Independent Low-Rank Matrix Analysis Based on Time-Variant Sub-Gaussian Source Model for Determined Blind Source Separation

Shinichi Mogami ⓘ, *Student Member, IEEE*, Norihiro Takamune, Daichi Kitamura ⓘ, *Member, IEEE*, Hiroshi Saruwatari ⓘ, *Member, IEEE*, Yu Takahashi, *Member, IEEE*, Kazunobu Kondo, and Nobutaka Ono, *Senior Member, IEEE*

Abstract—**Independent low-rank matrix analysis (ILRMA) is a fast and stable method of blind audio source separation. Conventional ILRMAs assume time-variant (super-)Gaussian source models, which can only represent signals that follow a super-Gaussian distribution. In this article, we focus on ILRMA based on a generalized Gaussian distribution (GGD-ILRMA) and propose a new type of GGD-ILRMA that adopts a time-variant sub-Gaussian distribution for the source model. We propose a new update scheme called generalized iterative projection for homogeneous source models (GIP-HSM) and obtain a convergence-guaranteed update rule for demixing spatial parameters by combining the GIP-HSM scheme and the majorization-minimization (MM) algorithm. Furthermore, a new extension of the MM algorithm is proposed for the convergence acceleration by applying the majorization-equalization algorithm to a multivariate case. In the experimental evaluation, we show the versatility of the proposed method, i.e., the proposed time-variant sub-Gaussian source model can be applied to various types of source signal.**

Index Terms—**Blind source separation, independent low-rank matrix analysis (ILRMA), generalized Gaussian distribution.**

I. INTRODUCTION

BLIND source separation (BSS) [1]–[14] is a technique of extracting specific sources from an observed multichannel mixture signal without knowing a priori information about the mixing system. Independent component analysis (ICA) and its extensions, such as independent vector analysis (IVA) [5], [6], are the most commonly used approaches for BSS in an

(over)determined case (the number of microphones is greater than or equal to that of sources). As a state-of-the-art ICA-based BSS method, Kitamura *et al.* proposed *independent low-rank matrix analysis (ILRMA)* [10], [14], which is a unification of IVA and nonnegative matrix factorization (NMF) [15]. ILRMA assumes both statistical independence between sources and a low-rank time-frequency structure for each source, and the frequency-wise demixing systems are estimated without encountering the permutation problem [3], [4]. ILRMA is a faster and more stable algorithm than multichannel NMF (MNMF) [16]–[18], which is an algorithm for BSS that estimates the mixing system on the basis of spatial covariance matrices.

The original ILRMA based on Itakura–Saito (IS) divergence assumes a time-variant isotropic (zero-mean and circularly symmetric) complex Gaussian distribution for a source generative model. Hereafter, we refer to the original ILRMA as *IS-ILRMA*. Recently, various types of source generative model have been proposed for ILRMA for high-performance BSS. In particular, t-ILRMA [11] and GGD-ILRMA [12], [13] have been proposed as generalizations of IS-ILRMA with a complex Student's t distribution and a complex generalized Gaussian distribution (GGD) [19], respectively. Their update rules for the demixing matrix can be derived by a combined technique of the majorization-minimization (MM) algorithm [20] and iterative projection (IP), where IP is a fast and convergence-guaranteed algorithm introduced first for auxiliary-function-based ICA (AuxICA) [8] and auxiliary-function-based IVA (AuxIVA) [9]. In t-ILRMA and GGD-ILRMA, the kurtosis of the source generative models' distributions can be parametrically changed along with the degree-of-freedom parameter in Student's t distribution and the shape parameter in the GGD. By changing the kurtosis of the distributions, we can control how often the source signal outputs outliers or its expected sparsity. In particular, in sub-Gaussian models, i.e., models that follow distributions with a *platykurtic* shape, the source signal rarely outputs outliers, where platykurtic means that the kurtosis is lower than that of a Gaussian distribution. Therefore, the sub-Gaussian modeling of sources is expected to accurately estimate the source spectrogram without ignoring its important spectral peaks. Furthermore, many audio sources follow platykurtic distributions; it is known that musical instrument signals obey sub-Gaussian distributions [21].
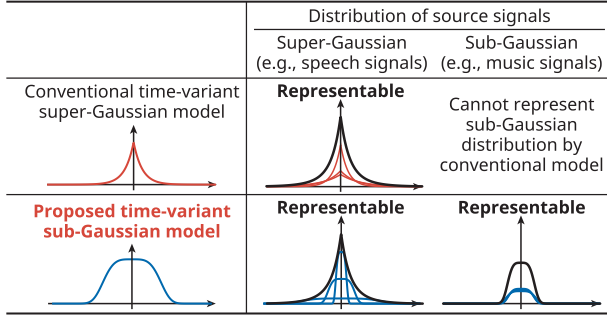
Fig. 1.   Versatility of proposed time-variant sub-Gaussian model for super- and sub-Gaussian source signals.

However, neither conventional t-ILRMA nor GGD-ILRMA assumes that the source generative model follows a sub-Gaussian distribution. Both t-ILRMA and GGD-ILRMA can adopt only a super-Gaussian (or Gaussian) model, i.e., a model that follows distributions with a *leptokurtic* shape, as the source model, where leptokurtic means that the kurtosis is higher than that of a Gaussian distribution. This is so in t-ILRMA because the complex Student's t distribution becomes only super-Gaussian for any degree-of-freedom parameter. On the other hand, this is so in GGD-ILRMA because the estimation algorithm for the demixing matrix has not yet been derived for a sub-Gaussian case, although the GGD itself can represent a sub-Gaussian distribution depending on its shape parameter. More specifically, the conventional algorithm, i.e., the combination of the MM algorithm and conventional IP, cannot be applied to sub-Gaussian-based GGD-ILRMA owing to mathematical difficulties.

In this paper, we propose a new type of ILRMA that assumes time-variant sub-Gaussian source models. This paper includes four novelties. The first novelty is a new update scheme for the demixing matrix called *generalized IP for homogeneous source models (GIP-HSM)*. The second novelty is a convergence-guaranteed update rule for the demixing matrix in GGD-ILRMA when the shape parameters are between two and four. To the best of our knowledge, this is the world's first attempt to model source signals with a time-variant sub-Gaussian distribution. We derive the update rule by combining the above-mentioned GIP-HSM scheme and the MM algorithm. The third novelty is that we propose a new extension of the MM algorithm for the convergence acceleration by applying the majorization-equalization (ME) algorithm [22] to a multivariate case. The fourth novelty is that we show the validity of the proposed sub-Gaussian GGD-ILRMA via BSS experiments on music and speech signals. In Fig. 1, we summarize the versatility of the proposed method in source modeling. The proposed time-variant sub-Gaussian model can represent super-Gaussian or Gaussian signals as well as sub-Gaussian signals owing to its time-variant nature, whereas conventional models can only represent super-Gaussian or Gaussian signals. The details of this versatility will be discussed in Section-IV.A.

The rest of this paper is organized as follows. In Section II, we formulate the problem of BSS and derive the cost function

to be minimized. In Section III, we introduce the conventional update rules in super-Gaussian GGD-ILRMA. In Section IV, we propose GIP-HSM and derive the update rules in sub-Gaussian GGD-ILRMA. In Section V, we propose the acceleration method of the update rule based on the ME algorithm. In Section VI, we show the results of experimental evaluations. The conclusions of this paper are presented in Section VII. Note that this paper is partially based on an international conference paper [23] written by the authors. This paper has the following additional contributions: the derivation of the more generalized update rules when the shape parameter ranges between two and four, the acceleration of the update rules using the ME algorithm, and more detailed experimental evaluations including the comparison with various conventional methods.

II. FORMULATION

A. Formulation of Demixing Model

Let N and M be the numbers of sources and channels, respectively. The short-time Fourier transforms (STFTs) of the multichannel source and observed signals are defined as

$$\boldsymbol{s}_{ij} = [s_{ij1} \ \cdots \ s_{ijN}]^\top \in \mathbb{C}^N, \qquad (1)$$

$$\boldsymbol{x}_{ij} = [x_{ij1} \ \cdots \ x_{ijM}]^\top \in \mathbb{C}^M, \qquad (2)$$

where $i = 1, \ldots, I$; $j = 1, \ldots, J$; $n = 1, \ldots, N$; and $m = 1 \ldots, M$ are the indices of the frequency bins, time frames, sources, and channels, respectively, and $^\top$ denotes the transpose. We assume the mixing system

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij}, \qquad (3)$$

where $\boldsymbol{A}_i = [\boldsymbol{a}_{i1} \ \cdots \ \boldsymbol{a}_{iN}] \in \mathbb{C}^{M \times N}$ is a frequency-wise mixing matrix and $\boldsymbol{a}_{in}$ is the steering vector for the nth source. When $M = N$ and $\boldsymbol{A}_i$ is not a singular matrix, the source signals can be estimated as

$$\boldsymbol{s}_{ij} \approx \boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij}, \qquad (4)$$

where

$$\boldsymbol{y}_{ij} = [y_{ij1} \ \cdots \ y_{ijN}]^\top \in \mathbb{C}^N \qquad (5)$$

is the STFT of the estimated signals, $\boldsymbol{W}_i = [\boldsymbol{w}_{i1} \ \cdots \ \boldsymbol{w}_{iN}]^\mathsf{H}$ is the demixing matrix, $\boldsymbol{w}_{in}$ is the demixing filter for the nth source, and $^\mathsf{H}$ denotes the Hermitian transpose. ILRMA estimates both $\boldsymbol{W}_i$ and $\boldsymbol{y}_{ij}$ from only the observation $\boldsymbol{x}_{ij}$ assuming statistical independence between s_{ijn} and $s_{ijn'}$, where $n \neq n'$.

B. Generative Model and Cost Function in GGD-ILRMA

GGD-ILRMA utilizes the isotropic complex GGD. The probability density function of the GGD is

$$p(z) = \frac{\beta}{2\pi r^2 \Gamma(2/\beta)} \exp\left(-\frac{|z|^\beta}{r^\beta}\right), \qquad (6)$$

where β is the shape parameter, r is the scale parameter, and $\Gamma(\cdot)$ is the gamma function. Fig. 2 shows the shapes of the GGD with $\beta = 2$ and $\beta = 4$. When $\beta = 2$, (6) corresponds to the probability density function of the complex Gaussian

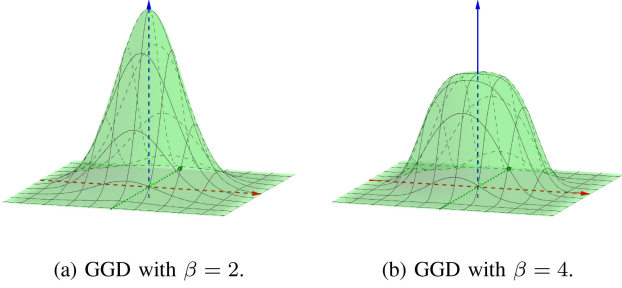(a) GGD with $\beta = 2$. (b) GGD with $\beta = 4$.

Fig. 2. Examples of shapes of complex GGD. (a) When $\beta = 2$, shape of GGD corresponds to that of Gaussian distribution. (b) When $\beta = 4$, GGD is platykurtic.

Fig. 3. Principle of source separation in GGD-ILRMA, where $\tilde{\boldsymbol{x}}_m$ and $\tilde{\boldsymbol{y}}_n$ are time-domain signals of x_{ijn} and y_{ijn}, respectively.

distribution with a *mesokurtic* shape, where mesokurtic means that the kurtosis is equal to that of a Gaussian distribution. In the case of $0 < \beta < 2$, the distribution becomes super-Gaussian with a leptokurtic shape. In the case of $\beta > 2$, the distribution becomes sub-Gaussian with a platykurtic shape.

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

$$p(\{y_{ijn}\}_{i,j}) = \prod_{i,j} p(y_{ijn})$$

$$= \prod_{i,j} \frac{\beta}{2\pi r_{ijn}^2 \Gamma(2/\beta)} \exp\left(-\frac{|y_{ijn}|^\beta}{r_{ijn}^\beta}\right), \quad (7)$$

$$r_{ijn}{}^\rho = \sum_k t_{ikn} v_{kjn}, \quad (8)$$

where $\{y_{ijn}\}_{i,j}$ is the set $\{y_{ijn} \mid i = 1, \ldots, I, j = 1, \ldots, J\}$ and the local distribution $p(y_{ijn})$ is defined as a circularly symmetric complex Gaussian distribution, i.e., the probability of $p(y_{ijn})$ only depends on the power of the complex value y_{ijn}. r_{ijn} is the time-frequency-varying scale parameter and ρ is the domain parameter in NMF modeling. Moreover, the variables t_{ikn} and v_{kjn} are the elements of the basis matrix $\boldsymbol{T}_n \in \mathbb{R}_{\geq 0}^{I \times K}$ and the activation matrix $\boldsymbol{V}_n \in \mathbb{R}_{\geq 0}^{K \times J}$, respectively, where $\mathbb{R}_{\geq 0}$ denotes the set of nonnegative real numbers. $k = 1, \ldots, K$ is the index, and K is set to a much smaller value than I and J, which leads to the low-rank approximation. From (7), the negative log-likelihood function $\mathcal{L}_{\mathrm{GGD}}$ of the observed signal $\boldsymbol{x}_{ij}$ can be obtained as follows by assuming independence between sources:

$$\mathcal{L}_{\mathrm{GGD}} := -\log p(\{x_{ijm}\}_{i,j,m})$$

$$= -2J \sum_i \log|\det \boldsymbol{W}_i| - \sum_n \log p(\{y_{ijn}\}_{i,j})$$

$$= -2J \sum_i \log|\det \boldsymbol{W}_i|$$

$$+ \sum_{i,j,n} \left(\frac{|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^\beta}{r_{ijn}^\beta} + 2\log r_{ijn}\right) + \mathrm{const.}, \quad (9)$$

where $\{x_{ijm}\}_{i,j,m}$ is the set $\{x_{ijm} \mid i = 1, \ldots, I, j = 1, \ldots, J, m = 1, \ldots, M\}$. We used the transformation of random
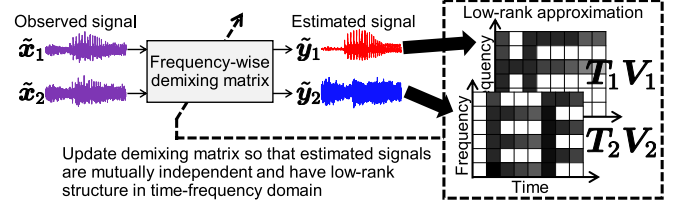
variables from $\boldsymbol{x}_{ij}$ to $\boldsymbol{y}_{ij}$ as denoted in (4). The cost function (9) of GGD-ILRMA coincides with that of IS-ILRMA when $\beta = \rho = 2$. By minimizing (9) w.r.t. $\boldsymbol{W}_i$ and r_{ijn} under limitation (8), we estimate the demixing system that maximizes the independence between sources.

Fig. 3 shows a conceptual model of GGD-ILRMA. When each of the original sources has a low-rank spectrogram, the spectrogram of their mixture should be more complicated, where the rank of the mixture spectrogram will be greater than that of the source spectrogram. On the basis of this assumption, in GGD-ILRMA, the low-rank constraint for each estimated spectrogram is introduced by NMF. The demixing matrix $\boldsymbol{W}_i$ is estimated so that the spectrogram of the estimated signal 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 (9) in a fully blind manner.

III. CONVENTIONAL METHOD

A. Update Rule for Demixing Matrix

In IS-ILRMA, the demixing matrix $\boldsymbol{W}_i$ can be efficiently updated by IP, which can be applied only when the cost function is the sum of $-\log|\det \boldsymbol{W}_i|$ and the quadratic form of $\boldsymbol{w}_{in}$ (this corresponds to GGD-ILRMA with $\beta = 2$). In GGD-ILRMA, the update rule of $\boldsymbol{W}_i$ is also derived using the MM algorithm [20]. When $0 < \beta \leq 2$, we can use the following inequality of weighted arithmetic and geometric means (weighted AM–GM inequality) to design the majorization function:

$$|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^\beta \leq \frac{\beta}{2} \frac{|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2}{\alpha_{ijn}^{2-\beta}} + \left(1 - \frac{\beta}{2}\right)\alpha_{ijn}{}^\beta, \quad (10)$$

where α_{ijn} is an auxiliary variable and the equality of (10) holds if and only if $\alpha_{ijn} = |y_{ijn}|$. By applying (10) to (9), we can design the majorization function of (9) as

$$\mathcal{L}_{\mathrm{GGD}} \leq -2J \sum_i \log|\det \boldsymbol{W}_i|$$

$$+ J \sum_{i,n} \boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{F}_{in} \boldsymbol{w}_{in} + \mathrm{const.}, \quad (11)$$

$$\boldsymbol{F}_{in} = \frac{\beta}{2J} \sum_j \frac{1}{\alpha_{ijn}^{2-\beta}\left(\sum_k t_{ikn} v_{kjn}\right)^{\frac{\beta}{\rho}}} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}}, \quad (12)$$

where the constant term is independent of $\boldsymbol{w}_{in}$. By applying IP to (11) and substituting the equality condition $\alpha_{ijn} = |y_{ijn}|$ into

(12), we can derive the update rule for $\boldsymbol{W}_i$ as

$$\boldsymbol{F}_{in} = \frac{\beta}{2J} \sum_j \frac{1}{|y_{ijn}|^{2-\beta} \left(\sum_k t_{ikn} v_{kjn}\right)^{\frac{\beta}{\rho}}} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}}, \quad (13)$$

$$\boldsymbol{w}_{in} \leftarrow \boldsymbol{F}_{in}^{-1} \boldsymbol{W}_i^{-1} \boldsymbol{e}_n, \quad (14)$$

$$\boldsymbol{w}_{in} \leftarrow \boldsymbol{w}_{in} \sqrt{1/(\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{F}_{in} \boldsymbol{w}_{in})}, \quad (15)$$

where $\boldsymbol{e}_n$ is an N-dimensional vector whose nth element is one and whose other elements are zero. When $\beta = \rho = 2$, these update rules (13)–(15) coincide with those in IS-ILRMA. The update rules (13)–(15) are sequentially calculated for each frequency and source (i and n). This calculation guarantees a monotonic decrease in the value of cost function (9), resulting in the maximization of independence between sources and the separation of mixed sources [9].

Note that the update rules (13)–(15) are valid only when $0 < \beta \leq 2$, which is equivalent to the condition that inequality (10) holds. In fact, when $\beta > 2$, it is thought to be impossible to design a majorization function to which we can apply IP because no quadratic function w.r.t. x can majorize x^β.

Conventional GGD-ILRMA achieves various types of source generative model: when $\beta = 2$, the entry of the source spectrogram follows a complex Gaussian distribution (the same model as that of IS-ILRMA), and when $\beta < 2$, the entry of the source spectrogram follows a complex leptokurtic distribution. However, a source generative model that follows a platykurtic complex GGD is yet to be achieved. Since the global distribution of the time-variant super-Gaussian or Gaussian model w.r.t. the time frame becomes only super-Gaussian, any signals that follow sub-Gaussian distributions, such as some music signals, cannot be appropriately dealt with by conventional GGD-ILRMA.

B. Update Rule for Low-Rank Source Model

The update rules for $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$ in IS-ILRMA and GGD-ILRMA can be derived by the MM algorithm, which is a popular approach for NMF. We obtain the following update rules:

$$t_{ikn} \leftarrow t_{ikn} \left(\frac{\beta \sum_j \frac{|y_{ijn}|^\beta}{\left(\sum_{k'} t_{ik'n} v_{k'jn}\right)^{\frac{\beta}{\rho}+1}} v_{kjn}}{2 \sum_j \frac{1}{\sum_{k'} t_{ik'n} v_{k'jn}} v_{kjn}} \right)^{\frac{\rho}{\beta+\rho}}, \quad (16)$$

$$v_{kjn} \leftarrow v_{kjn} \left(\frac{\beta \sum_j \frac{|y_{ijn}|^\beta}{\left(\sum_{k'} t_{ik'n} v_{k'jn}\right)^{\frac{\beta}{\rho}+1}} t_{ikn}}{2 \sum_j \frac{1}{\sum_{k'} t_{ik'n} v_{k'jn}} t_{ikn}} \right)^{\frac{\rho}{\beta+\rho}}. \quad (17)$$

See [12] for their detailed derivation.

In GGD-ILRMA, cost function (9) is minimized by alternately repeating the update of the demixing matrix $\boldsymbol{W}_i$ using (13)–(15) and the update of the low-rank source models $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$ using (16) and (17), respectively. The demixing matrix $\boldsymbol{W}_i$ is updated so that (9) monotonically decreases with the source models $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$ fixed, and then the source models $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$ are updated so that (9) monotonically decreases with the demixing matrix $\boldsymbol{W}_i$ fixed; i.e., these optimization steps are based on a coordinate descent algorithm. Thus, a monotonic decrease in cost function (9) is always guaranteed owing to the monotonic-decrease-guaranteed update rules (13)–(17) for all the variables $\boldsymbol{w}_{in}$, t_{ikn}, and v_{kjn}.

IV. PROPOSED METHOD

A. Motivation

The conventional methods [10], [12] have the limitation that a source signal cannot be appropriately represented when the signal follows a sub-Gaussian distribution. In this paper, we propose an MM-algorithm-based update rule for GGD-ILRMA to maximize the likelihood based on the sub-Gaussian source model.

In contrast to the time-variant super-Gaussian or Gaussian model, the global distribution of the time-variant sub-Gaussian model w.r.t. the time frame can be sub-Gaussian as well as Gaussian or super-Gaussian, depending on the time variance of the scale parameter r_{ijn}, as illustrated in Fig. 1, where the bold black curves depict the global distributions of time-variant models. When the scale parameter fluctuates w.r.t. the time frame, the global distribution always approaches the super-Gaussian side. This nature can be demonstrated via a simple experiment. The purpose of this experiment is that (a) we use a zero-mean uniform distribution as a typical sub-Gaussian distribution, (b) the fluctuation of the scale parameter is simulated by artificial modulation, and (c) we show that the uniform distribution with the time-variant scale parameter approaches the super-Gaussian distribution by calculating its kurtosis. We assume the following uniform distribution with a time-variant range parameter:

$$\chi_j \sim \mathcal{U}(-\gamma_j, \gamma_j), \quad (18)$$

where χ_j is a random variable with the time index j, $\mathcal{U}(-\gamma_j, \gamma_j)$ is the uniform distribution whose range is defined as $[-\gamma_j, \gamma_j]$, and γ_j is the time-variant range parameter. Note that the uniform distribution corresponds to the univariate GGD with $\beta \to \infty$, which has the most platykurtic shape in a GGD class. Fig. 4 shows the kurtosis, $\mathrm{Kurt} = \mathrm{E}[\chi_j^4]/\mathrm{E}[\chi_j^2]^2 - 3$, of the random variable χ_j for $j = 1, \ldots, 10^6$ ($J = 10^6$ random samples), where $\mathrm{E}[\cdot]$ denotes the expectation value. Thus, $\mathrm{Kurt} = 0$, $\mathrm{Kurt} > 0$, or $\mathrm{Kurt} < 0$ respectively represents a Gaussian, super-Gaussian, or sub-Gaussian distribution. In Fig. 4, we introduced a modulation parameter δ into γ_j and $\omega = 2\pi \times 5/J$. This result clearly indicates that the fluctuation of a scale parameter w.r.t. the time frame increases the kurtosis of the global distribution even if the fundamental distribution is sub-Gaussian. This supports that the proposed time-variant sub-Gaussian model covers distributions with a wider range between platykurtic and leptokurtic shapes than other conventional source models. Therefore, the proposed GGD-ILRMA is expected to have a robust performance against the variation of target signals.

For the above-mentioned reason, we can introduce a more robust and generalized source model for ILRMA by using the time-variant sub-Gaussian model. However, as described in
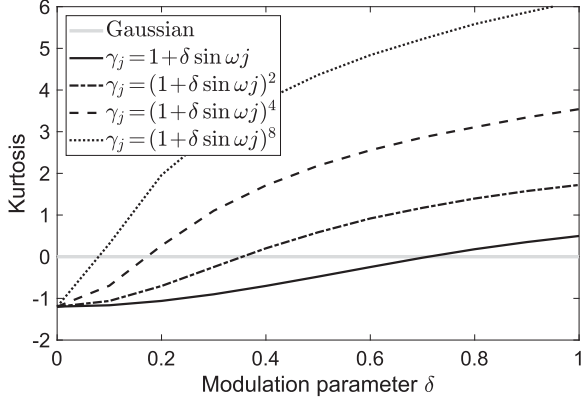
Fig. 4. Kurtosis drift of time-variant uniform distribution model with various fluctuations of range parameter, where kurtosis becomes zero, positive, or negative when distribution is Gaussian, super-Gaussian, or sub-Gaussian, respectively.

Section III-A, the conventional IP-based MM algorithm (13)–(15) cannot be utilized for GGD-ILRMA with the sub-Gaussian model ($\beta > 2$). To solve this problem, in Section IV-B, we extend the problem of demixing matrix estimation by replacing the quadratic form of $\boldsymbol{w}_{in}$ with a more generalized function $f_{in}(\boldsymbol{w}_{in})$, where the extended problem includes an optimization of the cost function in GGD-ILRMA with $\beta > 2$. Then, we derive an optimization algorithm called GIP-HSM that can solve this extended problem. In Section IV-C, the update rule of $\boldsymbol{W}_i$ for GGD-ILRMA with $\beta = 4$ is derived by combining GIP-HSM and the MM algorithm, where one majorization function is designed and applied to the cost function of GGD-ILRMA with $\beta = 4$. A computationally efficient version of this update rule is presented in Section IV-D. Finally, in Section IV-E, the update rules of $\boldsymbol{W}_i$ for GGD-ILRMA with $2 < \beta < 4$ are derived, where another majorization function is further applied to the majorization function of GGD-ILRMA with $\beta = 4$, namely, an MM algorithm is utilized twice to derive the update rules for $2 < \beta < 4$, which is an analogous derivation to the conventional IP for $0 < \beta < 2$.

B. Derivation of GIP-HSM

If we represent the second term in (9) as $f_{in}(\boldsymbol{w}_{in}) = (1/J) \sum_j (|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^\beta / r_{ijn}^\beta)$, the cost function of GGD-ILRMA can be rewritten as

$$\mathcal{L} = \sum_{i=1}^{I} \left[-2 \log |\det \boldsymbol{W}_i| + \sum_{n=1}^{N} f_{in}(\boldsymbol{w}_{in}) \right] + \text{const.} \quad (19)$$

up to the scale $1/J$, where the constant term is independent of $\boldsymbol{w}_{in}$. Note that we only focus on the optimization of $\boldsymbol{w}_{in}$, and the source scale r_{ijn} is assumed to be a fixed parameter in this section. By generalizing the term $f_{in}(\boldsymbol{w}_{in})$, we can define the extended optimization problem that includes the cost function (9) as a special case and covers various types of the source generative model. Here, we define the generalized term $f_{in}(\boldsymbol{w}_{in})$ as follows: $f_{in}: \mathbb{C}^N \to \mathbb{R}$ is a real-valued function that satisfies the following three conditions:

(C1) $f_{in}(\boldsymbol{w})$ is differentiable w.r.t. $\boldsymbol{w}$ at an arbitrary point.
(C2) $\forall c > 0, \{\boldsymbol{w} \in \mathbb{C}^N \mid f_{in}(\boldsymbol{w}) \leq c\}$ is convex (naturally satisfied when $f_{in}(\boldsymbol{w})$ is convex).
(C3) $\forall \eta, f_{in}(\eta \boldsymbol{w}) = \eta^d f_{in}(\boldsymbol{w})$, namely, f_{in} is a homogeneous function of degree d.

Next, we show that the optimization of (19) w.r.t. $\boldsymbol{w}_{in}$ is composed of "direction optimization" and "scale optimization" for each frequency bin. Let $\boldsymbol{u}_{in}$ be an N-dimensional vector that satisfies $f_{in}(\boldsymbol{u}_{in}) = 1$. Then, $\boldsymbol{w}_{in}$ can be uniquely represented as $\boldsymbol{w}_{in} = \eta_{in} \boldsymbol{u}_{in}$ by C3, where η_{in} is a positive real value. By substituting $\boldsymbol{w}_{in} = \eta_{in} \boldsymbol{u}_{in}$ into (19) and applying C3, we can rewrite cost function (19) as

$$\mathcal{L} = \sum_{i=1}^{I} \left[-2 \log \left| \det \left[\eta_{i1} \boldsymbol{u}_{i1} \quad \dots \quad \eta_{iN} \boldsymbol{u}_{iN} \right]^{\mathsf{H}} \right| \right.$$
$$\left. + \sum_{n=1}^{N} f_{in}(\eta_{in} \boldsymbol{u}_{in}) \right]$$
$$= \sum_{i=1}^{I} \left[-2 \log \left(\prod_n \eta_{in} \cdot |\det \boldsymbol{U}_i| \right) + \sum_{n=1}^{N} \eta_{in}{}^d f_{in}(\boldsymbol{u}_{in}) \right]$$
$$= \sum_{i=1}^{I} \left[-2 \log |\det \boldsymbol{U}_i| + \sum_{n=1}^{N} \left[-2 \log \eta_{in} + \eta_{in}{}^d \right] \right], \quad (20)$$

where $\boldsymbol{U}_i = [\boldsymbol{u}_{i1} \quad \dots \quad \boldsymbol{u}_{iN}]^{\mathsf{H}}$. Therefore, the minimization of the cost function can be interpreted as the minimization of $-\log |\det \boldsymbol{U}_i|$ for each frequency bin and the minimization of $-2 \log \eta_{in} + \eta_{in}{}^d$ for each source and frequency bin. Since the variables in (20) are split into two independent variables (direction vector variable $\boldsymbol{u}_{in}$ and scale variable η_{in}) and their related terms are linearly added, these direction optimization and scale optimization problems are independent of each other. The optimal η_{in} can be calculated using a closed form because the derivative of the cost function w.r.t. η_{in} can be written as

$$\frac{d}{d\eta_{in}}(-2 \log \eta_{in} + \eta_{in}{}^d) = -\frac{2}{\eta_{in}} + d\eta_{in}{}^{d-1}. \quad (21)$$

Hence, by letting the right side of (21) be zero, we can obtain the optimal η_{in} as

$$\eta_{in} = \sqrt[d]{2/d}. \quad (22)$$

The actual difficulty in the optimization of the demixing matrix is the direction optimization, i.e., the minimization of $-2 \log |\det \boldsymbol{U}_i|$. Since minimizing $-\log x^2$ is equivalent to maximizing x^2, we can reformulate this problem as

$$\text{maximize } |\det \boldsymbol{U}_i|^2 \quad \text{s.t. } \forall n, \; f_{in}(\boldsymbol{u}_{in}) = 1. \quad (23)$$

Since it is generally difficult to solve this problem using a closed form, we apply an approach called vectorwise coordinate descent. In this algorithm, we focus on $\boldsymbol{u}_{in}$, namely, the Hermitian transpose of a particular row vector of $\boldsymbol{U}_i$. By cofactor expansion, we can deform problem (23) as

$$\text{maximize } |\boldsymbol{b}_{in}^{\mathsf{H}} \boldsymbol{u}_{in}|^2 \quad \text{s.t. } f_{in}(\boldsymbol{u}_{in}) = 1, \quad (24)$$

where $\boldsymbol{b}_{in}$ is a column vector of the adjugate matrix $\boldsymbol{B}_i = [\boldsymbol{b}_{i1} \quad \dots \quad \boldsymbol{b}_{iN}]^{\mathsf{H}}$ of $\boldsymbol{U}_i$. Since $\boldsymbol{b}_{in}$ only depends on $\boldsymbol{u}_{in'}$
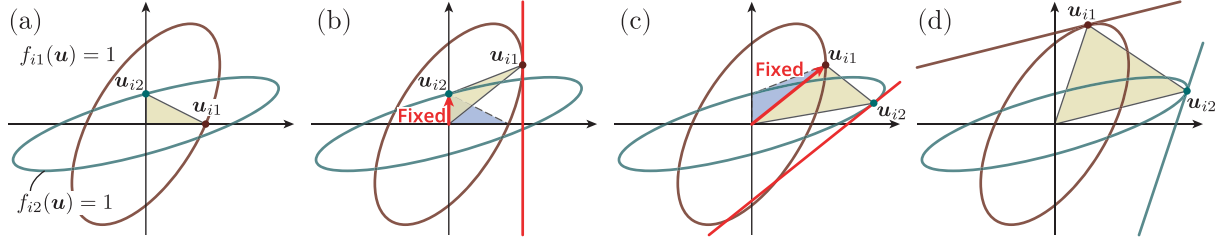
Fig. 5.   Intuitive explanation of GIP-HSM. Via steps (a)–(d), we can maximize yellow area that corresponds to (23).

$(n' \neq n)$ and is independent of $\boldsymbol{u}_{in}$ [24], (24) can be regarded as a function of $\boldsymbol{u}_{in}$ by fixing the other row vectors of $\boldsymbol{U}_i$. Using the method of Lagrange multipliers and C1, we obtain the stationary condition

$$\boldsymbol{b}_{in}(\boldsymbol{b}_{in}^{\mathsf{H}}\boldsymbol{u}_{in}) + \lambda \frac{\partial f_{in}}{\partial \boldsymbol{u}_{in}^{\mathsf{H}}}(\boldsymbol{u}_{in}) = 0, \qquad (25)$$

where λ is a Lagrange multiplier. Since $\boldsymbol{b}_{in}^{\mathsf{H}}\boldsymbol{u}_{in}$ is a scalar, the stationary condition can be rewritten as

$$\frac{\partial f_{in}}{\partial \boldsymbol{u}_{in}^{\mathsf{H}}}(\boldsymbol{u}_{in}) \parallel \boldsymbol{b}_{in}, \qquad (26)$$

where the binary relation "$\boldsymbol{x} \parallel \boldsymbol{y}$" means that $\boldsymbol{x}$ is parallel to $\boldsymbol{y}$. In (26), $\boldsymbol{b}_{in}$ is represented in terms of $\boldsymbol{W}_{in}$ as

$$\begin{aligned}
\boldsymbol{b}_{in} &= (\det \boldsymbol{U}_i)\boldsymbol{U}_i^{-1}\boldsymbol{e}_n \\
&= (\det \boldsymbol{U}_i)(\mathrm{diag}(\eta_{i1}^{-1}, \ldots, \eta_{iN}^{-1})\boldsymbol{W}_i)^{-1}\boldsymbol{e}_n \\
&= (\det \boldsymbol{U}_i)\boldsymbol{W}_i^{-1}\mathrm{diag}(\eta_{i1}, \ldots, \eta_{iN})\boldsymbol{e}_n \\
&= (\eta_{in} \det \boldsymbol{U}_i)\boldsymbol{W}_i^{-1}\boldsymbol{e}_n, \qquad (27)
\end{aligned}$$

where $\mathrm{diag}(c_1, \ldots, c_N)$ denotes the $N \times N$ diagonal matrix whose (n, n)th element is c_n. (27) shows that $\boldsymbol{b}_{in}$ and $\boldsymbol{W}_i^{-1}\boldsymbol{e}_n$ are unique up to scale and thus they are parallel. Since the level set of f_{in} is convex from C2, the stationary point of objective function (24) must also be the optimal point. In addition, (23) is the problem of estimating all the optimal (stationary) points given by (24) for all n. Therefore, cost function (20) that includes (23) monotonically decreases with each update of the direction $\boldsymbol{u}_{in}$.

In conclusion, to minimize cost function (19), we update the vector $\boldsymbol{w}_{in}$ by the following two steps in GIP-HSM: (a) Find a vector $\boldsymbol{w}_{in}'$ that satisfies

$$\frac{\partial f_{in}}{\partial \boldsymbol{w}_{in}^{\mathsf{H}}}(\boldsymbol{w}_{in}') \parallel \boldsymbol{W}_i^{-1}\boldsymbol{e}_n. \qquad (28)$$

(b) Update $\boldsymbol{w}_{in}$ as

$$\boldsymbol{w}_{in} \leftarrow \boldsymbol{w}_{in}' \sqrt[d]{2/(d \cdot f_{in}(\boldsymbol{w}_{in}'))}. \qquad (29)$$

The first step (a) and second step (b) correspond to the direction and scale optimizations, respectively. Note that $\boldsymbol{w}_{in}'$ calculated in the first step does not need to satisfy $f_{in}(\boldsymbol{w}_{in}') = 1$ because the scale is automatically adjusted in the second step. In fact, if $\boldsymbol{w}_{in}'$ is represented as $\boldsymbol{w}_{in}' = \eta_{in}'\boldsymbol{u}_{in}$, the second step results in

$$\boldsymbol{w}_{in} \leftarrow \eta_{in}'\boldsymbol{u}_{in} \cdot \sqrt[d]{2/(d \cdot \eta_{in}'^d)} = \boldsymbol{u}_{in}\sqrt[d]{2/d}, \qquad (30)$$

at which point both the direction and the scale are optimized.

Fig. 5 shows the intuitive explanation of GIP-HSM. On the basis of this figure, we discuss how to solve (23) when $N = 2$. In Fig. 5(a), two constraints, $f_{i1}(\boldsymbol{u}_{i1}) = 1$ and $f_{i2}(\boldsymbol{u}_{i2}) = 1$, are represented with two closed curves. Since $|\det \boldsymbol{U}_i|$ is proportional to the yellow area of the triangle surrounded by the two vectors $\boldsymbol{u}_{i1}$ and $\boldsymbol{u}_{i2}$, the goal is to maximize the area. In GIP-HSM, the vectors $\boldsymbol{u}_{i1}$ and $\boldsymbol{u}_{i2}$ are optimized as follows:
1) Maximize the yellow area by sliding the vector $\boldsymbol{u}_{i1}$ with $\boldsymbol{u}_{i2}$ fixed (see Fig. 5(b)).
2) Maximize the yellow area by sliding the vector $\boldsymbol{u}_{i2}$ with $\boldsymbol{u}_{i1}$ fixed (see Fig. 5(c)).

By iterating 1) and 2) until convergence, we obtain the maximal point, $\boldsymbol{u}_{i1}$ and $\boldsymbol{u}_{i2}$, as shown in Fig. 5(d).

C. Sub-Gaussian ILRMA Based on GIP-HSM

Using GIP-HSM, we propose a new update rule in GGD-ILRMA whose shape parameter β is set to four (time-variant sub-Gaussian model). The cost function of GGD-ILRMA with $\beta = 4$ is written as

$$\mathcal{J} = -2J\sum_i \log|\det \boldsymbol{W}_i| + \sum_{i,j,n} \frac{|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^4}{r_{ijn}^4} + \mathrm{const.}, \quad (31)$$

where the constant term is independent of $\boldsymbol{w}_{in}$. It seems possible to apply GIP-HSM to (31) by letting $f_{in}(\boldsymbol{w}_{in}) = (1/J)\sum_j(|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^4/r_{ijn}^4)$. In this case, however, it is difficult to solve (28), which is reduced to a cubic vector equation w.r.t. $\boldsymbol{w}_{in}'$. Instead, we apply an MM algorithm to derive an update rule that does not contain any cubic vector equations. Hereafter, we prove the following theorem, and then design a new type of majorization function of (31) on the basis of the theorem.

Theorem 1: Let $f(\boldsymbol{w}) = (1/J)\sum_{j=1}^J(|\boldsymbol{w}^{\mathsf{H}}\boldsymbol{x}_j|^4/r_j^4)$ and $g(\boldsymbol{w}) = (\boldsymbol{w}^{\mathsf{H}}\boldsymbol{G}\boldsymbol{w})^2$, where $\boldsymbol{G}$ is defined in terms of a vector $\tilde{\boldsymbol{w}}$ as

$$\boldsymbol{H} = \left[\frac{1}{r_1}\boldsymbol{x}_1 \quad \ldots \quad \frac{1}{r_J}\boldsymbol{x}_J \right], \qquad (32)$$

$$\tilde{\boldsymbol{q}} = \left[\tilde{q}_1 \quad \ldots \quad \tilde{q}_J \right]^{\top} = \boldsymbol{H}^{\mathsf{H}}\tilde{\boldsymbol{w}}, \qquad (33)$$

$$\tilde{\boldsymbol{Q}} = \begin{bmatrix} \|\tilde{\boldsymbol{q}}\|^2 & -\tilde{q}_1\tilde{q}_2^* & \cdots & -\tilde{q}_1\tilde{q}_J^* \\ -\tilde{q}_2\tilde{q}_1^* & \|\tilde{\boldsymbol{q}}\|^2 & \cdots & -\tilde{q}_2\tilde{q}_J^* \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{q}_J\tilde{q}_1^* & -\tilde{q}_J\tilde{q}_2^* & \cdots & \|\tilde{\boldsymbol{q}}\|^2 \end{bmatrix}, \qquad (34)$$

$$G = \frac{1}{\sqrt{J \sum_j |\tilde{q}_j|^4}} H\tilde{Q}H^{\mathsf{H}}, \tag{35}$$

where q^* denotes the complex conjugate of q. Then, $g(\boldsymbol{w})$ satisfies $f(\boldsymbol{w}) \le g(\boldsymbol{w})$ for arbitrary $\boldsymbol{w}$ and the equality holds when $\boldsymbol{w} = \tilde{\boldsymbol{w}}$.

The proof of Theorem 1 is described in Appendix. Applying Theorem 1, we can design a majorization function of (31) as

$$\mathcal{J} \le -2J\sum_i \log|\det \boldsymbol{W}_i| + J\sum_{i,n}(\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{G}_{in}\boldsymbol{w}_{in})^2 + \text{const.}$$

$$=: \mathcal{J}^+, \tag{36}$$

where

$$\boldsymbol{H}_{in} = \begin{bmatrix} \frac{1}{r_{i1n}}\boldsymbol{x}_{i1} & \cdots & \frac{1}{r_{iJn}}\boldsymbol{x}_{iJ} \end{bmatrix}, \tag{37}$$

$$\tilde{\boldsymbol{q}}_{in} = \begin{bmatrix} \tilde{q}_{i1n} & \cdots & \tilde{q}_{iJn} \end{bmatrix}^{\top} = \boldsymbol{H}_{in}^{\mathsf{H}}\tilde{\boldsymbol{w}}_{in}, \tag{38}$$

$$\tilde{\boldsymbol{Q}}_{in} = \begin{bmatrix} \|\tilde{\boldsymbol{q}}_{in}\|^2 & -\tilde{q}_{i1n}\tilde{q}_{i2n}^* & \cdots & -\tilde{q}_{i1n}\tilde{q}_{iJn}^* \\ -\tilde{q}_{i2n}\tilde{q}_{i1n}^* & \|\tilde{\boldsymbol{q}}_{in}\|^2 & \cdots & -\tilde{q}_{i2n}\tilde{q}_{iJn}^* \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{q}_{iJn}\tilde{q}_{i1n}^* & -\tilde{q}_{iJn}\tilde{q}_{i2n}^* & \cdots & \|\tilde{\boldsymbol{q}}_{in}\|^2 \end{bmatrix}, \tag{39}$$

$$\boldsymbol{G}_{in} = \frac{1}{\sqrt{J\sum_j |\tilde{q}_{ijn}|^4}} \boldsymbol{H}_{in}\tilde{\boldsymbol{Q}}_{in}\boldsymbol{H}_{in}^{\mathsf{H}}. \tag{40}$$

Here, $\tilde{\boldsymbol{w}}_{in}$ is an auxiliary variable and the equality of (36) holds when $\boldsymbol{w}_{in} = \tilde{\boldsymbol{w}}_{in}$. Since $g_{in}(\boldsymbol{w}_{in}) = (\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{G}_{in}\boldsymbol{w}_{in})^2$ is a differentiable, convex, and homogeneous function of $\boldsymbol{w}_{in}$, we can apply GIP-HSM to minimize $\mathcal{J}^+$. The optimal condition for the direction of $\boldsymbol{w}_{in}$ is determined as

$$\frac{\partial g_{in}}{\partial \boldsymbol{w}_{in}^{\mathsf{H}}}(\boldsymbol{w}_{in}') = (\boldsymbol{w}_{in}'^{\mathsf{H}}\boldsymbol{G}_{in}\boldsymbol{w}_{in}')\boldsymbol{G}_{in}\boldsymbol{w}_{in}' \parallel \boldsymbol{W}_i^{-1}\boldsymbol{e}_n. \tag{41}$$

Since $(\boldsymbol{w}_{in}'^{\mathsf{H}}\boldsymbol{G}_{in}\boldsymbol{w}_{in}')$ is a scalar, one of the solutions of (41) is

$$\boldsymbol{w}_{in}' = \boldsymbol{G}_{in}^{-1}\boldsymbol{W}_i^{-1}\boldsymbol{e}_n. \tag{42}$$

Substituting $\tilde{\boldsymbol{w}}_{in} = \boldsymbol{w}_{in}$ into (38)–(40), we obtain the following update rules for optimizing the direction of $\boldsymbol{w}_{in}$:

$$\boldsymbol{H}_{in} = \begin{bmatrix} \frac{1}{r_{i1n}}\boldsymbol{x}_{i1} & \cdots & \frac{1}{r_{iJn}}\boldsymbol{x}_{iJ} \end{bmatrix}, \tag{43}$$

$$\boldsymbol{q}_{in} = \begin{bmatrix} q_{i1n} & \cdots & q_{iJn} \end{bmatrix}^{\top} = \boldsymbol{H}_{in}^{\mathsf{H}}\boldsymbol{w}_{in}, \tag{44}$$

$$\boldsymbol{Q}_{in} = \begin{bmatrix} \|\boldsymbol{q}_{in}\|^2 & -q_{i1n}q_{i2n}^* & \cdots & -q_{i1n}q_{iJn}^* \\ -q_{i2n}q_{i1n}^* & \|\boldsymbol{q}_{in}\|^2 & \cdots & -q_{i2n}q_{iJn}^* \\ \vdots & \vdots & \ddots & \vdots \\ -q_{iJn}q_{i1n}^* & -q_{iJn}q_{i2n}^* & \cdots & \|\boldsymbol{q}_{in}\|^2 \end{bmatrix}, \tag{45}$$

$$\boldsymbol{G}_{in} = \frac{1}{\sqrt{J\sum_j |q_{ijn}|^4}} \boldsymbol{H}_{in}\boldsymbol{Q}_{in}\boldsymbol{H}_{in}^{\mathsf{H}}, \tag{46}$$

$$\boldsymbol{w}_{in} \leftarrow \boldsymbol{G}_{in}^{-1}\boldsymbol{W}_i^{-1}\boldsymbol{e}_n. \tag{47}$$

Finally, we operate the following scale optimization by applying (29):

$$\boldsymbol{q}_{in} = \begin{bmatrix} q_{i1n} & \cdots & q_{iJn} \end{bmatrix}^{\top} = \boldsymbol{H}_{in}^{\mathsf{H}}\boldsymbol{w}_{in}, \tag{48}$$

$$\boldsymbol{w}_{in} \leftarrow \boldsymbol{w}_{in} \sqrt[4]{J/(2\sum_j |q_{ijn}|^4)}, \tag{49}$$

which is the scale optimization w.r.t. f_{in}.

In GGD-ILRMA with $\beta = 4$, the demixing matrix $\boldsymbol{W}_i$ is updated using (43)–(49) with the low-rank source models $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$ fixed, and then the source models $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$ are updated using (16) and (17) with the demixing matrix $\boldsymbol{W}_i$ fixed; these optimization steps are alternately repeated. Since these update rules are derived using the MM algorithm and GIP-HSM and all the variables $\boldsymbol{W}_i, \boldsymbol{T}_n,$ and $\boldsymbol{V}_n$ are optimized on the basis of a coordinate descent algorithm, a monotonic decrease in cost function (31) is guaranteed as in conventional ILRMA.

Fig. 6 shows the intuitive explanation of the proposed update of $\boldsymbol{W}_i$ using GIP-HSM and the MM algorithm for (31). In Fig. 6(a), although the goal is to maximize the yellow area, it is difficult to find the optimal $\boldsymbol{u}_{i1}$ (red point) in a closed form. Instead, we design a majorization function (the square of the quadratic term) like Fig. 6(b) and maximize the yellow area. Since the level set of the majorization function is an ellipse, the optimal condition can be found as that shown in Fig. 6(c) by solving a linear equation. The update of $\boldsymbol{u}_{i1}$ finishes after aligning the scale of $\boldsymbol{u}_{i1}$ (see Fig. 6(d)), which becomes close to the optimal $\boldsymbol{u}_{i1}$.

D. Computationally Efficient Update Rule

From a practical viewpoint, the update rules (43)–(49) cost a large amount of calculation compared with IS-ILRMA. This is because we have to calculate (46), which contains the multiplication of $N \times J$, $J \times J$ and $J \times N$ matrices (computational cost: $O(NJ^2)$). This cost can be reduced by focusing on the behavior of $\boldsymbol{Q}_{in}$. The matrix $\boldsymbol{Q}_{in}$ can be rewritten using $\boldsymbol{q}_{in}$ as

$$\boldsymbol{Q}_{in} = \|\boldsymbol{q}_{in}\|^2 \boldsymbol{I} + \text{diag}(q_{i1n}, \ldots, q_{iJn}) - \boldsymbol{q}_{in}\boldsymbol{q}_{in}^{\mathsf{H}}. \tag{50}$$

Therefore, $\boldsymbol{G}_{in}$ can be calculated as

$$\boldsymbol{G}_{in} = \frac{1}{\sqrt{J\sum_j |q_{ijn}|^4}} \left(\|\boldsymbol{q}_{in}\|^2 \boldsymbol{H}_{in}\boldsymbol{H}_{in}^{\mathsf{H}} + \sum_j \frac{q_{ijn}}{r_{ijn}^2}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}} \right.$$
$$\left. - (\boldsymbol{H}_{in}\boldsymbol{q}_{in})(\boldsymbol{H}_{in}\boldsymbol{q}_{in})^{\mathsf{H}} \right), \tag{51}$$

and its computational cost is $O(N^2 J)$. Since the length of the time frame, J, is much greater than the number of sources, N, under practical conditions, the high computational cost can be reduced by applying (51) to the update rules instead of (45) and (46).
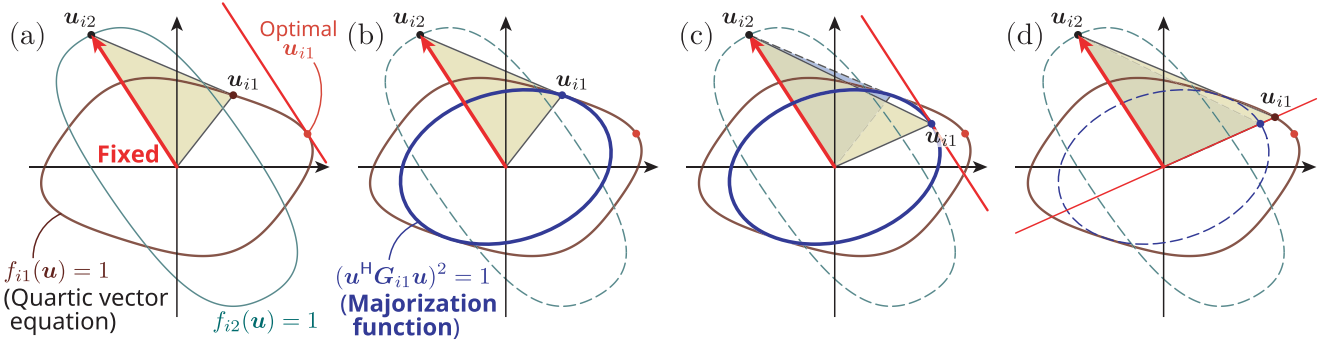
Fig. 6.   Intuitive explanation of MM-algorithm-based update rule in sub-Gaussian GGD-ILRMA with GIP-HSM.

E. Sub-Gaussian GGD-ILRMA With $2 < \beta < 4$

Hereafter, we discuss the minimization of the cost function of GGD-ILRMA

$$\mathcal{I} = -2J \sum_i \log |\det \boldsymbol{W}_i| + \sum_{i,j,n} \frac{|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^{\beta}}{r_{ijn}^{\beta}} + \text{const.} \quad (52)$$

with $2 < \beta < 4$. As noted in Section III-A, the majorization function of the cost function of conventional GGD-ILRMA for $0 < \beta < 2$ is designed on the basis of weighted AM–GM inequality, which bounds $|y|^{\beta}$ $(0 < \beta < 2)$ by $|y|^2$. Similarly, we can bound the term $|y|^{\beta}$ $(2 < \beta < 4)$ by $|y|^4$ using weighted AM–GM inequality and design the majorization function of (52), to which the proposed update rule can be applied. We apply the weighted AM–GM inequality

$$|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^{\beta} \leq \frac{\beta}{4} \frac{|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^4}{\gamma_{ijn}^{4-\beta}} + \left(1 - \frac{\beta}{4}\right) \gamma_{ijn}^{\beta} \quad (53)$$

to (52), where γ_{ijn} is an auxiliary variable and the equality of (53) holds when $\gamma_{ijn} = |y_{ijn}|$. Then, the majorization function of (52) can be designed as

$$\mathcal{I} \leq \sum_i \left[-2J \log |\det \boldsymbol{W}_i| + \sum_{n=1}^{N} \sum_{j=1}^{J} \frac{\beta}{4} \frac{|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^4}{\gamma_{ijn}^{4-\beta} r_{ijn}^{\beta}} \right]$$

$$+ \text{const.}$$

$$=: \mathcal{I}^+. \quad (54)$$

We can apply the combined technique of the MM algorithm and GIP-HSM proposed in Section IV-C to minimizing (54) and obtain the following update rules:

$$l_{ijn} = \sqrt[4]{|y_{ijn}|^{4-\beta} r_{ijn}^{\beta}}, \quad (55)$$

$$\boldsymbol{H}_{in} = \left[\frac{1}{l_{i1n}} \boldsymbol{x}_{i1} \quad \cdots \quad \frac{1}{l_{iJn}} \boldsymbol{x}_{iJ} \right], \quad (56)$$

$$\boldsymbol{q}_{in} = \left[q_{i1n} \quad \cdots \quad q_{iJn} \right]^{\top} = \boldsymbol{H}_{in}^{\mathsf{H}} \boldsymbol{w}_{in}, \quad (57)$$

$$\boldsymbol{Q}_{in} = \begin{bmatrix} \|\boldsymbol{q}_{in}\|^2 & -q_{i1n} q_{i2n}^* & \cdots & -q_{i1n} q_{iJn}^* \\ -q_{i2n} q_{i1n}^* & \|\boldsymbol{q}_{in}\|^2 & \cdots & -q_{i2n} q_{iJn}^* \\ \vdots & \vdots & \ddots & \vdots \\ -q_{iJn} q_{i1n}^* & -q_{iJn} q_{i2n}^* & \cdots & \|\boldsymbol{q}_{in}\|^2 \end{bmatrix}, \quad (58)$$

$$\boldsymbol{G}_{in} = \frac{\sqrt{\beta}}{2\sqrt{J \sum_j |q_{ijn}|^4}} \boldsymbol{H}_{in} \boldsymbol{Q}_{in} \boldsymbol{H}_{in}^{\mathsf{H}}, \quad (59)$$

$$\boldsymbol{w}_{in} \leftarrow \boldsymbol{G}_{in}^{-1} \boldsymbol{W}_i^{-1} \boldsymbol{e}_n, \quad (60)$$

$$\boldsymbol{w}_{in} \leftarrow \boldsymbol{w}_{in} \left(\frac{2J}{\beta \sum_j |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^{\beta} / r_{ijn}^{\beta}} \right)^{1/\beta}. \quad (61)$$

Note that we can also apply the update rules (55)–(61) to the $0 < \beta \leq 2$ case as well as the $2 < \beta \leq 4$ case. However, these update rules seem to be slower than the conventional update rule. The conventional method is derived with the majorization function including the quadratic form w.r.t. $\boldsymbol{w}_{in}$, whereas the proposed method is derived with that including the quartic form w.r.t. $\boldsymbol{w}_{in}$, which is far apart from the original function compared with that in the conventional method.

V. MULTIVARIATE ME-ALGORITHM-BASED UPDATE RULE FOR DEMIXING MATRIX

A. Motivation

In Sections IV-C and IV-E, we derived the convergence-guaranteed update rule of $\boldsymbol{W}_i$ in GGD-ILRMA with $2 < \beta \leq 4$ by the MM algorithm. However, from our experience, this rule is not as fast as that of conventional IS-ILRMA, which does not use the MM algorithm for derivation. To accelerate the convergence, in this section, we introduce a new ME algorithm, which is experimentally confirmed to be faster than the MM algorithm. Note that the basic principle of the original ME algorithm has been proposed in [22], but this literature presented only the application to NMF dealing with each scalar variable. The ME algorithm proposed in this paper, however, is the first attempt at updating the vector $\boldsymbol{w}_{in}$ directly to the best of our knowledge. This corresponds to a *multivariate* case and is generally difficult to solve, as described in Section V-C later.

B. Derivation of ME-Algorithm-Based Update Rule

The ME algorithm was proposed as an update algorithm based on a majorization function. In the MM algorithm, the solution is updated to the point that minimizes the majorization function. In the ME algorithm, on the other hand, the solution is updated to the point in which the value of the majorization function is equal

to that in the initial point. Since the cost function never increases during the update in the ME algorithm, the convergence of the cost function is also guaranteed. Let $\mathcal{Q}(\theta)$ be a cost function, θ a set of variables, $\mathcal{Q}^+(\theta, \tilde{\theta})$ a majorization function of $\mathcal{Q}(\theta)$ ($\mathcal{Q}^+(\theta, \tilde{\theta}) \geq \mathcal{Q}(\theta)$ holds for all θ and $\tilde{\theta}$), which touches with $\mathcal{Q}(\theta)$ at $\tilde{\theta} = \tilde{\theta}_0$, and $\tilde{\theta}$ a set of auxiliary variables. In the ME algorithm, the initial point $\theta = \theta_0$ is updated to $\theta = \theta_1$ that satisfies $\mathcal{Q}^+(\theta_0, \tilde{\theta}_0) = \mathcal{Q}^+(\theta_1, \tilde{\theta}_0)$. Then, the following inequality holds:

$$\mathcal{Q}(\theta_0) = \mathcal{Q}^+(\theta_0, \tilde{\theta}_0) = \mathcal{Q}^+(\theta_1, \tilde{\theta}_0) \geq \mathcal{Q}(\theta_1). \quad (62)$$

This proves the non-increase property in the ME algorithm. Although it is difficult to mathematically prove the advantage of the ME algorithm compared with the MM algorithm, by numerical experiments for several applications, it has been shown that the update rule by the ME algorithm tends to be faster than that by the MM algorithm [22].

We propose a faster update rule of the proposed sub-Gaussian GGD-ILRMA by applying the ME algorithm. In a multivariate case, the up-to-date variables that satisfy the ME condition (62) innumerably exist, and it is difficult to determine the optimal one. However, we discovered that the following update rule satisfies the ME condition and provides a relatively good update for the minimization of (19), which will be discussed in Section V-C. Let $\tilde{u}_{in}$ be the direction of a demixing filter w_{in} before the update and u_{in}^{MM} be the direction of w_{in} after the update by the MM algorithm. In the ME algorithm, we update the direction of w_{in} to

$$u_{in}^{\mathrm{ME}} = 2 \frac{u_{in}^{\mathrm{MM}\mathsf{H}} G_{in} \tilde{u}_{in}}{u_{in}^{\mathrm{MM}\mathsf{H}} G_{in} u_{in}^{\mathrm{MM}}} u_{in}^{\mathrm{MM}} - \tilde{u}_{in}. \quad (63)$$

The vector u_{in}^{ME} is a linear combination of $\tilde{u}_{in}$ and u_{in}^{MM}. The following theorem shows that the update rule based on (63) satisfies the requirement for the ME algorithm:

Theorem 2: When u_{in} is updated to $\eta_{in} u_{in}^{\mathrm{ME}}$, where u_{in}^{ME} is defined by (63), the values of the majorization function are identical before and after the update, i.e.,

$$\mathcal{J}^+|_{w_{in} = \eta_{in} u_{in}^{\mathrm{ME}}} = \mathcal{J}^+|_{w_{in} = \eta_{in} \tilde{u}_{in}}, \quad (64)$$

where (64) holds for any value of the scale η_{in}.

Proof: Let $\kappa_1 = u_{in}^{\mathrm{MM}\mathsf{H}} G_{in} u_{in}^{\mathrm{MM}}$ and $\kappa_2 = u_{in}^{\mathrm{MM}\mathsf{H}} G_{in} \tilde{u}_{in}$. From (36) and $\det U_{in} = b_{in}^{\mathsf{H}} u_{in}$, it is sufficient to prove the following two equations:

$$-\log \left| b_{in}^{\mathsf{H}} u_{in}^{\mathrm{ME}} \right|^2 = -\log \left| b_{in}^{\mathsf{H}} \tilde{u}_{in} \right|^2, \quad (65)$$

$$\left(u_{in}^{\mathrm{ME}\mathsf{H}} G_{in} u_{in}^{\mathrm{ME}} \right)^2 = (\tilde{u}_{in}^{\mathsf{H}} G_{in} \tilde{u}_{in})^2. \quad (66)$$

First, we prove (65). From (27) and (60), we have

$$u_{in}^{\mathrm{MM}} = \xi G_{in}^{-1} b_{in}, \quad (67)$$

where ξ is a complex number. Then,

$$b_{in}^{\mathsf{H}} u_{in}^{\mathrm{ME}} = \frac{1}{\xi^*} u_{in}^{\mathrm{MM}\mathsf{H}} G_{in} \left(2 \frac{\kappa_2}{\kappa_1} u_{in}^{\mathrm{MM}} - \tilde{u}_{in} \right)$$

$$= \frac{1}{\xi^*} \left(2 \frac{\kappa_2}{\kappa_1} \kappa_1 - \kappa_2 \right) = \frac{\kappa_2}{\xi^*} = b_{in}^{\mathsf{H}} \tilde{u}_{in}. \quad (68)$$

Therefore, (65) holds. Next, we prove (66). The following relation holds:

$$u_{in}^{\mathrm{ME}\mathsf{H}} G_{in} u_{in}^{\mathrm{ME}} - \tilde{u}_{in}^{\mathsf{H}} G_{in} \tilde{u}_{in}$$

$$= \left(2 \frac{\kappa_2}{\kappa_1} u_{in}^{\mathrm{MM}} - \tilde{u}_{in} \right)^{\mathsf{H}} G_{in} \left(2 \frac{\kappa_2}{\kappa_1} u_{in}^{\mathrm{MM}} - \tilde{u}_{in} \right)$$

$$\quad - \tilde{u}_{in}^{\mathsf{H}} G_{in} \tilde{u}_{in}$$

$$= 4 \frac{\kappa_2 \kappa_2^*}{\kappa_1^2} \kappa_1 - 2 \frac{\kappa_2}{\kappa_1} \kappa_2^* - 2 \frac{\kappa_2^*}{\kappa_1} \kappa_2 = 0. \quad (69)$$

Therefore, we obtain (66). ■

By inserting (63) into (55)–(61), we obtain the following ME-algorithm-based update rules of GGD-ILRMA with $2 < \beta \leq 4$:

$$l_{ijn} = \sqrt[4]{|y_{ijn}|^{4-\beta} r_{ijn}^{\beta}}, \quad (70)$$

$$H_{in} = \left[\frac{1}{l_{i1n}} x_{i1} \quad \cdots \quad \frac{1}{l_{iJn}} x_{iJ} \right], \quad (71)$$

$$q_{in} = \left[q_{i1n} \quad \cdots \quad q_{iJn} \right]^{\top} = H_{in}^{\mathsf{H}} w_{in}, \quad (72)$$

$$Q_{in} = \begin{bmatrix} \|q_{in}\|^2 & -q_{i1n} q_{i2n}^* & \cdots & -q_{i1n} q_{iJn}^* \\ -q_{i2n} q_{i1n}^* & \|q_{in}\|^2 & \cdots & -q_{i2n} q_{iJn}^* \\ \vdots & \vdots & \ddots & \vdots \\ -q_{iJn} q_{i1n}^* & -q_{iJn} q_{i2n}^* & \cdots & \|q_{in}\|^2 \end{bmatrix}, \quad (73)$$

$$G_{in} = \frac{\sqrt{\beta}}{2 \sqrt{J \sum_j |q_{ijn}|^4}} H_{in} Q_{in} H_{in}^{\mathsf{H}}, \quad (74)$$

$$w_{in}^{\mathrm{MM}} \leftarrow G_{in}^{-1} W_i^{-1} e_n, \quad (75)$$

$$w_{in} \leftarrow 2 \frac{w_{in}^{\mathrm{MM}\mathsf{H}} G_{in} w_{in}}{w_{in}^{\mathrm{MM}\mathsf{H}} G_{in} w_{in}^{\mathrm{MM}}} w_{in}^{\mathrm{MM}} - w_{in}, \quad (76)$$

$$w_{in} \leftarrow w_{in} \left(\frac{2J}{\beta \sum_j |w_{in}^{\mathsf{H}} x_{ij}|^{\beta} / r_{ijn}^{\beta}} \right)^{1/\beta}. \quad (77)$$

Note that (76) can be obtained by multiplying η_{in} to the both sides of (63). First, we calculate a temporal vector w_{in}^{MM}, which is the direction of the optimal point in $\mathcal{I}^+$, with (70)–(75). Second, we obtain the ME-algorithm-updated vector with (76). Finally, the scale of w_{in} is normalized with (77).

Compared with the MM-algorithm-based update rule, only (76) has to be calculated additionally. We can expect some acceleration of the update of W_i with low additional computational costs because some other calculations (e.g., (73)–(75)) are dominant.

C. Comparison Between MM and ME Algorithms

As mentioned above, Figs. 6(b) and (c) are the intuitive explanation of the update based on the MM algorithm. Fig. 7 shows the difference between the directions of the vector after the update of the MM and ME algorithms. As shown in this figure, in the case of $N = 2$, the angle between $\tilde{u}_{i1}$ and u_{i1}^{ME} tends to be greater than that between $\tilde{u}_{i1}$ and u_{i1}^{MM}. This implies that the proposed ME algorithm can be expected to converge with a smaller number of iterations than the MM algorithm.
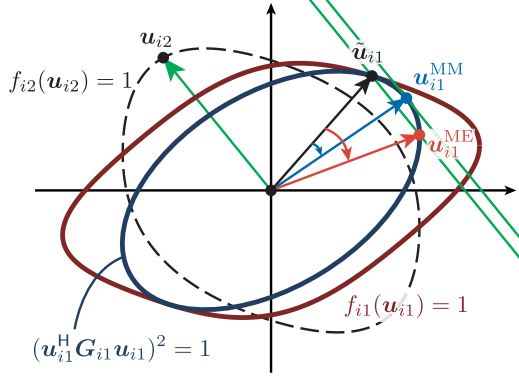
Fig. 7. Intuitive explanation of difference between ME and MM algorithms.

Note that there are innumerable points that satisfy the equalization condition when N is greater than or equal to three. Since the set of these points corresponds to the intersection of the hyperellipsoid (66) and the hyperplane (65), some points satisfying the condition are in the neighborhood of the initial point and thus useless. To realize a more effective update than the MM algorithm, we need to avoid a point closer to the initial point than that chosen in the MM algorithm.

Indeed, the proposed ME-algorithm-based update (63) corresponds to one of the above-mentioned innumerable points, but the following interpretation indicates that (63) tends to choose a direction whose variation of angle is greater than that in the MM-algorithm-based update.

Let $\phi \in [0, \pi]$ be the angle between two direction vectors $\boldsymbol{u}_1 \in \mathbb{C}^M$ and $\boldsymbol{u}_2 \in \mathbb{C}^M$, and the following equation holds:

$$\cos \phi = \frac{\mathrm{Re}[\boldsymbol{u}_1^{\mathsf{H}} \boldsymbol{u}_2]}{\|\boldsymbol{u}_1\|\|\boldsymbol{u}_2\|}, \tag{78}$$

where $\mathrm{Re}[\cdot]$ denotes the real part of an input complex value. Since the demixing filter has ambiguity w.r.t. phase rotation, we redefine the angle between $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ as the minimum value of possible angles as

$$\cos \phi = \frac{|\boldsymbol{u}_1^{\mathsf{H}} \boldsymbol{u}_2|}{\|\boldsymbol{u}_1\|\|\boldsymbol{u}_2\|}, \tag{79}$$

where the range of ϕ becomes $\phi \in [0, \pi/2]$.

We here analyze the condition that satisfies $\phi_{in}^{\mathrm{ME}} \geq \phi_{in}^{\mathrm{MM}}$, where ϕ_{in}^{ME} and ϕ_{in}^{MM} are the angles between $\tilde{\boldsymbol{u}}_{in}$ and $\boldsymbol{u}_{in}^{\mathrm{ME}}$ and between $\tilde{\boldsymbol{u}}_{in}$ and $\boldsymbol{u}_{in}^{\mathrm{MM}}$, respectively. From (63) and (79), we obtain

$$\cos^2 \phi_{in}^{\mathrm{MM}} - \cos^2 \phi_{in}^{\mathrm{ME}}$$
$$= \frac{|\tilde{\boldsymbol{u}}_{in}^{\mathsf{H}} \boldsymbol{u}_{in}^{\mathrm{MM}}|^2}{\|\tilde{\boldsymbol{u}}_{in}\|^2 \|\boldsymbol{u}_{in}^{\mathrm{MM}}\|^2} - \frac{|\tilde{\boldsymbol{u}}_{in}^{\mathsf{H}} \boldsymbol{u}_{in}^{\mathrm{ME}}|^2}{\|\tilde{\boldsymbol{u}}_{in}\|^2 \|\boldsymbol{u}_{in}^{\mathrm{ME}}\|^2}$$
$$= \frac{|\tilde{\boldsymbol{u}}_{in}^{\mathsf{H}} \boldsymbol{u}_{in}^{\mathrm{MM}}|^2}{\|\tilde{\boldsymbol{u}}_{in}\|^2 \|\boldsymbol{u}_{in}^{\mathrm{MM}}\|^2} - \frac{|\tilde{\boldsymbol{u}}_{in}^{\mathsf{H}} \left(2\frac{\kappa_2}{\kappa_1} \boldsymbol{u}_{in}^{\mathrm{MM}} - \tilde{\boldsymbol{u}}_{in} \right)|^2}{\|\tilde{\boldsymbol{u}}_{in}\|^2 \|2\frac{\kappa_2}{\kappa_1} \boldsymbol{u}_{in}^{\mathrm{MM}} - \tilde{\boldsymbol{u}}_{in}\|^2}$$
$$= \frac{\left(\frac{4\mathrm{Re}[\kappa_2 \tilde{\boldsymbol{u}}_{in}^{\mathsf{H}} \boldsymbol{u}_{in}^{\mathrm{MM}}]}{\kappa_1} - \|\tilde{\boldsymbol{u}}_{in}\|^2 \right)(\|\tilde{\boldsymbol{u}}_{in}\|^2 \|\boldsymbol{u}_{in}^{\mathrm{MM}}\|^2 - |\tilde{\boldsymbol{u}}_{in}^{\mathsf{H}} \boldsymbol{u}_{in}^{\mathrm{MM}}|^2)}{\|\tilde{\boldsymbol{u}}_{in}\|^2 \|\boldsymbol{u}_{in}^{\mathrm{MM}}\|^2 \|2\frac{\kappa_2}{\kappa_1} \boldsymbol{u}_{in}^{\mathrm{MM}} - \tilde{\boldsymbol{u}}_{in}\|^2}. \tag{80}$$

Signal	Parts	Instruments (1/2)
Music 1	Bass/Melody 1	Bassoon/oboe
Music 2	Midrange/Melody 1	Piano/oboe
Music 3	Melody 1/Melody 2	Oboe/flute
Music 4	Bass/Melody 1	Bassoon/trumpet
Music 5	Midrange/Melody 1	Piano/trumpet
Music 6	Melody 1/Melody 2	Trumpet/flute

Since

$$\|\tilde{\boldsymbol{u}}_{in}\|^2 \|\boldsymbol{u}_{in}^{\mathrm{MM}}\|^2 - |\tilde{\boldsymbol{u}}_{in}^{\mathsf{H}} \boldsymbol{u}_{in}^{\mathrm{MM}}|^2$$
$$= \|\tilde{\boldsymbol{u}}_{in}\|^2 \|\boldsymbol{u}_{in}^{\mathrm{MM}}\|^2 (1 - \cos^2 \phi_{in}^{\mathrm{MM}})$$
$$\geq 0,$$

we obtain the condition that satisfies $\phi_{in}^{\mathrm{ME}} \geq \phi_{in}^{\mathrm{MM}}$ as follows:

$$\frac{4\mathrm{Re}[\kappa_2 \tilde{\boldsymbol{u}}_{in}^{\mathsf{H}} \boldsymbol{u}_{in}^{\mathrm{MM}}]}{\kappa_1} - \|\tilde{\boldsymbol{u}}_{in}\|^2 \geq 0$$
$$\Leftrightarrow \frac{\mathrm{Re}[\kappa_2 \tilde{\boldsymbol{u}}_{in}^{\mathsf{H}} \boldsymbol{u}_{in}^{\mathrm{MM}}]}{\kappa_1 \|\tilde{\boldsymbol{u}}_{in}\|^2} \geq \frac{1}{4}$$
$$\Leftrightarrow \frac{\mathrm{Re}[\boldsymbol{u}_{in}^{\mathrm{MM}\mathsf{H}} \boldsymbol{G}_{in} \tilde{\boldsymbol{u}}_{in} \tilde{\boldsymbol{u}}_{in}^{\mathsf{H}} \boldsymbol{u}_{in}^{\mathrm{MM}}]}{\boldsymbol{u}_{in}^{\mathrm{MM}\mathsf{H}} \boldsymbol{G}_{in} \boldsymbol{u}_{in}^{\mathrm{MM}} \|\tilde{\boldsymbol{u}}_{in}\|^2} \geq \frac{1}{4}. \tag{81}$$

When $\boldsymbol{u}_{in}^{\mathrm{MM}}$ is equal to an eigenvector of $\boldsymbol{G}_{in}$, the condition (81) can be rewritten as

$$\frac{\zeta |\tilde{\boldsymbol{u}}_{in}^{\mathsf{H}} \boldsymbol{u}_{in}^{\mathrm{MM}}|^2}{\zeta \|\boldsymbol{u}_{in}^{\mathrm{MM}}\|^2 \|\tilde{\boldsymbol{u}}_{in}\|^2} \geq \frac{1}{4}$$
$$\Leftrightarrow \cos^2 \phi_{in}^{\mathrm{MM}} \geq \frac{1}{4}, \tag{82}$$

where $\zeta > 0$ is the eigenvalue of $\boldsymbol{G}_{in}$ that corresponds to $\boldsymbol{u}_{in}^{\mathrm{MM}}$. Thus, $\phi_{in}^{\mathrm{ME}} \geq \phi_{in}^{\mathrm{MM}}$ is satisfied when $\phi_{in}^{\mathrm{MM}} \leq \pi/3$. Although $\boldsymbol{u}_{in}^{\mathrm{MM}}$ does not always become equal to the eigenvector of $\boldsymbol{G}_{in}$, $\boldsymbol{u}_{in}^{\mathrm{MM}}$ is generally close to the eigenvector of $\boldsymbol{G}_{in}$ that has the minimum eigenvalue when the sources are sufficiently separated. In addition, the angle between $\tilde{\boldsymbol{u}}_{in}$ and $\boldsymbol{u}_{in}^{\mathrm{MM}}$ becomes small around the converged point, i.e., $\phi_{in}^{\mathrm{MM}} \leq \pi/3$ probably holds. For these reasons, we can expect that the proposed ME algorithm is faster than the MM algorithm. This issue will experimentally be confirmed in the next section.

VI. EXPERIMENTAL EVALUATION

A. Datasets

We prepared two types of signal for evaluation: music and speech signals. For the music signals, we artificially produced monaural dry music sources of four melody parts (melody 1, main melody; melody 2, counter melody; midrange; and bass) using Microsoft GS Wavetable Synth, where several musical instruments were chosen to play these melody parts [25], [26]. Scores we used for generating the sources are shown in Fig. 8. For the speech signals, we used the SiSEC2011 dataset [27]. Tables I and II show the pairs of music and speech sources used

Fig. 8. Scores of each part used in music-signal separation. Length of dry source is 4 s.

TABLE II
PAIRS OF DRY SOURCES USED IN TWO-SOURCE
SEPARATION OF SPEECH SIGNALS

Signal	Data name	Sources (1/2)
Speech 1	dev1_female4/dev1_female4	src_1/src_2
Speech 2	dev1_female4/dev1_female4	src_3/src_4
Speech 3	dev1_male4/dev1_male4	src_1/src_2
Speech 4	dev1_male4/dev1_male4	src_3/src_4
Speech 5	dev1_male4/dev1_female4	src_1/src_2
Speech 6	dev1_male4/dev1_female4	src_3/src_4

TABLE III
TRIPLETS OF DRY SOURCES USED IN THREE-SOURCE
SEPARATION OF MUSIC SIGNALS

Signal	Parts	Instruments (1/2/3)
Music 1	Bass/Melody 1/Melody 2	Bassoon/trumpet/flute
Music 2	Bass/Melody 1/Melody 2	Bassoon/oboe/flute
Music 3	Midrange/Melody 1/Melody 2	Piano/trumpet/flute
Music 4	Midrange/Melody 1/Melody 2	Piano/oboe/flute

in a two-source separation experiment, respectively. Tables III and IV show the triplets of music and speech sources used in a three-source separation experiment, respectively.

B. Comparison Between MM and ME Algorithms

We compared the convergence speed between the MM-algorithm-based and ME-algorithm-based update rules in the proposed sub-Gaussian GGD-ILRMA ($\beta = 4$) by conducting a two-source separation experiment. In this experiment, 12 pairs of signals were used: six types of music signal (shown in Table I) and six types of speech signal (shown in Table II). To simulate reverberant mixtures, the observed signals were produced by convolving two types of impulse response, IR1 and IR2, which were obtained from the RWCP database [28], as shown in Figs. 9(a) and (b). As the evaluation score, we used the improvement of the source-to-distortion ratio (SDR) [29], which indicates the overall separation quality. An STFT was performed using a 128-ms-long Hamming window with a 64-ms-long shift in music signals and a 256-ms-long Hamming window with a 128-ms-long shift in speech signals. The other conditions are shown in Table V.

Fig. 10 shows the average SDR improvements for all sources and recording conditions, where error bars depict standard errors for each averaged value. There is a negligible difference

in the final score between the MM-algorithm-based and ME-algorithm-based updates. However, the ME algorithm converges with a smaller number of iterations than the MM algorithm. For example, the MM-algorithm-based update takes 690 iterations to achieve the separation performance of $6 dB$, whereas the ME-algorithm-based update takes only 420 iterations to achieve the same performance. Therefore, the ME-algorithm-based update rule reduces the number of iterations by about 39% compared with the MM-algorithm-based update rule. Considering this finding, hereafter, we only show the performance of the proposed method with the ME algorithm.

C. BSS Experiment on Two-Source Signals

We compared the separation performance of the proposed sub-Gaussian GGD-ILRMA ($\beta = 3$ and 4) with those of conventional methods: frequency-domain ICA [4] with a sub-Gaussian model (sub-Gauss ICA), IVA [9], MNMF [18], IS-ILRMA [10], and super-Gaussian GGD-ILRMA [12]. We used the same signals as Section VI-B for evaluation. The shape parameter β in GGD-ILRMA was set to 1, 1.99, and 2 for conventional super-Gaussian GGD-ILRMA, and to 3 and 4 for the proposed sub-Gaussian GGD-ILRMA. The case of $\beta = 2$ is identical to that in IS-ILRMA [10], and the case of $\beta = 1.99$ is the best parameter in a previous study [12]. In sub-Gauss ICA, we optimized the demixing matrix by using a natural-gradient-based update algorithm [4]. Its contrast function was set to the most typical one for a global (time-*invariant*) sub-Gaussian source model [30], which was extended to a complex-valued case, i.e., $(\mathrm{Re}[z] - \tanh(\mathrm{Re}[z])) + \mathrm{j}(\mathrm{Im}[z] - \tanh(\mathrm{Im}[z]))$, where z is a complex variable, $\mathrm{Im}[z]$ denotes the imaginary part of z, and j is the imaginary unit. To solve the source-permutation problem, we employed an interfrequency-correlation-based method [31] that does not utilize information on the position of microphones as in the other methods. In MNMF, we used the same initial values as the original study in [18]. We used the same experimental conditions as those in Section VI-B (shown in Table V).

Figs. 11–14 show the experimental results in two-source separation, where error bars depict standard errors for each averaged value. Figs. 11 and 12 show the average of six music signals in the recording conditions IR1 and IR2, respectively. Figs. 13 and 14 show the average of six speech signals in the recording conditions IR1 and IR2, respectively.

First of all, we compare the performances inside the GGD-ILRMA family. From Figs. 11 and 12, we can confirm that sub-Gaussian ILRMAs ($\beta = 3$ and $\beta = 4$) provide more than 1.5 dB SDR improvement compared with Gaussian ILRMA ($\beta = 2$). This is because the proposed source model, namely, the sub-Gaussian source model, is more appropriate than the other conventional super-Gaussian source models. On the other hand, super-Gaussian ILRMAs ($\beta = 1$ and $\beta = 1.99$) also slightly outperform Gaussian ILRMA, although the improvements are less than 0.5 dB. This improvement might be provided by the low-rank property of music signals; by setting β to a small value, low-rankness is injected into the source model (NMF) of GGD-ILRMA [12], which is a desirable property for the music signals used in this experiment. Furthermore, from Figs. 13

TABLE IV
TRIPLETS OF DRY SOURCES USED IN THREE-SOURCE SEPARATION OF SPEECH SIGNALS

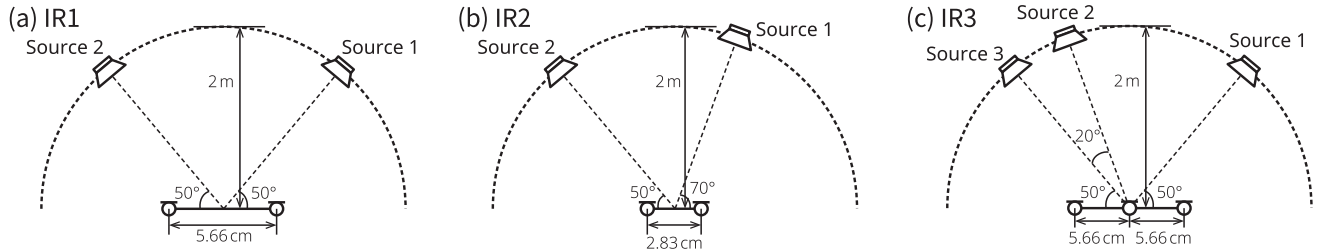| Signal	Data name	Sources (1/2/3)
Speech 1	dev1_male4/dev1_male4/dev1_male4	src_1/src_2/src_3
Speech 2	dev1_female4/dev1_female4/dev1_female4	src_1/src_2/src_3
Speech 3	dev1_male4/dev1_male4/dev1_female4	src_3/src_4/src_4
Speech 4	dev1_female4/dev1_female4/dev1_male4	src_3/src_4/src_4

Fig. 9. Recording conditions of impulse response (reverberation time: $T_{60} = 300$ ms) obtained from RWCP database [28]: (a) IR1, (b) IR2, and (c) IR3.

TABLE V
EXPERIMENTAL CONDITIONS FOR MUSIC AND SPEECH SOURCE SEPARATION

Sampling frequency	16 kHz
Number of iterations	1000
Number of bases	5
Domain parameter	$\rho = 0.5$
Initial demixing matrix $\boldsymbol{W}_i$	identity matrix
Entries of initial source model matrices $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$	uniformly distributed random values
Number of trials	10 trials changing initial values

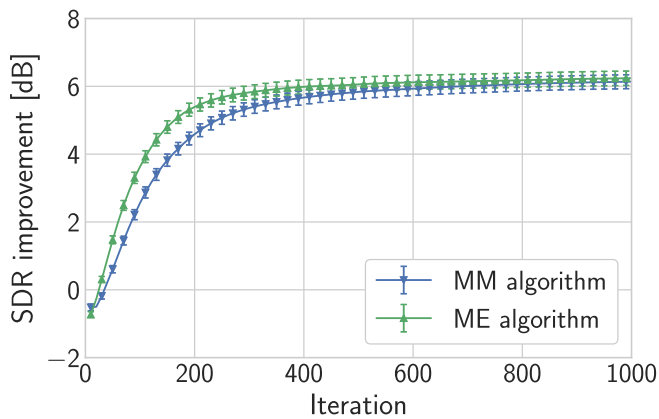Fig. 10. Average SDR improvement in GGD-ILRMA using MM and ME algorithms. Error bar represents standard error.
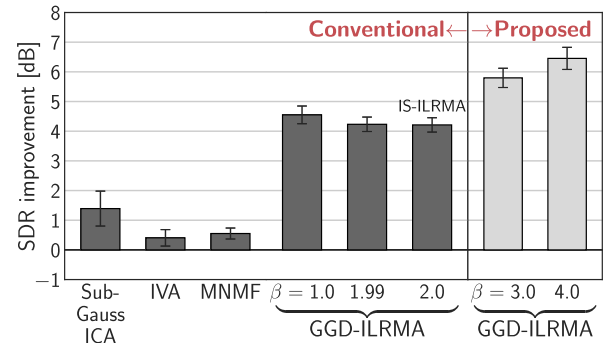

Fig. 11. Average SDR improvement of sub-Gauss ICA, IVA, MNMF, and GGD-ILRMA ($\beta = 1, 1.99, 2, 3, 4$) for two-music-signal separation under recording condition IR1. Error bar represents standard error.
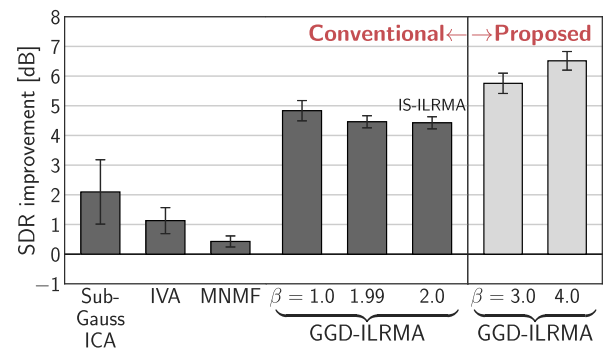

Fig. 12. Average SDR improvement of sub-Gauss ICA, IVA, MNMF, and GGD-ILRMA ($\beta = 1, 1.99, 2, 3, 4$) for two-music-signal separation under recording condition IR2. Error bar represents standard error.

and 14, the proposed GGD-ILRMA with $\beta = 4$ also outperforms or is comparable to the other conventional ILRMAs even for speech signals, which are expected to be sparse and follow super-Gaussian distributions. This finding shows that the proposed time-variant sub-Gaussian model can appropriately model super-Gaussian signals as well as sub-Gaussian signals owing to its time-variant property, as described in Section IV-A. On the other hand, conventional ILRMA with $\beta = 1$ does not provide a better separation result even though the source model is suitable for speech signals. This phenomenon has already been

reported in [12], namely, excessive low-rank injection degrades the separation performance of speech signals because they do not have a low-rank time-frequency structure in the strict sense. Compared with sub-Gaussian ILRMA with $\beta = 4$ that gives the best SDRs, the performance of sub-Gaussian ILRMA with $\beta = 3$ is slightly worse regardless of the types of source and
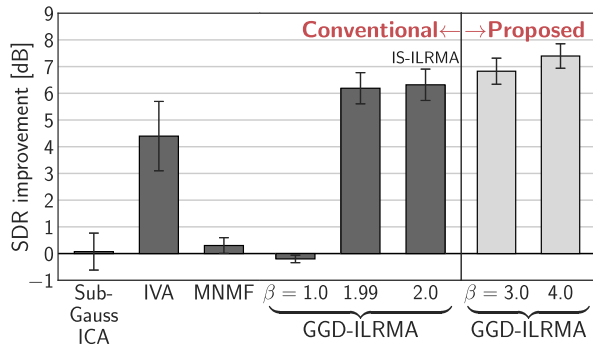
Fig. 13. Average SDR improvement of sub-Gauss ICA, IVA, MNMF, and GGD-ILRMA ($\beta = 1, 1.99, 2, 3, 4$) for two-speech-signal separation under recording condition IR1. Error bar represents standard error.
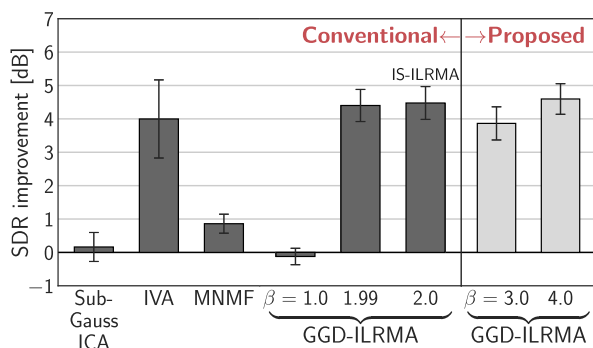

Fig. 15. Average SDR improvement of sub-Gauss ICA, IVA, MNMF, and GGD-ILRMA ($\beta = 1, 1.99, 2, 3, 4$) for three-music-signal separation under recording condition IR3. Error bar represents standard error.

Fig. 14. Average SDR improvement of sub-Gauss ICA, IVA, MNMF, and GGD-ILRMA ($\beta = 1, 1.99, 2, 3, 4$) for two-speech-signal separation under recording condition IR2. Error bar represents standard error.

Fig. 16. Average SDR improvement of sub-Gauss ICA, IVA, MNMF, and GGD-ILRMA ($\beta = 1, 1.99, 2, 3, 4$) for three-speech-signal separation under recording condition IR3. Error bar represents standard error.

recording condition. This is expected to be due to the overuse of the majorization functions, i.e., (36) and (54), which become far from the original cost function to be minimized and generally give a slow convergence.

Next, we compare the performance of sub-Gaussian GGD-ILRMA with those of ICA with a time-invariant sub-Gaussian model (sub-Gauss ICA), IVA with a time-variant super-Gaussian (Laplace) model, and MNMF with a time-variant Gaussian model. For both music and speech signals, the proposed sub-Gaussian GGD-ILRMA markedly outperforms sub-Gauss ICA, IVA, and MNMF, showing the efficacy of the proposed method. Sub-Gauss ICA slightly separates music signals (see Figs. 11 and 12) owing to the sub-Gaussianity of the signals, but cannot separate speech signals at all (see Figs. 13 and 14). Overall, the performance of sub-Gauss ICA is insufficient; one possible reason is that the gradient-based parameter update algorithm does not always guarantee a monotonic decrease in the cost function, whereas the other methods used in this experiment guarantee it. IVA can work in the speech-signal separation task to some extent (see Figs. 13 and 14) because the source model is super-Gaussian, but cannot be applicable to music signals (see Figs. 11 and 12). As reported in [10], MNMF is worse than the original IS-ILRMA especially in the determined BSS because of the difficulties in optimizing a large number of parameters,
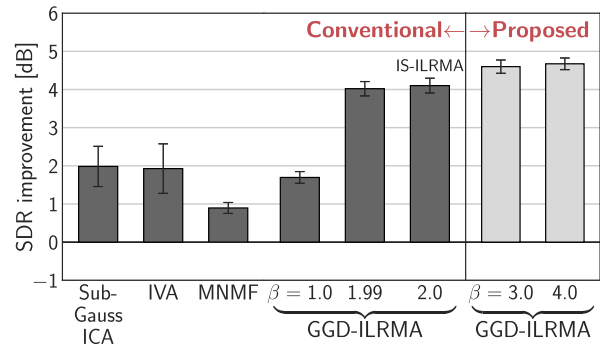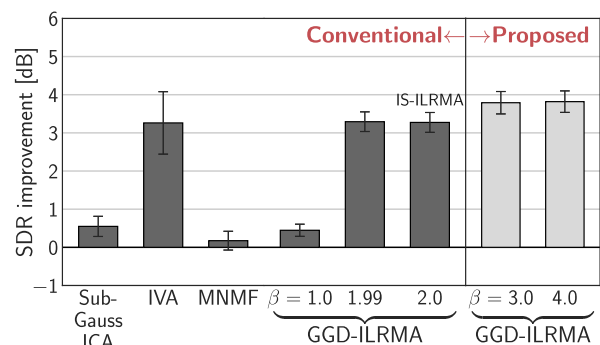
whereas the proposed sub-Gaussian ILRMA maintains a good feature of the ILRMA family, i.e., efficient parameter optimization.

D. BSS Experiment on Three-Source Signals

We compared the separation performance of the proposed sub-Gaussian GGD-ILRMA ($\beta = 3$ and 4) with those of conventional methods: sub-Gauss ICA, IVA, MNMF, IS-ILRMA, and super-Gaussian GGD-ILRMA. In this experiment, eight signals were used: four types of music signal (shown in Table III) and four types of speech signal (shown in Table IV). To simulate reverberant mixtures, the observed signals were produced by convolving the impulse response IR3, which was obtained from the RWCP database [28], shown in Fig. 9(c). The other experimental conditions were the same as those in Section VI-C.

Figs. 15 and 16 show the averages of music and speech signals, respectively. The proposed sub-Gaussian GGD-ILRMA outperforms the other conventional methods for both music and speech signals. From these results, we can confirm that the proposed sub-Gaussian ILRMA can work even in the case of $N = 3$, which generally increases the difficulties in parameter optimization, and is versatile for various combinations of sources.
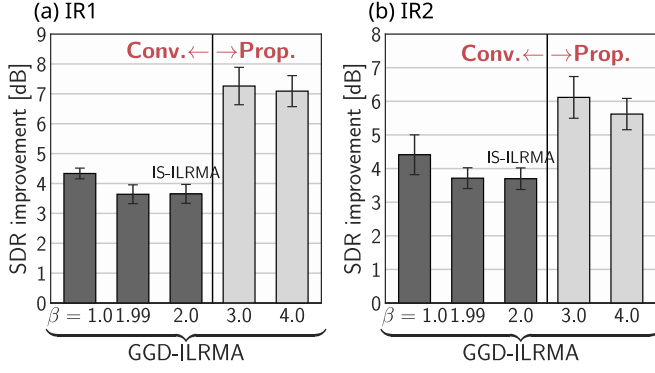
Fig. 17. Average SDR improvement of GGD-ILRMA ($\beta = 1, 1.99, 2, 3, 4$) for two-music-signal separation under recording conditions (a) IR1 and (b) IR2 with realistic audio sources. Error bar represents standard error.

VII. EVALUATION USING REALISTIC MUSIC SOURCE

In the experiments in the previous section, we used an artificial PCM-based MIDI synthesizer (Microsoft GS Wavetable Synth) to produce the simulated music signals. In this section, we show some examples of separation performance using more realistic music-instrument source signals. We used Garritan Personal Orchestra 4, which imitates more realistic sounds based on professionally recorded sample sounds. The other experimental conditions are the same as those in Section VI.

Fig. 17 shows the performance for the realistic signals in two-music-signal separation under recording conditions IR1 and IR2. From this result, we can confirm that the proposed GGD-ILRMA markedly outperforms the conventional methods, similarly to in the previous section. The proposed method separates the sources with high accuracy even for the sampling-based realistic music signals.

VIII. CONCLUSION

In this paper, we proposed a new type of ILRMA, which assumes that the source signal follows the time-variant isotropic complex sub-Gaussian GGD. By combining a new update scheme called GIP-HSM and an MM algorithm, we obtained a convergence-guaranteed update rule for the demixing matrix. Furthermore, we accelerated it using a new ME algorithm extended to a multivariate case, which is expected to be faster than the MM algorithm. In the experimental evaluations, we revealed the versatility of the proposed method, i.e., the proposed time-variant sub-Gaussian source model can deal with various types of source signal, ranging from sub-Gaussian music signals to super-Gaussian speech signals.

APPENDIX
PROOF OF THEOREM 1

Proof: Let $\boldsymbol{q} = [q_1 \quad \cdots \quad q_J]^\top = \boldsymbol{H}^{\mathsf{H}} \boldsymbol{w}$. $f(\boldsymbol{w})$ and $g(\boldsymbol{w})$ can be written as

$$f(\boldsymbol{w}) = \frac{1}{J} \sum_j |q_j|^4, \tag{83}$$

$$g(\boldsymbol{w}) = \frac{1}{J \sum_j |\tilde{q}_j|^4} (\boldsymbol{q}^{\mathsf{H}} \tilde{\boldsymbol{Q}} \boldsymbol{q})^2. \tag{84}$$

Then, the objective inequality $f(\boldsymbol{w}) \le g(\boldsymbol{w})$ holds if and only if

$$\left(\sum_j |q_j|^4 \right) \left(\sum_j |\tilde{q}_j|^4 \right) \le (\boldsymbol{q}^{\mathsf{H}} \tilde{\boldsymbol{Q}} \boldsymbol{q})^2. \tag{85}$$

The quadratic form of the right side in (85) can be deformed as

$$\boldsymbol{q}^{\mathsf{H}} \tilde{\boldsymbol{Q}} \boldsymbol{q} = \mathrm{tr}(\tilde{\boldsymbol{Q}} \boldsymbol{q} \boldsymbol{q}^{\mathsf{H}})$$

$$= \left(\sum_j |q_j|^2 \right) \|\tilde{\boldsymbol{q}}\| - \sum_{i \ne j} q_i q_j^* \tilde{q}_i^* \tilde{q}_j$$

$$= \left(\sum_j |q_j|^2 \right) \left(\sum_j |\tilde{q}_j|^2 \right) - \sum_{i \ne j} q_i q_j^* \tilde{q}_i^* \tilde{q}_j. \tag{86}$$

Hence, we prove the following inequality hereafter:

$$\left(\sum_j |q_j|^4 \right) \left(\sum_j |\tilde{q}_j|^4 \right)$$

$$\le \left(\left(\sum_j |q_j|^2 \right) \left(\sum_j |\tilde{q}_j|^2 \right) - \sum_{i \ne j} q_i q_j^* \tilde{q}_i^* \tilde{q}_j \right)^2. \tag{87}$$

Let

$$x_1 = \sum_j |q_j|^2, \tag{88}$$

$$x_2 = \sqrt{\sum_{i \ne j} |q_i|^2 |q_j|^2}, \tag{89}$$

$$y_1 = \sum_j |\tilde{q}_j|^2, \tag{90}$$

$$y_2 = \sqrt{\sum_{i \ne j} |\tilde{q}_i|^2 |\tilde{q}_j|^2}. \tag{91}$$

Since

$$x_1{}^2 - x_2{}^2 = \sum_j |q_j|^4 \ge 0, \tag{92}$$

$$y_1{}^2 - y_2{}^2 = \sum_j |\tilde{q}_j|^4 \ge 0, \tag{93}$$

and $x_1, x_2, y_1, y_2 \ge 0$, it is obvious that

$$x_1 y_1 - x_2 y_2 \ge 0. \tag{94}$$

Furthermore, we obtain the following inequality by applying the Cauchy–Schwarz inequality:

$$x_2 y_2 = \sqrt{\left(\sum_{i \ne j} |q_i|^2 |q_j|^2 \right) \left(\sum_{i \ne j} |\tilde{q}_i|^2 |\tilde{q}_j|^2 \right)}$$

$$\ge \left| \sum_{i \ne j} q_i q_j^* \tilde{q}_i^* \tilde{q}_j \right| \ge \sum_{i \ne j} q_i q_j^* \tilde{q}_i^* \tilde{q}_j, \tag{95}$$

where we used the fact that

$$\sum_{i \ne j} q_i q_j^* \tilde{q}_i^* \tilde{q}_j = \sum_{i < j} \left(q_i q_j^* \tilde{q}_i^* \tilde{q}_j + q_i^* q_j \tilde{q}_i \tilde{q}_j^* \right)$$

$$= 2 \sum_{i < j} \mathrm{Re}[q_i q_j^* \tilde{q}_i^* \tilde{q}_j] \in \mathbb{R}. \tag{96}$$

From (94) and (95), we have

$$(x_1 y_1 - x_2 y_2)^2 \leq \left(x_1 y_1 - \sum_{i \neq j} q_i q_j^* \tilde{q}_i^* \tilde{q}_j \right)^2. \qquad (97)$$

Therefore, (87) can be proven by using the Brahmagupta identity, (92), (93), and (97) as follows:

$$\left(\sum_j |q_j|^4 \right) \left(\sum_j |\tilde{q}_j|^4 \right)$$

$$= (x_1{}^2 - x_2{}^2)(y_1{}^2 - y_2{}^2)$$

$$= (x_1 y_1 - x_2 y_2)^2 - (x_1 y_2 - x_2 y_1)^2$$

$$\leq (x_1 y_1 - x_2 y_2)^2$$

$$\leq \left(x_1 y_1 - \sum_{i \neq j} q_i q_j^* \tilde{q}_i^* \tilde{q}_j \right)^2$$

$$= \left(\left(\sum_j |q_j|^2 \right) \left(\sum_j |\tilde{q}_j|^2 \right) - \sum_{i \neq j} q_i q_j^* \tilde{q}_i^* \tilde{q}_j \right)^2. \qquad (98)$$

It is easy to prove that the equality of (87) holds if $w = \tilde{w}$ because then $q = \tilde{q}$ holds. ∎

References

[1] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

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

[3] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 530–538, Sep. 2004.

[4] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 666–678, 2006.

[5] A. Hiroe, "Solution of permutation problem in frequency domain ICA using multivariate probability density functions," in *Proc. 6th Int. Conf. Independent Component Anal.*, 2006, pp. 601–608.

[6] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 1, pp. 70–79, Jan. 2007.

[7] H. Kameoka, T. Yoshioka, M. Hamamura, J. L. Roux, and K. Kashino, "Statistical model of speech signals based on composite autoregressive system with application to blind source separation," in *Proc. 10th Int. Conf. Latent Variable Anal. Signal Separation*, 2010, pp. 245–253.

[8] N. Ono and S. Miyabe, "Auxiliary-function-based independent component analysis for super-Gaussian sources," in *Proc. 10th Int. Conf. Latent Variable Anal. Signal Separation*, 2010, pp. 165–172.

[9] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2011, pp. 189–192.

[10] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, Sep. 2016.

[11] S. Mogami, D. Kitamura, Y. Mitsui, N. Takamune, H. Saruwatari, and N. Ono, "Independent low-rank matrix analysis based on complex Student's *t*-distribution for blind audio source separation," in *Proc. IEEE 27th Int. Workshop Mach. Learn. Signal Process.*, 2017.

[12] D. Kitamura *et al.*, "Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source separation," *EURASIP J. Adv. Signal Process.*, vol. 2018, no. 28, pp. 1–25, 2018.

[13] R. Ikeshita and Y. Kawaguchi, "Independent low-rank matrix analysis based on multivariate complex exponential power distribution," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 741–745.

[14] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation with independent low-rank matrix analysis," in *Audio Source Separation*, S. Makino, Ed., Cham, Switzerland: Springer, Mar. 2018, ch. 6, pp. 125–155.

[15] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[16] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 550–563, Mar. 2010.

[17] A. Ozerov, C. Févotte, R. Blouet, and J. L. Durrieu, "Multichannel non-negative tensor factorization with structured constraints for user-guided audio source separation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2011, pp. 257–260.

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

[19] G. Box and G. Tiao, *Bayesian Inference in Statistical Analysis*. Reading, MA, USA: Addison Wesley, 1973.

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

[21] G. R. Naik and W. Wang, "Audio analysis of statistically instantaneous signals with mixed Gaussian probability distributions," *Int. J. Electron.*, vol. 99, no. 10, pp. 1333–1350, 2012.

[22] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural Comput.*, vol. 23, no. 9, pp. 2421–2456, 2011.

[23] S. Mogami *et al.*, "Independent low-rank matrix analysis based on time-variant sub-Gaussian source model," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2018, pp. 1684–1691.

[24] G. Strang, *Introduction to Linear Algebra*. Wellesley Hills, MA, USA: Wellesley-Cambridge Press, 1993, vol. 3.

[25] D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo, and S. Nakamura, "Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 4, pp. 654–669, Apr. 2015.

[26] D. Kitamura, "Open dataset: songkitamura," Accessed : May 27, 2018. [Online]. Available: http://d-kitamura.net/en/dataset_en.htm

[27] S. Araki *et al.*, "The 2011 signal separation evaluation campaign (SiSEC2011): —audio source separation—," in *Proc. 10th Int. Conf. Latent Variable Anal. Signal Separation*, 2012, pp. 414–422.

[28] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," in *Proc. 2nd Int. Conf. Lang. Resour. Eval.*, 2000, pp. 965–968.

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

[30] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources," *Neural Comput.*, vol. 11, no. 2, pp. 417–441, 1999.

[31] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in *Proc. Int. Workshop ICA BSS*, 1999, pp. 365–371.

Shinichi Mogami received the B.E. degree in engineering from the University of Tokyo, Tokyo, Japan, in 2017. He is currently working toward the M.S. degree in information physics and computing with the University of Tokyo. His research interests include audio signal processing, statistical signal processing, source separation, and machine learning. He is a member of the IEEE Signal Processing Society and the Acoustical Society of Japan.

Norihiro Takamune received the B.E. degree in engineering, and the M.S. degree in information science and technology, from the University of Tokyo, Tokyo, Japan, in 2012 and 2015, respectively. He is currently a Researcher with the University of Tokyo, Tokyo, Japan. His research interests include music information analysis audio source separation and machine learning.

Yu Takahashi received the B.E. degree in information engineering from the Himeji Institute of Technology, Kobe, Japan and the M.E. and the Ph.D. degrees in information science from the Nara Institute of Science and Technology, Ikoma, Japan, in 2007 and 2010, respectively. He is currently a Researcher with Yamaha corporation. His research interest includes statistical signal processing for audio and music.

Daichi Kitamura received the Ph.D. degree from SOKENDAI, Hayama, Japan. He joined the University of Tokyo, Tokyo, Japan, in 2017 as a Research Associate, and he moved to the National Institute of Technology, Kagawa Collage, Mitoyo, Japan, as an Assistant Professor, in 2018. His research interests include audio source separation, statistical signal processing, and machine learning. He was the recipient of the Awaya Prize Young Researcher Award from The Acoustical Society of Japan (ASJ) in 2015, Ikushi Prize from Japan Society for the Promotion of Science in 2017, Best Paper Award from IEEE Signal Processing Society Japan in 2017, and Itakura Prize Innovative Young Researcher Award from ASJ in 2018.

Kazunobu Kondo received the M.E. and Ph.D. degrees from Nagoya University, Nagoya, Japan, in 1993 and 2014, respectively. He joined the Electronics Development Center, Yamaha Co., Ltd., in 1993. He is currently a Senior Engineer with Yamaha Corporate R&D Center. His research interests include blind source separation, noise reduction, and dereverberation. He is a member of the IEICE and the Acoustical Society of Japan.

Hiroshi Saruwatari received the B.E., M.E., and Ph.D. degrees from Nagoya University, Nagoya, Japan, in 1991, 1993, and 2000, respectively. He joined SECOM Intelligent Systems Laboratory, Tokyo, Japan, in 1993, and the Nara Institute of Science and Technology, Ikoma, Japan, in 2000. Since 2014, he has been a Professor with the University of Tokyo, Tokyo. His research interests include statistical audio signal processing, blind source separation, and speech enhancement. He was the recipient of the paper awards from IEICE in 2001 and 2006, from TAF in 2004, 2009, 2012, and 2018, from IEEE-IROS2005 in 2006, and from APSIPA in 2013 and 2018. He was also the recipient of the DOCOMO Mobile Science Award in 2011, Ichimura Award in 2013, The Commendation for Science and Technology by the Minister of Education in 2015, Achievement Award from IEICE in 2017, and Hoko-Award in 2018. He has been professionally involved in various volunteer works for IEEE, EURASIP, IEICE, and ASJ.

Nobutaka Ono received the B.E., M.S., and Ph.D degrees in mathematical engineering and information physics from the University of Tokyo, Tokyo, Japan, in 1996, 1998, 2001, respectively. He has been with Tokyo Metropolitan University as a Professor, since October 2017. He is the author or coauthor of more than 250 articles in international journal papers and peer-reviewed conference proceedings. His research interests include microphone array processing, blind source separation, and optimization algorithms for them. He is a Senior Member of the IEEE signal processing society, and has been a member of IEEE Audio and Acoustic Signal Processing (AASP) technical committee, since 2014. He was the recipient of the best paper awards from IEEE ISIE in 2008 and from IEEE IS3C in 2014, the Unsupervised Learning ICA Pioneer Award from SPIE.DSS in 2015, the Sato Paper Award from the Acoustic Society of Japan and two telecom system technology awards from the Telecommunications Advancement Foundation in 2018.