Circuits Syst Signal Process
DOI 10.1007/s00034-015-0162-x

A Promising Technique for Blind Identification: The Generic Statistics

Fanglin Gu[1] · Hang Zhang[2] · Wenwu Wang[3] · Chunlin Xiong[1]

Received: 15 December 2014 / Revised: 27 August 2015 / Accepted: 27 August 2015

Abstract The generic statistics, related to lower-order derivatives of the log characteristic function (CAF) evaluating at the processing points located away from the origin, are frequently used in multivariate statistical signal processing. In comparison with the conventional statistics, e.g., cumulants, the generic statistics have the following advantages: (a) they can offer the structural simplicity and controllable statistical stability of lower-order statistics, and retain higher-order statistical information; (b) if the derivatives of the log CAF were evaluated at all (infinitely many) possible processing-points, a complete description of the joint CAF would be obtained. Furthermore, we show in this paper that even if a random process is symmetrically distributed, the odd-order generic statistics are not equal to zero, while in such a case the odd-order cumulants are equal to zero. For these reasons, a family of blind identification (BI) methods, in which the mixing matrix is obtained by decomposing the tensor constructed by the higher order derivatives of the log CAF of the observations, is proposed to achieve BI

✉ Fanglin Gu
gu.fanglin@gmail.com

Hang Zhang
hangzh_2002@163.com

Wenwu Wang
w.wang@suurey.ac.uk

Chunlin Xiong
xchlzju@nudt.edu.cn

1 College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, People's Republic of China

2 College of Communication Engineering, PLA University of Science and Technology, Nanjing 210007, People's Republic of China

3 Department of Electronic Engineering, University of Surrey, Guildford, GU2 7XH, UK

Published online: 15 September 2015

of underdetermined mixtures. Simulation results show that the BI methods based on generic statistics have superior performance to the existing cumulant-based method, such as FOOBI method, especially, when the SNR of the observations is high and/or the data block is short.

Keywords Blind identification · Characteristic function · Cumulant · Tensor decomposition · Underdetermined mixture

1 Introduction

Blind identification (BI) of linear mixtures has recently attracted a lot of research interest for its wide range of applications, such as blind source separation (BSS) [5,18,31] and direction of arrivals (DOA) estimation [4,28]. Although some progress has been made in this area, e.g., the works presented in [1,5,18,31,32], there are still several open problems. Firstly, most existing work is based on the assumption that the mixture is overdetermined/determined, i.e., there should be at least as many sensors as sources. Secondly, for underdetermined mixtures, the majority of existing algorithm turns to the assumption that the sources are sparse. However, these assumptions are not always realistic in many applications. For example, (i) the underdetermined mixtures are common in practical situations, such as in the radio communication context, the reception of more sources than sensors is possible and its probability increases with the reception bandwidth; (ii) many signals such as the majority of communication signals do not possess the sparsity property.

To address the above issues, algebraic methods have been developed for BI of the mixing channels from underdetermined mixtures by using various data structures such as correlation [23,30] and higher-order cumulants [2,3,8,13,19,22] matrices. The main idea of these algebraic algorithms is to estimate the mixing matrix by decomposing a tensor formed from the cumulants of the observations. This is notably the case for the second-order blind identification of underdetermined mixtures (SOBIUM) [23], fourth-order blind identification of underdetermined mixtures (FOBIUM) [13], fourth-order-only blind identification (FOOBI) [22], FOOBI-2 [2], and BI of mixtures of sources using redundancies in the data hexacovariance matrix (BIRTH) [8] algorithms, in which only the second-order, fourth-order, or sixth-order cumulant-based tensors were used, respectively. A family of methods named blind identification of over-complete mixtures of sources (BIOME), using an arbitrary even-order $2K$ ($K > 2$) cumulants, is proposed in [3]. From a practical point of view, for a given underdetermined mixture, it would be better to use the method involving lower-order (rather than higher-order) cumulant (if both can be used to identify the given underdetermined mixture). The reason for choosing the method involving the lower-order cumulant is twofold [25]. First, the computational complexity will increase dramatically with the increase in the order of statistics. For example, computing the third-order cumulant only requires 24 multiplications per sample, while computing the fourth-order cumulant requires 64 multiplications per sample. Second, as the order increases, estimating higher-order (HO) statistics from sample data generally requires larger data sets to obtain sufficient estimation accuracy. Therefore, using odd-order $(2K-1)$ rather than

even-order ($2K$) cumulants may render a simpler BI method due to its lower dimension. Nevertheless, BI methods based on odd-order cumulants are rarely mentioned in the existing literatures. Actually, if a random process is symmetrically distributed, then its odd-order cumulants are equal to zero. Hence, for such a process we must use even-order cumulants. For example, uniform, Laplace, Gaussian, and Bernoulli–Gaussian distributions are symmetric. Additionally, some processes have extremely small odd-order cumulants and larger even-order cumulants, for such processes we would also use the latter.

On the other hand, the cumulant-based BI methods presented in [2,3,8,13,19,22,23,30] only exploit the information contained in the second order (SO) or certain HO statistic of the data, which only provides a partial statistical description. A complete statistical description of the data can be obtained by using the joint probability density function (PDF), joint characteristic function (CAF), or cumulant generating function (CGF). However, the true joint PDF, CAF, or CGF are often unavailable. Recently, the generic statistics, which are related to the derivatives of the log CAF evaluated at non-origin processing points, are used in multivariate statistical signal processing [3,17]. The motivations for using the generic statistics are summarized as follows: First, it offers the structural simplicity and controllable statistical stability of lower-order statistics and retains HO statistical information. Second, if the derivatives of the log CAF were evaluated at all (infinitely many) possible processing points, a complete description of the joint CAF would be obtained. Moreover, it turns out that, with just a few processing points, it is often adequate to extract the required statistical information from the data.

A family of BI approaches was proposed in [7,9–11,24,27,33] by exploiting the information contained in the generic statistics. In these works, the authors showed that the mixing matrix can be estimated up to trivial scaling and permutation indeterminacies by decomposing the tensor formed from partial derivatives of the CGF away from the origin. It is worth mentioning that the algorithms in [7,11,27,33] have only been applied to BI problems involving real sources and mixing process. In [24], the CAF approach was extended to the case of complex signals, which often occurs in digital communications. An improved version [9,10] was proposed recently to reduce the computational complexity. The complex form CAF, named generalized generating function (GGF), was proposed to exploit the information carried on complex random variables [14]. It has been shown that the mixing matrix can be estimated by jointly diagonalizing the Hessian matrices of the log GGF of the observations. The method was extended to address the underdetermined mixtures in [15,16]. However, the similarities and differences between the generic statistics and cumulants, which result in the superiority of the CAF-based BI algorithms, have not yet been studied comprehensively in these literatures.

In this paper, we study the problem of BI of underdetermined mixtures with non-sparse sources using CAF. First, we systematically analyze the similarities and differences between the generic statistics and cumulants. Then, we formulate the BI problem as a tensor decomposition problem in which the tensor is formed from the generic statistics of the observations. Thanks to the uniqueness of tensor decomposition, the mixing matrix can be estimated uniquely up to scale and permutation indeterminacy. To summarize, the main contributions of this article are as follows:

- Under the assumption that the $2K$th-order cumulants are not null, we prove that the odd-order derivatives of the CGF evaluated at the off-origin processing points are not equal to zero even if the random process is symmetrically distributed.
- Due to the fact that the cumulants and generic statistics have the similar structure (the former is related to the derivatives of the CGF at the origin, while the latter is related to the derivatives of the CGF at the non-origin), a family of generic statistics-based BI methods by referring to the cumulant-based BI methods. In comparison with the cumulant-based BI methods, the BI method based on generic statistics may be simpler for a typical BI problem since it is no longer restricted to an even order. The identification condition of the generic statistics-based BI method is also discussed.

The remainder of this article is organized as follows: In Sect. 2, we present the problem formulation and introduce relevant hypotheses and data statistics. In Sect. 3, the relationship and difference between the cumulants and the derivatives of generating function are discussed, and simulations are used to illustrate our findings. In Sect. 4, a new algorithm based on alternating least squares (ALS) and the CAF, called ALSCAF, is proposed to solve the BI problem, especially the underdetermined mixtures. The algorithm is implemented via tensor decomposition where the tensor is formed from the high-order partial derivatives of the log CAF of the observations. Furthermore, the identification condition is also discussed. In Sect. 5, systematic simulations are used to illustrate the performance of the proposed method. Finally, conclusions are drawn in Sect. 6.

2 Problem Formulation and Hypotheses

Consider the following linear mixture model

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

where the stochastic vector $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_Q(t)]^{\mathrm{T}}$ contains multi-channel observation signals, the components of the stochastic vector $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_P(t)]^{\mathrm{T}}$ correspond to unobserved source signals, $\mathbf{n}(t) = [n_1(t), n_2(t), \ldots, n_Q(t)]^{\mathrm{T}}$ denotes additive Gaussian noise, and t is discrete time index. The mixing matrix $\mathbf{A}$ characterizes the way that the sources are combined into the observations. The goal of BI is to estimate the mixing matrix from the observations up to scale and permutation indeterminacies, assuming that the sources are statistically independent. To clarify our idea, we restrict our discussions to the real-valued signals. Nevertheless, generalization of some of our results to complex signals is possible, which is, however, beyond the scope of this paper. In Sects. 3–5, we further make the following assumptions:

i) Noise vector $\mathbf{n}(t)$ is stationary, ergodic, and Gaussian.
ii) The mixing matrix is full (row) rank and full (row) k-rank [21].
iii) $2K$th-order marginal source cumulants are not null.

3 The Generic Statistics and Cumulant

For a real-valued stochastic vector $\mathbf{x}(t) \in R^Q$, the GF $\phi_{\mathbf{x}}(\mathbf{u})$ is defined as following

$$\phi_x(\mathbf{u}) = E\left[\exp\left(\mathbf{u}^{\mathrm{T}}\mathbf{x}(t)\right)\right], \quad \mathbf{u} \in R^Q, \tag{2}$$

where $\mathbf{u} \in R^Q$ is an arbitrary vector (to be termed as a "processing point"), $E[\cdot]$ denotes expectation operator, exp(.) is an exponential function, and $(\cdot)^{\mathrm{T}}$ denotes transpose operator. The Kth-order cumulant of these random variables is defined as the coefficient of $(u_1, \ldots, u_Q)$ in the Taylor series expansion (provided that it exists) of the CGF

$$\psi_x(\mathbf{u}) = \log E\left[\exp\left(\mathbf{u}^{\mathrm{T}}\mathbf{x}(t)\right)\right], \quad \mathbf{u} \in R^Q \tag{3}$$

Replacing $\mathbf{x}(t)$ by its model and neglecting the noise contribution leads to the decomposition of the observation CGF into the sum of the sources individual CGFs

$$\psi_x(\mathbf{u}) = \log E\left[\exp\left(\sum_{q,p} u_q A_{qp} s_p\right)\right] \tag{4}$$

Using the source independence assumption, we get

$$\psi_x(\mathbf{u}) = \sum_p \varphi_p\left(\sum_q A_{qp} u_q\right), \tag{5}$$

where $\varphi_p(v) = \log E[\exp(v s_p(t))]$, $v \in R$ is the CGF of the source. The Kth-order cumulant of the observation vector can be obtained by differentiating (5) K times with respect to K components of R^Q, denoted by $u_{q_1}, u_{q_2}, \ldots, u_{q_K}$, and evaluated at the origin

$$\mathcal{C}_{k,x}\left(x_{q_1}(t), x_{q_2}(t), \ldots, x_{q_K}(t)\right) = \partial^{(K)}\psi_x(\mathbf{u})/\partial u_{q_1}\partial u_{q_2}\ldots\partial u_{q_K}\,|_{\mathbf{u}=\mathbf{0}} \tag{6}$$

where $q_k = 1, \ldots, Q$. After a series of derivations, we obtain (more detail can be found in "Appendix A")

$$\mathcal{C}_{k,x}\left(x_{q_1}(t), x_{q_2}(t), \ldots, x_{q_K}(t)\right) = \sum_{p=1}^{P} A_{q_1 p} A_{q_2 p}\ldots A_{q_K p} C_{K,p} \tag{7}$$

where $p = 1, \ldots, P$, and $C_{K,p} = \mathrm{d}^{(K)}\varphi_p(v)/\mathrm{d}v^K\,|_{v=0}$ is the Kth-order auto-cumulant of the pth source. The Kth-order cumulant could be stored in a Kth-order tensor or represented as Q^K-element vector. There are several cumulant-based BI approaches in the literature. For example, a BI method exploiting the information contained in the FO statistics was proposed in [13]. This method is able to process underdetermined mixtures, provided that the sources have different tri-spectrum and

nonzero kurtosis with the same sign. A similar BI method, exploiting the sixth-order statistics, is proposed in [8]. Its extension to an arbitrary even-order $2K$ ($K > 2$) was presented in [3]. Despite being very powerful, and easy to implement, these methods may suffer from a high numerical complexity since the data statistics used are restricted to an even order.

To address the above limitations, we propose a new and alternative BI method, based on generic statistics, for underdetermined mixtures of sources. To proceed, we first show the relationship and the difference between the cumulants and generic statistics.

Theorem 1 *If a random process is symmetrically distributed, such as Laplacian, uniform, Gaussian, and Bernoulli–Gaussian distributions, then its odd-order cumulant is equal to zero.*

Proof Refer to "Appendix B."

It should be pointed out that some natural signals are symmetrically distributed, for example speech signals and modulated communication signals. The probability density of a speech signal and a modulated BPSK signal is shown in Fig. 1. The modulated BPSK signal is shaped by a raised cosine pulse shape filter with a roll-off $\rho = 0.5$, the baud rate is 1000 Baud/s, the carrier frequency is 3000 Hz, and the sampling frequency is 30 kHz. □

As shown in Fig. 1, it is clear that the probability density of a speech signal has a Laplacian distribution, which can be well approximated by a mixture of Gaussian distributions with 3 Gaussian components, and is symmetrically distributed. A similar conclusion also holds true for the modulated BPSK signal.

Additionally, we simulate random variables that are distributed as zero-mean uniform, Gaussian, Laplacian, and mixed Gaussians, respectively, and check whether the odd-order cumulants of these variables follow Theorem 1. The shape parameter λ in Laplace distribution is set as 1, and the way to generate a Laplacian random variable is described in [12]. The weight, mean, and variance in the mixed Gaussians are chosen as $\mathbf{w} = [0.5, 0.5]$, $\boldsymbol{\mu} = [-0.5, 0.5]$ and $\boldsymbol{\sigma}^2 = [0.4, 0.4]$, respectively. The third-order cumulants of random variables are shown in Fig. 2.

From Fig. 2, it can be observed that, as stated in Theorem 1, the third-order cumulants of the symmetrically distributed random variable are 0.

Theorem 2 *If a random process is symmetrically distributed, and its $2K$th-order cumulants are not null, then the odd-order $(2K-1)$ derivative of its CGF evaluated at the off-origin processing points is not equal to zero.*

Proof Refer to "Appendix C".

A simulation example is used to explicitly explain Theorem 2. In the simulation, random variables are distributed as zero-mean uniform, Gaussian, Laplacian, and mixed Gaussian, respectively. The coefficient λ in Laplacian distribution and the weight, mean, and variance in the mixed Gaussian distribution are the same as those in the last experiment. The third-order derivatives of the CGF evaluated at the non-origin processing points are shown in Fig. 3. □

From Fig. 3, it can be observed that, as stated in Theorem 2, the third-order derivatives of the CGF of the symmetrically distributed (except the Gaussian) random

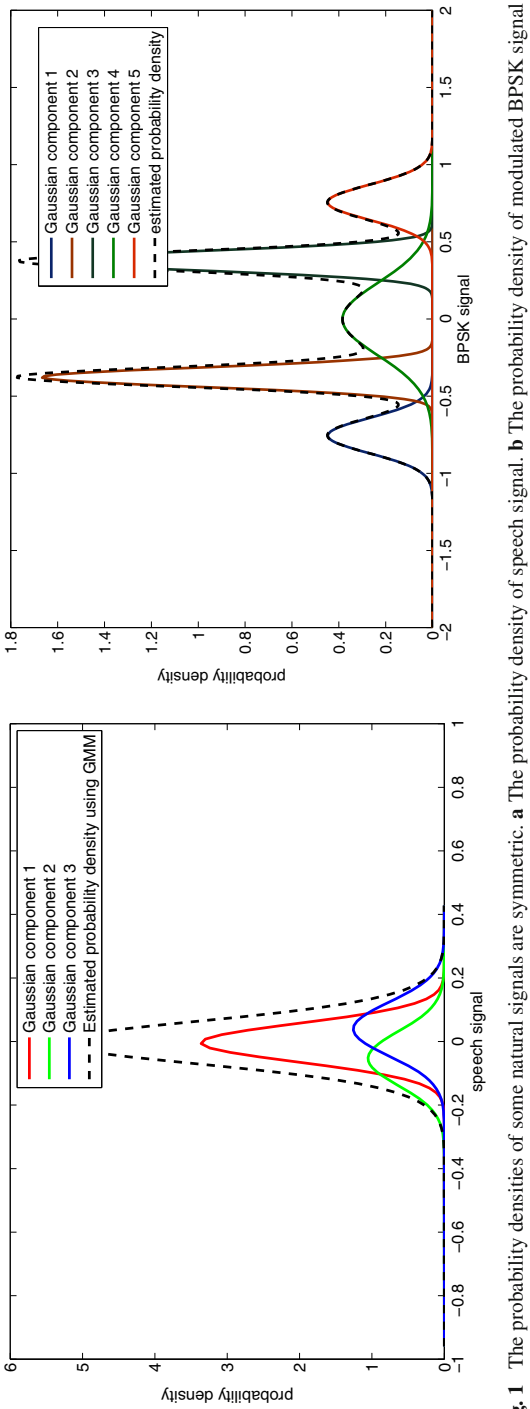

Fig. 1 The probability densities of some natural signals are symmetric. **a** The probability density of speech signal. **b** The probability density of modulated BPSK signal

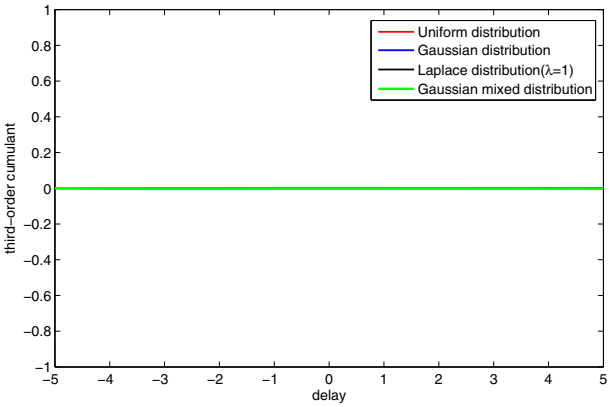

Fig. 2 The third-order cumulant of uniform, Gaussian, and Laplace distributed random variables

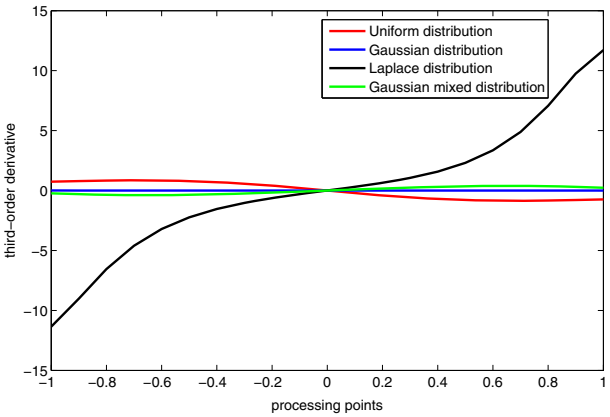

Fig. 3 The third-order derivatives of the generating function of uniform, Gaussian, and Laplace random variables

variables are not equal to 0 when evaluated at the non-origin processing points. Note that the third-order derivatives of the CGF of a Gaussian random variable are always equal to 0 due to its inherent property.

4 Blind Identification Based on Characteristic Function

The CAF approach involves the following steps: (i) forming the associated tensor decomposition by differentiating the CGF of the observations and (ii) deriving an efficient algorithm to estimate the mixing matrix from the obtained tensor decomposition.

4.1 Identification of the Mixing Matrix via Tensor Decomposition

Differentiating (5) K times with respect to K components of R^{Q}, denoted by $u_{q_1}, u_{q_2}, \ldots, u_{q_K}$, we obtain

$$\frac{\partial^{(K)}\psi_x(\mathbf{u})}{\partial u_{q_1}\partial u_{q_2}\cdots\partial u_{q_K}} = \sum_{p=1}^{P} A_{q_1 p}A_{q_2 p}\cdots A_{q_K p}G_p \tag{8}$$

where $G_p = \mathrm{d}^{(K)}\varphi_p(v)/\mathrm{d}v^K$. It is clear that Eq. (8) is equivalent to Eq. (7) when the processing points take values at 0. Nevertheless, it should be pointed out that the equations obtained when the processing points are near but not at origin can also be used to estimate the mixing matrix. Furthermore, the generic statistics-based method has the following advantages as compared with cumulant-based methods:

- It can offer structural simplicity and controllable statistical stability and retain the HO statistical information as the processing points move away from the origin;
- It can offer more flexible and simple solutions for sources with some typical distributions. For example, the tensor constructed by the odd-order (e.g., third-order) derivatives of the CGF is enough to address the BI problem even if the sources are symmetrically distributed. Therefore, it is unnecessary to resort to the even-order, for example fourth-order, cumulant which involves a HO tensor decomposition;
- The number of equations can be increased for a fixed differentiation order, by computing partial derivatives of ψ_x in S different processing points $(\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(S)})$ of R^Q. Correspondingly, it can achieve a better estimation quality by increasing the order of the tensor.

Note that (8) is nothing else but the rank-P canonical decomposition (CAND) of the data tensor, which enables the identification of the mixing matrix $\mathbf{A}$. Indeed, when the number of sources is smaller than the generic rank of the tensor, this decomposition admits an essentially unique solution of $\mathbf{A}$ (i.e., up to scaling and permutation ambiguities of their columns).

Several algorithms exist for the computation of tensor decomposition. The standard way for computing the tensor decomposition is by the means of an "alternating least squares" (ALS) algorithm [12]. Several improved versions, such as the enhanced line search (ELS) [21] and extrapolating search direction (ESD) [26], are proposed to accelerate the rate of convergence of ALS. Thus, the ALS is chosen here to compute the CAND, and the proposed algorithm is named as ALSCAF-O, where "O" indicates the order of differentiation.

Note that the differentiation order K is an input parameter of the algorithm. The higher the differentiation order, the higher the tensor order, and hence, its generic rank for these dimensions. Consequently, increasing the differentiation order should in principle allow one to identify mixtures involving a larger number of sources without increasing the number of sensors. The price to pay is, of course, an increase in the algorithm complexity and possibly a loss in robustness and accuracy.

To summarize, the general structure of ALSCAF algorithms can be summarized as follows:

- Choose S processing points of R^Q;
- Compute, for each processing point, the K-order partial derivatives of second CAF ψ_x and store the results in a K-order tensor $\mathcal{M}$;

- Estimate mixing matrix $\mathbf{A}$ from the rank-P decomposition of $\mathcal{M}$ by using the ALS algorithm.

4.2 Constructing Tensor Based on the Characteristic Function

In this subsection, we discuss the way to compute, for each processing point, the order K partial derivatives of ψ_x and store the results in a tensor $\mathcal{M}$. As in [11], it is preferred to compute formal derivatives and estimate the obtained expressions with the help of sample means.

For notational simplicity, let us define $D(\mathbf{u})$ as the partial derivatives of the GF of the observations $\phi_x(\mathbf{u})$ with respect to the components of vector $\mathbf{u}^{\mathrm{T}}$. For instance, $D_{q_1q_2}(\mathbf{u})$ is the second-order derivative with respect to components q_1 and q_2 of $\mathbf{u}^{\mathrm{T}}$. Examples of the first-, second-, and third-order derivatives are

$$D_{q_1}(\mathbf{u}) = \frac{\partial \phi_x(\mathbf{u})}{\partial u_{q_1}} = E\left[\exp\left(\mathbf{u}^{\mathrm{T}}\mathbf{x}\right) x_{q_1}\right] \tag{9}$$

$$D_{q_1q_2}(\mathbf{u}) = \frac{\partial^2 \phi_x(\mathbf{u})}{\partial u_{q_1}\partial u_{q_2}} = E\left[\exp\left(\mathbf{u}^{\mathrm{T}}\mathbf{x}\right) x_{q_1}x_{q_2}\right] \tag{10}$$

$$D_{q_1q_2q_3}(\mathbf{u}) = \frac{\partial^3 \phi_x(\mathbf{u})}{\partial u_{q_1}\partial u_{q_2}\partial u_{q_3}} = E\left[\exp\left(\mathbf{u}^{\mathrm{T}}\mathbf{x}\right) x_{q_1}x_{q_2}x_{q_3}\right] \tag{11}$$

Hence, the first-order derivatives of ψ_x are given by

$$\frac{\partial \psi_x(\mathbf{u})}{\partial u_{q_1}} = \frac{D_{q_1}(\mathbf{u})}{\phi_x(\mathbf{u})} \tag{12}$$

At order 2, the second-order derivatives of ψ_x are obtained by differentiating (12)

$$\frac{\partial \psi_x(\mathbf{u})}{\partial u_{q_1}\partial u_{q_2}} = \frac{D_{q_1q_2}(\mathbf{u})}{\phi_x(\mathbf{u})} - \frac{D_{q_2}(\mathbf{u})D_{q_2}(\mathbf{u})}{\phi_x^2(\mathbf{u})} \tag{13}$$

At order 3, the third-order derivatives of ψ_x are obtained by differentiating the order 2 equation

$$\begin{aligned}\frac{\partial^3 \psi_x(\mathbf{u})}{\partial u_{q_1}\partial u_{q_2}\partial u_{q_3}} = {} & \frac{D_{q_1q_2q_3}(\mathbf{u})}{\phi_x(\mathbf{u})} - \frac{D_{q_1q_2}(\mathbf{u})D_{q_3}(\mathbf{u})}{\phi_x^2(\mathbf{u})} \\ & - \frac{D_{q_1q_3}(\mathbf{u})D_{q_2}(\mathbf{u}) + D_{q_2q_3}(\mathbf{u})D_{q_1}(\mathbf{u})}{\phi_x^2(\mathbf{u})} + \frac{2D_{q_1}(\mathbf{u})D_{q_2}(\mathbf{u})D_{q_3}(\mathbf{u})}{\phi_x^3(\mathbf{u})}\end{aligned} \tag{14}$$

Note that we may have to use multi-linearity properties of higher orders of differentiation when the observation diversity is not sufficient. The objective is to increase the order of the tensor, with the goal of achieving a better estimation quality. Nevertheless, in the theoretical part of this study, we limit ourselves to second and third orders, since

equations associated with higher differentiation orders can be obtained in a similar manner.

4.3 Uniqueness and Complexity

To a large extent, the practical importance of tensor decomposition stems from its uniqueness properties. It is clear that the tensor decomposition can only be unique up to a permutation of the rank-1 terms and scaling of the factors of the rank-1 terms. Therefore, we call the tensor decomposition (8) essentially unique if any other matrix pair $\mathbf{A}'$ that satisfies (8) related to $\mathbf{A}$ via

$$\mathbf{A} = \mathbf{A}'\mathbf{P}\Delta \tag{15}$$

with $\Delta \in R^{P\times P}$ diagonal matrices, and $\mathbf{P} \in R^{P\times P}$ is a permutation matrix. The most general and well-known result on uniqueness is due to Kruskal [6] and depends on the concept of the k-rank.

The k-rank: The Kruskal rank or k-rank of a matrix $\mathbf{A}$, denoted by $\kappa_{\mathbf{A}}$, is the maximal number λ such that any set of λ columns of $\mathbf{A}$ is linearly independent.

Theorem 3 *The tensor decomposition of* (8) *is essentially unique if* [6]

$$K\kappa_{\mathbf{A}} + \kappa_{\mathbf{G}} \geq 2P + K \tag{16}$$

According to the assumptions presented in Sect. 2, the mixing matrix is full (row) rank and full k-rank. Hence, in practice, $\kappa_{\mathbf{A}} = \min(Q, P)$ *and* $\kappa_{\mathbf{G}} = \min(S, P)$. *In conclusion, when* $Q < P$, *if* $S \geq P$, *then the identifiable condition is* $P \leq (Q-1)K$; *if* $S < P$, *then the identifiable condition is* $P \leq ((Q-1)K + S)/2$.

On the other hand, we aim at giving an insight into the numerical complexity of the proposed algorithm. For the ALSCAF algorithm, its numerical complexity is related to the number of processing points S, the order of derivative K, the number of sources P, the number of sensors Q, and the sample size of data T. Take the ALSCAF-3 algorithm as an example, its computational load is consist of two parts. One part is attributed to implementing the ALS algorithm, and the other is attributed to computing the tensor $\mathcal{M}$. The ALS algorithm is directly used to decompose the tensor $\mathcal{M} \in R^{Q\times Q\times QS}$; therefore, the computational complexity is $\mathcal{O}(3PQ^3S + P^2Q^2S + P^2Q^2)$ per iteration. The elements of tensor $\mathcal{M}$ are calculated as shown in Eqs. (9,10,11,14). Hence, each element of tensor $\mathcal{M}$ requires $\mathcal{O}(Q + 24)$ multiplications and one exponent operation.

5 Simulations and Analysis

In this section, systematic simulations are used to illustrate the performance of the proposed CAF approaches for underdetermined mixtures. First, the performance of the proposed ALSCAF algorithms is investigated in terms of relative error performance

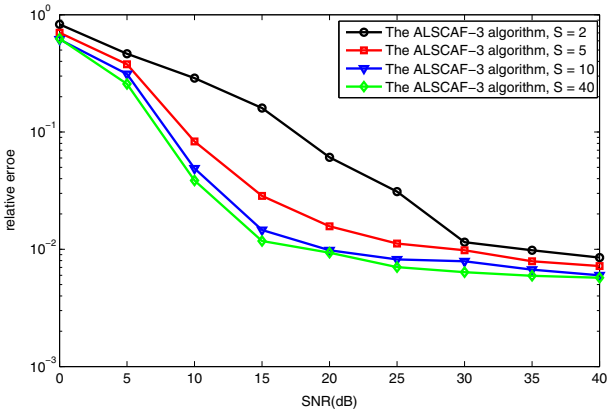

Fig. 4 The average relative error curves of the ALSCAF-3 algorithm versus the number of processing points ($Q = 3$, $P = 4$)

index (PI) versus the number of processing points. Then, some related algorithms are also used to identify the underdetermined mixtures, and the performances of the tested algorithms are evaluated and compared in terms of relative error versus sample size, the signal-to-noise ratio (SNR) of the observations, and the number of sources. Here, the relative error PI is defined as [32] PI $= E\{||\mathbf{A} - \hat{\mathbf{A}}||/||\mathbf{A}||\}$, in which the norm is the Frobenius norm and $\hat{\mathbf{A}}$ represents the optimally ordered and scaled estimate of $\mathbf{A}$. At the end, for each simulation and algorithm, the PI values are computed from 100 independent realizations of sources and mixing matrices.

First, we investigate the influence of the number of processing points S on the performance of the proposed ALSCAF algorithm. The mixture with $P = 4$ sources and $Q = 3$ observations is considered. The sources are 4-PAM signals, which are randomly drawn from $\{-3, -1, 1, 3\}$. The sources are mixed by a mixing matrix whose elements are randomly drawn from [0, 2]. Additive zero-mean Gaussian noise is added to the observations. The SNR of the observations ranges from 0 to 40 dB, and the number of observation samples is 10,000. The choice of the processing points and the value of S are two important factors that influence the performance of the proposed ALSCAF algorithm. The selection rule of the processing points is complicated, since it depends on the distribution of the stochastic variables as presented in [29]. For example, the minimum variance of the estimates of Hessian matrices is obtained when the processing point is relatively close to the origin if the random variable follows Gaussian mixture distribution, while the minimum variance of the estimates of Hessian matrices is obtained when the processing point is resided far away from the origin if the random variable follows exponential distribution (nonzero mean). Hence, without loss of generality, the processing points in the simulations are randomly drawn from interval $[-1, 1]$. On the other hand, we set the number of processing points $S = 2, 5, 10, 40$, respectively. The ALSCAF-3 algorithm is used to identify the 3×4 underdetermined mixture.

The average relative error curves of the ALSCAF-3 algorithm versus the number of processing points are shown in Fig. 4. We can see that the performance of the

ALSCAF-3 algorithm improves with the increase in the number of processing points. Nevertheless, as shown in Fig. 4, there is little improvement when the number of processing points is greater than 10.

Second, we investigate the influences of the sample size and SNR of the observations on the performance of the proposed ALSCAF algorithm. We also consider the mixture with $P = 4$ sources and $Q = 3$ observations. The sources are also 4-PAM signals, and the sources are mixed by a mixing matrix whose elements are randomly drawn from [0, 2]. Additive zero-mean Gaussian noise is added to the observations. Here, both the ALSCAF-3 and ALSCAF-4 are implemented to identify the 3×4 underdetermined mixtures. Concerning the ALSCAF-3 and ALSCAF-4, the number of processing points S is set as 10, and the processing points are drawn randomly from interval $[-1, 1]$. Besides, in order to verify the advantages of the generic statistics, the well-known FOOBI algorithm based on fourth-order cumulant [22] is implemented as the baseline algorithm.

Figure 5 shows the average relative error of the tested algorithms as a function of the SNR when $N = 10,000$ symbols are transmitted. Figure 6 shows the average relative error of the tested algorithms as a function of the number of the observation samples, when the SNR is 20 dB. We can see that on average the proposed ALSCAF algorithms have better performance than the FOOBI algorithm in terms of the relative error when the SNR of the observations is high, and the performances of ALSCAF-3 and ALSCAF-4 algorithms perform similarly in terms of relative error. It agrees with the fact that generic statistics offer the structural simplicity and controllable statistical stability of lower-order statistics and retains HO statistical information. Meanwhile, it should be pointed out that the FOOBI algorithm slightly outperforms the ALSCAF algorithms for low SNRs ($<$15 dB) due to the fact that the fourth-order cumulant of Gaussian noise is zero while the generic statistics of Gaussian noise is nonzero. Moreover, the ALSCAF-3 algorithm slightly outperforms the ALSCAF-4 algorithm when the data block is short, while the ALSCAF-4 algorithm has better performance than the ALSCAF-3 when the data block is long. The main reason is that estimating HO statistics from sample data generally requires longer data sets to obtain sufficient estimation accuracy.

Finally, we investigate the impact of the number of sources on the performance of the proposed ALSCAF algorithm. We consider two cases: $P = 5$ and $P = 6$ sources are received by 4 sensors. The sources are also 4-PAM signals, and the sources are mixed by a mixing matrix whose elements are randomly drawn from [0, 2]. Additive zero-mean Gaussian noise is added to the observations. The SNR of the observations ranges from 0 to 40 dB, and the number of observation samples is 10,000. Here, both the ALSCAF-3 and FOOBI algorithms are used to identify the mixtures. Concerning the ALSCAF-3, the number of processing points S is set as 10, and the processing points are drawn randomly from interval $[-1, 1]$.

Figure 7 shows the average relative error of the tested algorithms as a function of the SNR of the observations when the number of sensors is fixed, but the number of sources is varied. It also appears that ALSCAF-3 usually provides a lower identification error than FOOBI when the SNR of observations is high. Nevertheless, the FOOBI algorithm outperforms the ALSCAF-3 algorithm for low SNRs ($<$15 dB). The main reason is that the generic statistics are more sensitive to noise. Moreover, both the

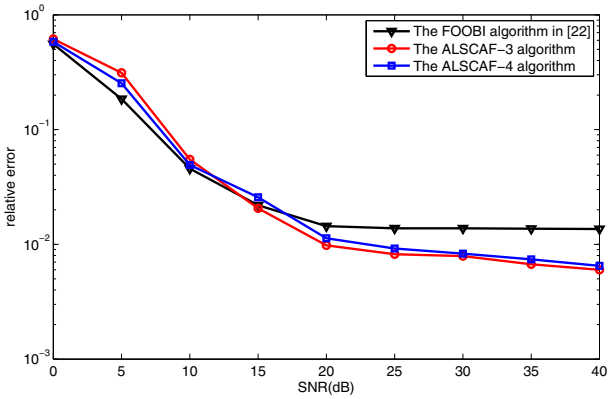

Fig. 5 The average relative error curves of the tested algorithms versus the SNR of the observations ($Q = 3$, $P = 4$)

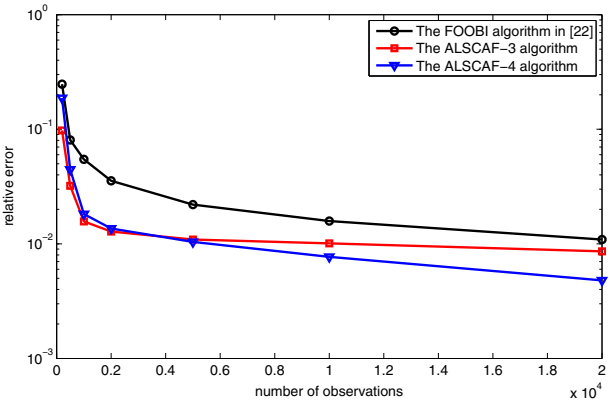

Fig. 6 The average relative error curves of the tested algorithms versus the number of the observation samples ($Q = 3$, $P = 4$)

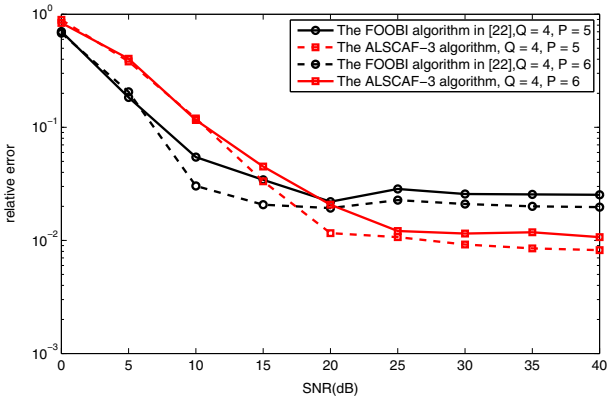

Fig. 7 The average relative error curves of the tested algorithms versus the SNR of the observations ($Q = 4$, $P = 5, 6$)

performances of FOOBI and ALSCAF-3 algorithms deteriorate with the increases in the underdeterminacy level (i.e., the increase in the number of sources).

6 Conclusions

In this paper, we have presented a new statistics named generic statistics, which is related to the derivatives of the CGF that are evaluated from the non-original processing points. In comparison with the cumulants, the generic statistics have been shown to have the following advantages: (a) it offers the structural simplicity and controllable statistical stability of lower-order statistics and retains HO statistical information; (b) if the derivatives of the log CAF were evaluated at all (infinitely many) possible processing points, a complete description of the joint CAF would be obtained; (c) the odd-order generic statistics are not equal to zero even if a random process is symmetrically distributed. Then, a family of BI method based on generic statistics is presented for underdetermined mixtures. In this method, the mixing matrix is identified by decomposing a tensor formed from the generic statistics. Simulations results show that the proposed ALSCAF method has superior performance to the existing cumulant-based methods, such as FOOBI, especially when the SNR of observations is high and the data block is short.

Acknowledgments The authors would like to thank Lieven De Lathauwer for sharing the Matlab codes of FOOBI algorithm. They also thank the anonymous reviewers for their careful reading and helpful remarks, which have contributed to improving the clarity of the paper. This work was supported in part by the Major Projects of the National Natural Science Foundation of China under Grant 91338105, and the foundation of Science and Technology on Information Transmission and Dissemination in Comm. Networks Lab.

Appendix A

Proof of Eq. (7)

In this Appendix, we show the computational details of equation in (7). First, the differentiation of (7) with respect to u_{q_1} gives

$$\begin{aligned}\frac{\partial \psi_x(\mathbf{u})}{\partial u_{q_1}} &= \sum\nolimits_{p=1}^{P} \frac{\partial\left(\varphi_p\left(\sum_q A_{qp}u_q\right)\right)}{\partial\left(\sum_q A_{qp}u_q\right)} \frac{\partial\left(\sum_q A_{qp}u_q\right)}{\partial u_{q_1}} \\ &= \sum\nolimits_{p=1}^{P} A_{q_1 p} \frac{\partial\left(\varphi_p\left(\sum_q A_{qp}u_q\right)\right)}{\partial\left(\sum_q A_{qp}u_q\right)} \end{aligned} \tag{17}$$

Similarly, the differentiation of (7) with respect to $(u_{q_1}, u_{q_2}, \ldots, u_{q_K})$ gives

$$\frac{\partial^{(K)}\psi_x(\mathbf{u})}{\partial u_{q_1}\partial u_{q_2}\cdots\partial u_{q_K}} = \sum\nolimits_{p=1}^{P} A_{q_1 p}\cdots A_{q_{K-1}p} \frac{\partial^{(K)}\left(\varphi_p\left(\sum_q A_{qp}u_q\right)\right)}{\partial\left(\sum_q A_{qp}u_q\right)^K} \frac{\partial\left(\sum_q A_{qp}u_q\right)}{\partial u_{q_K}}$$

$$= \sum_{p=1}^{P} A_{q_{1p}} \cdots A_{q_{Kp}} \frac{\partial^{(K)} \left(\varphi_p \left(\sum_q A_{qp} u_q \right) \right)}{\partial \left(\sum_q A_{qp} u_q \right)^K} \tag{18}$$

Defining $C_{K,p} = \partial^{(K)} \left(\varphi_p (\sum_q A_{qp} u_q) \right) / \partial \left(\sum_q A_{qp} u_q \right)^K$, we can rewrite (18) in a more compact form as in (7).

Appendix B

Proof of Theorem 1

Lemma *Suppose z is a random variable with a Kth-order cumulant. Then, for any $b \in R$, $z + b$ has a Kth-order cumulant and*

$$C_{K,z}(z+b) = \begin{cases} C_{K,z}(z) + b, & \text{if } K = 1 \\ C_{K,z}(z), & \text{if } K \neq 1 \end{cases} \tag{19}$$

Proof According to the definition of the generating function in (2), we obtain

$$\begin{aligned} \phi_{z+b}(u) &= E[\exp(v(z+b))] = \exp(vb) E[\exp(vz)] \\ &\Rightarrow \varphi_{z+b}(v) = vb + \varphi_z(v) \\ &\Rightarrow \mathrm{d}^{(K)} \varphi_{z+b}(v) / \mathrm{d}v^K = \mathrm{d}^{(K)}(vb) / \mathrm{d}v^K + \mathrm{d}^{(K)} \varphi_z(v) / \mathrm{d}v^K \end{aligned}$$

Since $C_{K,z} = \mathrm{d}^{(K)} \varphi_z(v) / \mathrm{d}v^K \,|_{v=0}$, Eq. (19) holds. □

Suppose z is a random process which is symmetrically distributed. According to *Lemma*, the cumulants of z are not affected by adding a constant to z. Hence, we can further assume that z has zero mean.
Because the probability density function $f(z)$ and the mean of random variable z are symmetric and zero, respectively, then, we obtain

$$\phi_z^{(2k+1)}(v) \,|_{v=0} = E[z^{2k+1}] = \int_z z^{2k+1} f(z) dz = 0 \tag{20}$$

On the other hand, exploiting the relationship between GF and CGF, we derive

$$\varphi_z(v) = \log E(\exp(vz)) = \log(\phi_z(v)) \tag{21}$$

Differentiating (21) with respect to u gives

$$\phi_z'(v) = \phi_z(v) \varphi_z'(v) \tag{22}$$

and evaluating (22) at $v = 0$ gives $M_{1,z} = C_{1,z} = 0$. Thus, Theorem 1 holds when the order is 1.

Differentiating (21) $2k+1$ times and evaluating at $v=0$ gives

$$\phi_z^{(2k+1)}(0) = \sum_{j=0}^{2k} \phi_z^{(j)}(0)\varphi_z^{(2k+1-j)}(0) \tag{23}$$

If $j = 2m+1, m = 1, \ldots, k$, using the property shown in Eq. (20), we obtain $\phi_z^{(j)}(0)\varphi_z^{(2k+1-j)}(0) = 0$; else if $j = 2m, m = 1, \ldots, k$, then, $2k+1-j$ is odd. Therefore, $\phi_z^{(j)}(0)\varphi_z^{(2k+1-j)}(0) = 0$ also holds true based on assumption that Theorem 1 holds true when the order is lower than $2k+1$ and odd. It is interesting to mention that it agrees with the mathematical induction idea.

In all, Theorem 1 follows.

Appendix C

Proof of Theorem 2

Suppose z is a random process which is symmetrically distributed. According to Theorem 1, its odd-order cumulant is equal to zero, i.e.,

$$C_{2k-1,z} = \mathrm{d}^{(2k-1)}\varphi_z(v)\big/\mathrm{d}v^{2k-1}\,|_{v=0} = 0 \tag{24}$$

Note that its $2k$-order cumulants are not null. That is

$$C_{2k,z} = \mathrm{d}^{(2k)}\varphi_z(v)\big/\mathrm{d}v^{2k}\,|_{v=0} \neq 0 \tag{25}$$

Therefore, $\mathrm{d}^{(2k-1)}\varphi_z(v)\big/\mathrm{d}v^{2k-1}$ is a monotone function for v near 0. Thus, we obtain

$$\begin{aligned} \mathrm{d}^{(2k-1)}\varphi_z(v)\big/\mathrm{d}v^{2k-1}\,|_{v\to 0} &= \mathrm{d}^{(2k-1)}\varphi_z(0) + \mathrm{d}^{(2k)}\varphi_z(0) \times v \\ &\neq 0 \end{aligned} \tag{26}$$

Thus, Theorem 2 holds true.

References

1. A. Aissa-EI-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, Y. Grenier, Underdetermined blind separation of nondisjoint sources in the time-frequency domain. IEEE Trans. Signal Process. **55**(3), 897–907 (2007)
2. L. Albera, A. Ferreol, P. Comon, P. Chevalier, Sixth order blind identification of under-determined mixtures (BIRTH) of sources, in *4th International Symposium on Independent Component Analysis and Blind Signal Separation*, Nara, Japan, pp. 909–914 (2003)
3. L. Albera, A. Ferréol, P. Comon, P. Chevalier, Blind identification of over-complete mixtures of sources (BIOME). Linear Algebra Appl. **391**, 1–30 (2004)
4. D. Astèly, A.L. Swindlehurst, B. Ottersten, Spatial signature estimation for uniform linear arrays with unknown receivers gains and phases. IEEE Trans. Signal Process. **47**(8), 2128–2138 (1999)

5. A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, E. Moulines, A blind source separation technique using second-order statistics. IEEE Trans. Signal Process. **45**(2), 434–444 (1997)
6. Y. Chen, D. Han, L. Qi, New ALS methods with extrapolating search directions and optimal step size for complex-valued tensor decompositions. IEEE Trans. Signal Process. **59**(12), 5888–5898 (2011)
7. P. Comon, M. Rajih, Blind identification of underdetermined mixtures based on the characteristic function. Signal Process. **86**(9), 2271–2281 (2006)
8. A.L.F. de Almeida, X. Luciani, P. Comon, Blind identification of underdetermined mixtures based on the hexacovariance and higher-order cyclostationarity, in *IEEE Workshop on Statistical Signal Processing*, Cardiff, UK, pp. 669–672 (2009)
9. A.L.F. de Almeida, X. Luciani, P. Comon, Fourth-order CONFAC decomposition approach for blind identification of underdetermined mixtures, in *Proceedings of 20th European Signal Processing Conference*, Bucarest, Romania, pp. 709–712 (2012)
10. A.L.F. de Almeida, X. Luciani, A. Stegeman, P. Comon, CONFAC decomposition approach to blind identification of underdetermined mixtures based on generating function derivatives. IEEE Trans. Signal Process. **60**(11), 5698–5713 (2012)
11. E. Eidinger, A. Yeredor, Blind MIMO identification using the second characteristic function. IEEE Trans. Signal Process. **53**(11), 4067–4079 (2005)
12. T. Eltoft, T. Kim, T. Lee, On the multivariate Laplace distribution. IEEE Signal Process. Lett. **13**(5), 300 (2006)
13. A. Ferréol, L. Albera, P. Chevalier, Fourth-order blind identification of underdetermined mixtures of sources (FOBIUM). IEEE Trans. Signal Process. **53**(5), 1640–1653 (2005)
14. F. Gu, H. Zhang, D. Zhu, Blind separation of complex sources using the generalized generating function. IEEE Signal Process. Lett. **20**(1), 74–77 (2013)
15. F. Gu, H. Zhang, W. Wang, D. Zhu, Generalized generating function with tucker decomposition and alternating least squares for underdetermined blind identification. EURASIP J. Adv. Signal Process. (2013). doi:10.1086/1687-6180-2013-124
16. F. Gu, H. Zhang, S. Wang, D. Zhu, Blind identification of underdetermined mixtures with complex sources using the generalized generating function. Circuits Syst. Signal Process. (2014). doi:10.1007/s00034-014-9858-6
17. M.I. Gurelli, C.L. Nikias, Blind identification and array processing applications of generalized higher-order statistics, in *Proceedings of the Military Communication Conference*, pp. 838–842 (1996)
18. A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis. Neural Comput. **9**, 1483–1492 (1997)
19. A. Karfoul, L. Albera, G. Birot, Blind underdetermined mixture identification by joint canonical decomposition of HO cumulants. IEEE Trans. Signal Process. **58**(2), 638–649 (2010)
20. T.G. Kolda, B.W. Bader, Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
21. L.D. Lathauwer, B.D. Moor, J. Vandewalle, On the best rank-1 and rank-(R_1, R_2, …, R_N) approximation of higher-order tensors. SIAM J. Matrix Anal. Appl. **21**, 1324–1342 (2000)
22. L.D. Lathauwer, J. Castaing, J.F. Cardoso, Fourth-order cumulant-based blind identification of underdetermined mixtures. IEEE Trans. Signal Process. **55**(6), 2965–2973 (2007)
23. L.D. Lathauwer, J. Castaing, Blind identification of underdetermined mixtures by simultaneous matrix diagonalization. IEEE Trans. Signal Process. **56**(3), 1096–1105 (2008)
24. X. Luciani, A.L.F. de Almeida, P. Comon, Blind identification of underdetermined mixtures based on the characteristic function: the complex case. IEEE Trans. Signal Process. **59**(2), 540–553 (2011)
25. J.M. Mendel, Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. Proc. IEEE **79**(3), 278–315 (1991)
26. D. Nion, L.D. Lathauwer, An enhanced line search scheme for complex-valued tensor decompositions: application in DS-CDMA. Signal Process. **88**(3), 749–755 (2008)
27. M. Rajih, P. Comon, Alternating least squares identification of underdetermined mixtures based on the characteristic function, in *IEEE ICASSP 2006*, Toulouse, France, pp. III–117/III–120
28. Y. Rong, S.A. Vorobyov, A.B. Gershman, N.D. Sidiropoulos, Blind spatial signature estimation via time-varying user power loading and parallel factor analysis. IEEE Trans. Signal Process. **53**(5), 1697–1710 (2005)
29. A. Slapak, A. Yeredor, Charrelation and charm: generic statistics incorporating higher-order information. IEEE Trans. Signal Process. **60**(10), 5089–5106 (2012)
30. P. Tichavský, Z. Koldovský, Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources. IEEE Trans. Signal Process. **59**(3), 1037–1047 (2011)

31. K. Todros, J. Tabrikian, Blind separation of independent sources using Gaussian mixture model. IEEE Trans. Signal Process. **55**(7), 3645–3658 (2007)
32. S. Xie, L. Yang, J. Yang, G. Zhou, Y. Xiang, Time-frequency approach to underdetermined blind source separation. IEEE Trans. Neural Netw. Learn. Syst. **23**(2), 306–316 (2012)
33. A. Yeredor, Blind source separation via the second characteristic function. Signal Process. **80**(5), 897–902 (2000)