrangement in Figs. 11, 12, 13, and 14, respectively. Table 5 indicates the overall average results for all music and speech signals with `E2A_1` and `E2A_2`, respectively, with the best parameter settings. From these results, we can confirm that the conventional and proposed ILRMA mostly outperform the other methods and that there are several settings of p and β or ν that outperform the conventional ILRMA based on the Gaussian distribution. In particular, the proposed methods with $p = 1.0$ or $p = 0.5$ often outperform the same methods with other values of p. However, regarding the parameters β and ν, smaller values produce poor separation results except for t-ILRMA in Fig. 8 (Music 4). This is because the NMF source model with the heavy-tailed distribution excessively enhances the low-rankness in the early stage of the iterative optimization, which can cause the serious problem of the sourcewise NMF model incorrectly capturing the spectrogram of the mixture signal by ignoring the important components for discriminating the sources, and the estimated signals become a distorted mixture signal and an artificial residue.

4.2.3 Results using parameter tempering

To solve the problem described in Section 4.2.2, we applied a tempering approach to the parameters in GGD- or t-ILRMA. The detailed tempering process is shown in Fig. 15. In the first half of the optimization, we perform GGD-ILRMA with $\beta = 2$ and $p = 1$. Then, the NMF source model $\boldsymbol{T}_n \boldsymbol{V}_n$ is retrained using a temporary estimated signal. After that, ILRMA with the desired distribution (desired parameters p_T and β_T or ν_T) is performed using the pretrained $\boldsymbol{W}_i$, $\boldsymbol{T}_n$, and $\boldsymbol{V}_n$. The intermediate NMF process is based on the same parameters (p_T and β_T or ν_T) as the subsequent ILRMA in the second half of the optimization. This can be considered as a binary tempering approach that avoids overfitting of the source model to the mixture signal. Note that we also attempted a more precise tempering approach involving continuously changing the parameters in every iteration, but the binary tempering approach shown in Fig. 15 achieved the most accurate and stablest results. The reason why we started from not $p = 2$ but $p = 1$ is that a small exponent value, $p/(\beta + p)$ in (41) and (42) or $p/(p + 2)$ in (55) and (57), in the NMF update rules provides better separation as revealed in [54], where the exponent value monotonically decreases as a value of p decreases. Indeed, the results in Section 4.2.2 showed outstanding performance for $p = 1$ rather than $p = 2$.

Figures 16, 17, 18, 19, 20, 21, 22, and 23 show examples of results with the proposed tempering approach, where the signals correspond to Figs. 7, 8, 9, 10, 11, 12, 13, and 14 with parameter tempering, and Table 6 shows the overall average results for all the signals. The results show that the parameter tempering improves the separation, particularly in ILRMA with heavy-tailed source models. Also, it further improves the results obtained using fixed values of the parameters. In total, the proposed generalization of ILRMA can achieve approximately 1.2 dB improvement in the SDR compared with the conventional ILRMA with the Gaussian model, which is a significant gain in BSS tasks with two sources.

4.2.4 Performance for various signal lengths

In BSS framework, the length of observed signal is important to achieve the better separation performance. This is because the accuracy of statistical estimation decreases when the number of time frames J is insufficient [60, 61]. In the extreme case, the demixing matrix $\boldsymbol{W}_i$ cannot be updated by IP when $J = 1$ because the rank of $\boldsymbol{U}_{i,n}$ in (16), $\boldsymbol{G}_{i,n}$ in (34), or $\boldsymbol{H}_{i,n}$ in (49) becomes unity. However, it is not clarified whether the heavy-tailed source distribution provides more robust statistical estimation for fewer time

Table 10 Relative computational times normalized by Laplace IVA based on IP, where the length of the observed signal is 10 s

Method	Two-source case	Three-source case
Laplace IVA based on IP [49]	1.00	1.00
GGD-IVA based on IP [17]	1.04	1.20
MNMF based on MM algorithm [32]	49.25	51.42
t-MNMF based on MM algorithm [40]	57.87	60.31
ILRMA based on IP [34]	1.10	1.19
GGD-ILRMA based on IP	1.32	1.38
t-ILRMA based on IP	1.20	1.27

frames or not. Thus, in this subsection, we experimentally compare the separation performance of ILRMA, GGD-ILRMA, and t-ILRMA for the observed signals with fewer and more time frames.

To simulate the short and long source signals, we utilized the dry sources described in Table 3. As the short dry sources, the music and speech signals were trimmed only to the former half, and their signal lengths were 2.5 s (music) and 5.0 s (speech), respectively. In contrast, the long music and speech signals were produced by repeating the entire length of the dry sources twice, namely, the signal lengths of the long music and speech dry sources become 10.0 s (music) and 20.0 s (speech), respectively. These dry sources were convoluted with `E2A_1` to produce the observed mixture signal with two sources, where the combinations of dry sources were the same as those described in Table 3. The other experimental conditions were the same as those in Section 4.2.2.

Figures 24 and 25 show the results of Music 1 for shorter and longer signals, respectively. Also, Table 7 shows the overall average results for all the signals. By comparing these figures and Fig. 7 (the results of Music 1 with the original length), we can confirm that the separation performance of all the methods improves in proportion to the number of time frames J. Similarly to ILRMA, GGD- and t-ILRMA also suffer from the degradation of separation performance depending on the decrease of J regardless of the heavy tail property.

4.3 BSS experiment with three sources

To emphasize the advantage of the proposed methods, we investigated a more difficult situation with three sources. In this experiment, for the sake of simplicity, we only compared the conventional ILRMA and the proposed GGD- and t-ILRMA. The used dry sources are shown in Table 8, which were convoluted with the impulse response depicted in Fig. 26. The other conditions were the same as those in Section 4.2.

Table 9 shows the overall average results of each method. Similar to the previous results, the proposed methods outperform the conventional ILRMA, and the tempering approach slightly improves the quality of separation compared with GGD- or t-ILRMA with fixed parameters.

4.4 Comparison of computational times

To demonstrate the optimization efficiency of ILRMA, we compared the computational times of Laplace IVA, GGD-IVA, MNMF, t-MNMF, ILRMA, GGD-ILRMA, and t-ILRMA. The update calculation for the NMF parameters in each algorithm was almost the same, but the estimation of the spatial parameter ($\boldsymbol{W}_i$ for ILRMA-based methods and the spatial covariance for MNMF-based methods) was different. Although ILRMA-based methods require one inverse of $\boldsymbol{W}_i\boldsymbol{U}_{i,n}$ for each i and n, MNMF-based methods require J inverses and two eigenvalue decompositions of the $M \times M$ matrix. Table 10 shows relative computational times normalized by that of Laplace IVA based on IP [49], where the conditions are the same as in Table 4 and we used MATLAB 9.2 (64-bit) with an AMD Ryzen 7 1800X (8 cores and 3.6 GHz) CPU. From this table, we can confirm that the computational time of ILRMA-based methods is not significantly larger than that of IVA, whereas that of MNMF-based methods is significantly larger.

5 Conclusions

In this paper, we proposed two generalizations of the source distribution assumed in ILRMA that introduce a heavy-tailed property by using the GGD and Student's t distribution. The GGD can be considered as a natural extension of the conventional Gaussian source model, and Student's t distribution partially satisfies the stable property of complex-valued random variables, which is desirable for NMF-based low-rank decomposition. We derived efficient optimization algorithms for GGD- and t-ILRMA, which ensure a monotonic decrease in the objective function and provide faster computation than existing MNMF-based BSS methods. Also, we revealed an interesting relationship between GGD- and t-NMF: GGD-NMF is equivalent to IS-NMF upon assuming the geometric mean of the data and the low-rank model as an observation, whereas t-NMF corresponds to the same algorithm with the harmonic mean of the data and the low-rank model as previously mentioned. These properties lead to more accurate parameter estimation in an ILRMA-based BSS framework, resulting in higher separation accuracy than the conventional ILRMA with the Gaussian source distribution. From the experiments, it is confirmed that the proposed generalized ILRMA improves the separation accuracy, especially for the music mixture signals. However, the improvement for speech mixture signals is still limited. This is because typical speech sources do not have an apparent low-rank time-frequency structure, and NMF-based source model in ILRMA cannot capture the precise spectral structures in speech sources even if the source model is generalized by the heavy-tailed distributions. The better modeling for speech sources remains as a future work.

Endnote

[1] Note that ILRMA was originally called rank-1 MNMF in [33, 34]. After the original publications, we renamed the method to clarify that ILRMA is a natural extension of IVA.

Abbreviations

BSS: Blind source separation; FDICA: Frequency-domain independent component analysis; GGD: Generalized Gaussian distribution; GGD-ILRMA: Independent low-rank matrix analysis based on generalized Gaussian distribution; GGD-IVA: Independent vector analysis based on spherical multivariate generalized Gaussian distribution; ICA: Independent component analysis; ILRMA: Independent low-rank matrix analysis; IP: Iterative projection; IS-NMF: Nonnegative matrix factorization based on Itakura–Saito divergence; IVA: Independent vector analysis; Laplace IVA: Independent vector analysis based on spherical multivariate Laplace distribution; MM: Majorization-minimization; MNMF: Multichannel nonnegative matrix factorization; NMF: Nonnegative matrix factorization; p.d.f.: Probability density function; SDR: Signal-to-distortion ratio STFT: Short-time Fourier transform; *t*-ILRMA: Independent low-rank matrix analysis based on Student's *t* distribution; *t*-MNMF: Multichannel nonnegative matrix factorization based on Student's *t* distribution; *t*-NMF: Nonnegative matrix factorization based on Student's *t* distribution; Time-varying Gaussian IVA: Independent vector analysis based on Gaussian distribution having a time-varying variance

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions that helped improve the quality of this manuscript.

Funding

This work was partly supported by the ImPACT Program of Council for Science, SECOM Science and Technology Foundation, and JSPS KAKENHI Grant Numbers JP16H01735, JP17H06101, and JP17H06572.

Availability of data and materials

Not available online. Please contact author for data requests.

Authors' contributions

All authors have contributed equally. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]National Institute of Technology, Kagawa College, 355 Chokushi, Takamatsu, Kagawa 761-8058, Japan. [2]The University of Tokyo, 7-3-1 Hongo, Bunkyo, 113-8656 Tokyo, Japan. [3]Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, 191-0065 Tokyo, Japan. [4]Yamaha Corporation, 203 Matsunokijima, Iwata, 438-0192 Shizuoka, Japan.

Received: 6 November 2017 Accepted: 19 April 2018
Published online: 02 May 2018

References

1. P Bofill, M Zibulevsky, Underdetermined blind source separation using sparse representations. Signal Process. **81**(11), 2353–2362 (2001)
2. S Araki, H Sawada, R Mukai, S Makino, Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors. Signal Process. **87**(8), 1833–1847 (2007)
3. L Zhen, D Peng, Z Yi, Y Xiang, P Chen, Underdetermined blind source separation using sparse coding. IEEE Trans. Neural Netw. Learn. Syst. **28**(12), 3102–3108 (2017)
4. P Comon, Independent component analysis, a new concept? Signal Process. **36**(3), 287–314 (1994)
5. P Smaragdis, Blind separation of convolved mixtures in the frequency domain. Neurocomputing. **22**(1), 21–34 (1998)
6. S Kurita, H Saruwatari, S Kajita, K Takeda, F Itakura, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. Evaluation of blind signal separation method using directivity pattern under reverberant conditions, (2000), pp. 3140–3143
7. H Sawada, R Mukai, S Araki, S Makino, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. Convolutive blind source separation for more than two sources in the frequency domain, (2004), pp. 885–888
8. H Saruwatari, T Kawamura, T Nishikawa, A Lee, K Shikano, Blind source separation based on a fast-convergence algorithm combining ICA and beamforming. IEEE Trans. Audio Speech Lang. Process. **14**(2), 666–678 (2006)
9. N Murata, S Ikeda, A Ziehe, An approach to blind source separation based on temporal structure of speech signals. Neurocomputing. **41**(1–4), 1–24 (2001)
10. H Sawada, R Mukai, S Araki, S Makino, A robust and precise method for solving the permutation problem of frequency-domain blind source separation. IEEE Trans. Speech Audio Process. **12**(5), 530–538 (2004)
11. H Sawada, S Araki, S Makino, in *Proc. IEEE Int. Symp. Circuits Syst*. Measuring Dependence of Bin-wise Separated Signals for Permutation Alignment in Frequency-Domain BSS, (2007), pp. 3247–3250
12. A Hiroe, in *Proc. Int. Conf. Independent Compon. Anal. Blind Source Separation*. Solution of permutation problem in frequency domain ICA using multivariate probability density functions, (2006), pp. 601–608
13. T Kim, T Eltoft, T-W Lee, in *Proc. Int. Conf. Independent Compon. Anal. Blind Source Separation*. Independent vector analysis: an extension of ICA to multivariate components, (2006), pp. 165–172
14. T Kim, HT Attias, S-Y Lee, T-W Lee, Blind source separation exploiting higher-order frequency dependencies. IEEE Trans. Audio Speech Lang. Process. **15**(1), 70–79 (2007)
15. G Box, G Tiao, *Bayesian Inference in Statistical Analysis*. (Addison Wesley, Reading, Mass, 1973)
16. T Itahashi, K Matsuoka, Stability of independent vector analysis. Signal Process. **92**(8), 1809–1820 (2012)
17. N Ono, in *Proc. Asia-Pacific Signal and Info. Process. Assoc. Annual Summit and Conf*. Auxiliary-function-based independent vector analysis with power of vector-norm type weighting functions, (2012)
18. M Anderson, GS Fu, R Phlypo, T Adalı, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. Independent vector analysis, the Kotz distribution, and performance bounds, (2013), pp. 3243–3247
19. Y Liang, J Harris, SM Naqvi, G Chen, JA Chambers, Independent vector analysis with a generalized multivariate Gaussian source prior for frequency domain blind source separation. Signal Process. **105**, 175–184 (2014)
20. Z Boukouvalas, GS Fu, T Adalı, in *Proc. Annual Conf. Info. Sci. and Syst*. An efficient multivariate generalized Gaussian distribution estimator: application to IVA, (2015)
21. T Ono, N Ono, S Sagayama, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. User-guided independent vector analysis with source activity tuning, (2012), pp. 2417–2420
22. DD Lee, HS Seung, Learning the parts of objects by non-negative matrix factorization. Nature. **401**(6755), 788–791 (1999)
23. DD Lee, HS Seung, in *Proc. Neural Info. Process. Syst*. Algorithms for non-negative matrix factorization, (2000), pp. 556–562
24. T Virtanen, Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. IEEE Trans. Audio, Speech, Lang. Process. **15**(3), 1066–1074 (2007)
25. P Smaragdis, B Raj, M Shashanka, in *Proc. Int. Conf. Independent Compon. Anal. Signal Separation*. Supervised and semi-supervised separation of sounds from single-channel mixtures, (2007), pp. 414–421
26. A Ozerov, C Févotte, M Charbit, in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust*. Factorial scaled hidden Markov model for polyphonic audio representation and source separation, (2009), pp. 121–124
27. D Kitamura, H Saruwatari, K Yagi, K Shikano, Y Takahashi, K Kondo, Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **E97-A**(5), 1113–1118 (2014)
28. D Kitamura, H Saruwatari, H Kameoka, Y Takahashi, K Kondo, S Nakamura, Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration. IEEE/ACM Trans. Audio, Speech, Lang. Process. **23**(4), 654–669 (2015)
29. C Févotte, N Bertin, J-L Durrieu, Nonnegative matrix factorization with the Itakura–Saito divergence. With application to music analysis. Neural Comput. **21**(3), 793–830 (2009)

30. A Ozerov, C Févotte, Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. IEEE Trans. Audio, Speech, Lang. Process. **18**(3), 550–563 (2010)
31. H Kameoka, T Yoshioka, M Hamamura, JL Roux, K Kashino, in *Proc. Int. Conf. Latent Variable Anal. Signal Separation*. Statistical model of speech signals based on composite autoregressive system with application to blind source separation, (2010), pp. 245–253
32. H Sawada, H Kameoka, S Araki, N Ueda, Multichannel extensions of non-negative matrix factorization with complex-valued data. IEEE Trans. Audio, Speech, Lang. Process. **21**(5), 971–982 (2013)
33. D Kitamura, N Ono, H Sawada, H Kameoka, H Saruwatari, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model, (2015), pp. 276–280
34. 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. Audio, Speech, Lang. Process. **24**(9), 1626–1641 (2016)
35. D Kitamura, N Ono, H Sawada, H Kameoka, H Saruwatari, in *Audio Source Separation*, ed. by S Makino. Determined blind source separation with independent low-rank matrix analysis (Springer, Cham, 2018), pp. 125–155. https://link.springer.com/chapter/10.1007%2F978-3-319-73031-8_6#citeas
36. C Févotte, SJ Godsill, A Bayesian approach for blind separation of sparse sources. IEEE Trans. Audio, Speech, Lang. Process. **14**(6), 2174–2188 (2006)
37. S Leglaive, R Badeau, G Richard, in *Proc. Eur. Signal Process. Conf*. Semi-blind Student's t source separation for multichannel audio convolutive mixtures, (2017)
38. A Liutkus, D FitzGerald, R Badeau, in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust*. Cauchy nonnegative matrix factorization, (2015)
39. K Yoshii, K Itoyama, M Goto, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. Student's t nonnegative matrix factorization and positive semidefinite tensor factorization for single-channel audio source separation, (2016), pp. 51–55
40. K Kitamura, Y Bando, K Itoyama, K Yoshii, in *Proc. Int. Workshop Acoust. Signal Enh*. Student's t multichannel nonnegative matrix factorization for blind source separation, (2016)
41. G Samorodnitsky, MS Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. (Chapman & Hall/CRC Press, Florida, 1994)
42. A Liutkus, R Badeau, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. Generalized Wiener filtering with fractional power spectrograms, (2015), pp. 266–270
43. S Leglaive, U Simsekli, A Liutkus, R Badeau, G Richard, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. Alpha-stable multichannel audio source separation, (2017), pp. 576–580
44. S Mogami, D Kitamura, Y Mitsui, N Takamune, H Saruwatari, N Ono, in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process*. Independent low-rank matrix analysis based on complex Student's t-distribution for blind audio source separation, (2017)
45. NQK Duong, E Vincent, R Gribonval, Under-determined reverberant audio source separation using a full-rank spatial covariance model. IEEE Trans. Audio Speech Lang. Process. **18**(7), 1830–1840 (2010)
46. D Kitamura, N Ono, H Sawada, H Kameoka, H Saruwatari, in *Proc. Eur. Signal Process. Conf*. Relaxation of rank-1 spatial constraint in overdetermined blind source separation, (2015), pp. 1271–1275
47. DR Hunter, K Lange, Quantile regression via an MM algorithm. J. Comput. Graph. Stat. **9**(1), 60–77 (2000)
48. N Ono, S Miyabe, in *Proc. Int. Conf. Latent Variable Anal. Signal Separation*. Auxiliary-function-based independent component analysis for super-Gaussian sources, (2010), pp. 165–172
49. N Ono, in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust*. Stable and fast update rules for independent vector analysis based on auxiliary function technique, (2011), pp. 189–192
50. M Nakano, H Kameoka, JL Roux, Y Kitano, N Ono, S Sagayama, in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process*. Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with beta-divergence, (2010), pp. 283–288
51. N Murata, S Ikeda, A Ziehe, An approach to blind source separation based on temporal structure of speech signals. Neurocomputing. **41**(1–4), 1–24 (2001)
52. D Kitamura, Algorithms for Independent Low-rank Matrix Analysis. http://d-kitamura.net/pdf/misc/AlgorithmsForIndependentLowRankMatrixAnalysis.pdf. Accessed 27 Apr 2018
53. C Févotte, J Idier, Algorithms for nonnegative matrix factorization with the β-divergence. Neural Comput. **23**(9), 2421–2456 (2011)
54. Y Mitsui, D Kitamura, N Takamune, H Saruwatari, Y Takahashi, K Kondo, in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process*. Independent low-rank matrix analysis based on parametric majorization-equalization algorithm, (2017), pp. 98–102
55. D Kitamura, Open Dataset: songKitamura. http://d-kitamura.net/en/dataset_en.htm. Accessed 27 Apr 2018
56. S Araki, F Nesta, E Vincent, Koldovský, G Nolte, A Ziehe, A Benichoux, in *Proc. Int. Conf. Latent Variable Anal. Signal Separation*. The 2011 signal separation evaluation campaign (SiSEC2011):-audio source separation, (2012), pp. 414–422
57. Third Community-based Signal Separation Evaluation Campaign (SiSEC 2011). http://sisec2011.wiki.irisa.fr. Accessed 27 Apr 2018
58. S Nakamura, K Hiyane, F Asano, T Nishiura, T Yamada, in *Proc. Int. Conf. Lang. Res. Eval*. Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition, (2000), pp. 965–968
59. E Vincent, R Gribonval, C Févotte, Performance measurement in blind audio source separation. IEEE Trans. Audio, Speech, Lang. Process. **14**(4), 1462–1469 (2006)
60. S Araki, R Mukai, S Makino, T Nishikawa, H Saruwatari, The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech. IEEE Trans. Speech and Audio Process. **11**(2), 109–116 (2003)
61. D Kitamura, N Ono, H Saruwatari, in *Proc. Eur. Signal Process. Conf*. Experimental analysis of optimal window length for independent low-rank matrix analysis, (2017), pp. 1210–1214