計測自動制御学会東北支部 第228回研究集会 (2006.5.17)
資料番号 228-3

Independent Component Analysis using an Empirical Characteristic Function Based Contrast Function

○ Muhammad Tufail, Masahide Abe, and Masayuki Kawamata

Tohoku University

キーワード： independent component analysis, characteristic function, non-Gaussianity measure, contrast function

連絡先： Kawamata Laboratory, Department of Electronic Engineering, Tohoku University, 6-6-05, Aoba, Aramaki, Aoba-ku, Sendai, 980-8579

Muhammad Tufail, Tel.: (022)795-7095, Fax.: (022)263-9169, E-mail:tufail@mk.ecei.tohoku.ac.jp

1. Abstract

This paper proposes to employ an empirical characteristic function based non-Gaussianity measure as a one-unit contrast function for independent component analysis. This non-Gaussianity measure is a weighted distance between the characteristic function of a random variable and a Gaussian characteristic function at some adequately chosen sample points. Independent component analysis of an observed random vector is performed by optimizing the above mentioned contrast function (for different units) using a fixed-point algorithm. Moreover, in order to obtain a better separation performance, we employ a mechanism to choose appropriate sample points from an initially selected sample vector. Finally, some computer simulations are presented to demonstrate the validity and effectiveness of the proposed method.

2. Introduction

In the problem of blind source separation (BSS), we observe m signals $\overline{\mathbf{x}}(k) = [\overline{x}_1(k), \overline{x}_2(k), \cdots, \overline{x}_m(k)]^T$ at different sensors that are generated by the following multi-input multi-output system

$$\overline{\mathbf{x}}(k) = \overline{\mathbf{A}}\mathbf{s}(k) + \overline{\boldsymbol{\nu}}(k), \tag{1}$$

where k is a (discrete) time index, $\overline{\mathbf{A}} \in \mathbf{R}^{m\times n}$ $(m \geqq n)$ is a full rank mixing matrix, $\mathbf{s}(k) = [s_1(k), s_2(k), \cdots, s_n(k)]^T$ is the vector of original source signals, and $\overline{\boldsymbol{\nu}}(k) = [\overline{\nu}_1(k), \overline{\nu}_2(k), \cdots, \overline{\nu}_m(k)]^T$ denotes the vector of (Gaussian) noise signals present at m sensors. The objective of BSS is to estimate the input vector $\mathbf{s}(k)$, given only the observed vector $\overline{\mathbf{x}}(k)$, and certain assumptions about the statistics of sources. The primary task in BSS is to obtain a de-mixing matrix $\overline{\mathbf{W}} \in \mathbf{R}^{n\times m}$ that estimates the pseudo-inverse ($\mathbf{A}\dagger$) of the mixing matrix with the exceptions of some scaling and permutation ambiguities [1]. The output of this de-mixing

system can be written as

$$\bar{\mathbf{y}}(k) = \overline{\mathbf{W}}\bar{\mathbf{x}}(k) = \overline{\mathbf{W}\mathbf{A}}\mathbf{s}(k) + \overline{\mathbf{W}}\bar{\boldsymbol{\nu}}(k), \quad (2)$$

$$= \mathbf{H}\mathbf{s}(k) + \overline{\mathbf{W}}\bar{\boldsymbol{\nu}}(k), \quad (3)$$

where $\mathbf{H} = \overline{\mathbf{W}\mathbf{A}} \in \mathbf{R}^{n\times n}$ is the global transformation matrix from $\mathbf{s}(k)$ to $\bar{\mathbf{y}}(k)$. Once an unbiased estimate of $\mathbf{A}\dagger$ is available, we can employ some auxiliary filters or optimal nonlinear estimators in order to reconstruct the sources from the observed data (in the presence of noise) [1, 2].

Independent component analysis (ICA) (see [1]) is a statistical technique that can be used for BSS, provided the sources are statistically independent and at most one source is Gaussian. Although, there exist many approaches for ICA, our main concern in this paper is ICA by maximizing non-Gaussianity. In order to implement the maximum non-Gaussianity approach, we require some quantitative measure of non-Gaussianity such as negentropy. The well known FastICA algorithm (see [3]) maximizes certain approximations of negentropy in order to perform ICA. Although, practically any smooth non-quadratic function can be employed for this purpose [4], the optimal nonlinear functions depend on the (unknown) pdfs of sources [3]. This implies that if some preselected (parametric) nonlinearity differs considerably from the optimal function, the FastICA algorithm may perform poorly for these sources.

It is therefore necessary to employ some direct measure of non-Gaussianity that works well for signals with wide range of pdfs. In this paper, we propose to utilize an empirical characteristic function (ecf) based non-Gaussianity measure (contrast function) in order to perform ICA. It may be noted that the ecf, being the Fourier-Stieltjes transform of the empirical distribution, retains all the information about the data. Consequently, the estimation methods based on the ecf can be made as efficient as the likelihood-based approaches [5]. In ICA estimation, the ecf has already been used in [6] to construct an objective function for measuring statistical independence between random variables. In this contribution, we employ the ecf to (directly) measure the distance of an arbitrary empirical distribution from the Gaussian distribution (at some adequately chosen sample points). Such a contrast function can be easily maximized by using a fixed-point algorithm. Furthermore, we also suggest a procedure for choosing appropriate sample points (from an initially chosen sample vector) in order to obtain somewhat better separation performance. Finally, some simulation results are given in order to show that the proposed approach works well for both symmetric and asymmetric distributions.

3. Pre-whitening

Pre-whitening is a commonly employed pre-processing technique in many ICA algorithms including JADE [7] and FastICA [3]. It is mainly used to reduce the complexity of the BSS problem. Considering the noisy mixture model given by Eq. 1, a robust pre-whitening stage linearly transform the observed vector into another n-dimensional vector $\mathbf{x}(k) = \mathbf{Q}\bar{\mathbf{x}}(k)$, given by

$$\mathbf{x}(k) = \mathbf{Q}\overline{\mathbf{A}}\mathbf{s}(k) + \mathbf{Q}\overline{\boldsymbol{\nu}}(k) = \mathbf{A}\mathbf{s}(k) + \boldsymbol{\nu}(k), \quad (4)$$

where $\mathbf{Q} \in \mathbf{R}^{n\times m}$ is a whitening matrix, $\boldsymbol{\nu}(k) = \mathbf{Q}\overline{\boldsymbol{\nu}}(k)$ is the transformed noise vector, and $\mathbf{A} = \mathbf{Q}\overline{\mathbf{A}} \in \mathbf{R}^{n\times n}$ is an orthogonal matrix i.e. $\mathbf{A}\mathbf{A}^T = \mathbf{I}_n$. Therefore, in order to achieve BSS, we are now required to obtain an orthogonal de-mixing matrix $\mathbf{W}$ that estimates $\mathbf{A}^{-1} = \mathbf{A}^T$. The corresponding

output vector is given by

$$\mathbf{y}(k) = \mathbf{W}\mathbf{x}(k). \tag{5}$$

For simplicity, we assume that $m = n$, and the covariance matrix $\mathbf{C}_{\overline{\boldsymbol{\nu}}}$ of $\overline{\boldsymbol{\nu}}$ is known. In this case, a robust pre-whitening matrix can be obtained as [8]

$$\mathbf{Q} = (\mathbf{C}_{\overline{\mathbf{x}}} - \mathbf{C}_{\overline{\boldsymbol{\nu}}})^{-1/2}, \tag{6}$$

where $\mathbf{C}_{\overline{\mathbf{x}}} = E\overline{\mathbf{x}}\overline{\mathbf{x}}^T$ is the covariance matrix of $\overline{\mathbf{x}}$. It can be easily seen from Eq. 4 that the covariance matrix $\mathbf{C}_{\mathbf{x}}$ of the transformed vector $\mathbf{x}$ is given by

$$\mathbf{C}_{\mathbf{x}} = \mathbf{I}_n + \mathbf{C}_{\boldsymbol{\nu}}, \tag{7}$$

where $\mathbf{C}_{\boldsymbol{\nu}} = \mathbf{Q}\mathbf{C}_{\overline{\boldsymbol{\nu}}}\mathbf{Q}^T$ is the covariance matrix of $\boldsymbol{\nu}$.

4. ICA by Maximizing Non-Gaussianity

ICA is usually achieved by minimizing a contrast function that attains its minimum value when the output signals $\{y_i = \mathbf{w}_i^T\mathbf{x}\}_{i=1}^n$ become mutually statistical independent, where $\mathbf{w}_i^T$ denotes the ith row of the de-mixing matrix. A natural choice for such a contrast function is mutual information, which is considered as the most satisfying information-theoretic measure of statistical dependence between random variables. In case of the output vector given by Eq. 5 (with $\mathbf{W}\mathbf{W}^T = \mathbf{I}_n$), it leads to the following contrast function [1]

$$I(\mathbf{y}, \mathbf{W}) = \sum_{i=1}^{n} H(y_i) - H(\mathbf{x}), \tag{8}$$

where $H(y_i) = -E\log(q_i(y_i))$ is the entropy of y_i, $q_i(y_i)$ is the pdf of y_i, and $H(\mathbf{x})$ is the entropy of $\mathbf{x}$. After some simple manipulations, we can write Eq. 8 as [3]

$$I(\mathbf{y}, \mathbf{W}) = C - \sum_{i=1}^{n} J(y_i), \tag{9}$$

where $J(y_i)$ is the negentropy of y_i and C is some irrelevant constant. Negentropy of a random variable u is defined as

$$J(u) = H(u_{\text{gauss}}) - H(u), \tag{10}$$

where u_{gauss} denotes a Gaussian random variable with the same variance as u. Negentropy is a measure of non-Gaussianity in the sense that it is always nonnegative, and attains its minimum value (of 0) if and only if u has a Gaussian pdf. Therefore, in the light of CLT, Eq. 9 implies that ICA by minimization of mutual information is equivalent to finding directions in which the outputs y_i are uncorrelated and maximally non-Gaussian.

Finally, it may be noted that negentropy is difficult to estimate from its definition since it requires an estimate of the pdf. However, we can still perform consistent ICA estimation by replacing negentropy by any other (good) measure of non-Gaussianity. The FastICA algorithm, for instance, employs certain simpler approximations of negentropy to perform ICA. These approximations can be written as [4]

$$\hat{J}(u) \propto [E\{G(u) - G(u_{\text{gauss}})\}]^2, \tag{11}$$

where G is any smooth no-quadratic function. Some choices of these nonlinear functions include $\log\cosh(u)$, $\exp(-u^2)$, u^4, and u^3. Nevertheless, the best performance is obtained if we choose G proportional to the log of the pdf of u [3]. This implies that any fixed general purpose function may not work well for all the pdfs. We can overcome this drawback by employing more than one (appropriate) nonlinearities or utilize a non-Gaussianity measure that works well for different pdfs.

5. The Proposed Approach

5.1 An Empirical Characteristic Function based Non-Gaussianity Measure

The characteristic function (cf) of a random variable u is defined as the Fourier-Stieltjes transform of its cumulative distribution function (cdf),

$$c(u,\lambda) = \int_{-\infty}^{\infty} \exp(j\lambda u) dF(u), \qquad (12)$$

where $j = \sqrt{-1}$, λ is a real valued frequency parameter, and $F(u)$ is the cdf of u. The cf can be easily estimated from the data by replacing it with the ecf. Given N independent samples $\{u(k)\}_{k=1}^{N}$, the ecf is defined as

$$c_N(u,\lambda) = \frac{1}{N}\sum_{k=1}^{N} \exp(j\lambda u(k)). \qquad (13)$$

From the above equation, we see that $c_N(u,\lambda)$ is a sum of bounded i.i.d. random variables. Therefore, it follows from the strong law of large numbers that $c_N(u,\lambda)$ converges almost surely to $c(u,\lambda)$ for every λ. This implies that $c_N(u,\lambda)$ is a consistent estimate of $c(u,\lambda)$ $\forall\lambda$. Furthermore, it is shown in [9] that as a random process, $Y_N(\lambda) = \sqrt{N}(c_N(u,\lambda) - c(u,\lambda))$ converges weakly to a zero mean complex Gaussian process satisfying $Y(\lambda) = Y(-\lambda)$ and

$$EY(\lambda_1)Y(\lambda_2) = c(\lambda_1+\lambda_2) - c(\lambda_1)c(\lambda_2). \qquad (14)$$

In order to measure non-Gaussianity by ecf, we note that $c_N(u,\lambda)$ is the Fourier-Stieltjes transform of the empirical cdf $F_N(u) = P(u)/N$, where $P(u)$ is the number of $u(i) \leqq u$ with $1 \leqq i \leqq N$. This implies that there exists a one-one correspondence between $F_N(u)$ and $c_N(u,\lambda)$. More specifically, if u_1 and u_2 are two random variables then we have $F_N(u_1) = F_N(u_2)$ if and only if $c_N(u_1,\lambda) = c_N(u_2,\lambda)$ $\forall\lambda$. Consequently, we can utilize a distance measure in the ecf domain to determine how far two distributions are from each other. In particular, we employ a weighted distance between $c_N(u,\lambda)$ and $c(u_{\text{gauss}},\lambda)$ in order to obtain a non-Gaussianity measure. Here, $c(u_{\text{gauss}},\lambda) = \exp(-\lambda^2/2)$ is the cf of the Gaussian distribution and we have assumed that both u and u_{gauss} have unit variances.

Such a distance measure, in comparison to approximations of negentropy given by Eq. 11, is expected to work for wider range of pdfs. For instance, it is possible to observe a non-Gaussian random variable u that has a zero kurtosis value but nonzero higher order cumulants (or the pdf of u is asymmetric). For these signals, kurtosis will not work as a measure of non-Gaussianity. On the other hand, as indicated above, $c_N(u,\lambda)$ becomes equal to $c(u_{\text{gauss}},\lambda)$ $\forall\lambda$ only when u has a Gaussian pdf.

If we use an L_2 type distance measure, the resulting test statistics can be written as [10]

$$T_N(u) = \int_{-\infty}^{\infty} \beta(\lambda) \mid c_N(u,\lambda) - c(u_{\text{gauss}},\lambda) \mid^2 d\lambda, \qquad (15)$$

where $\beta(\lambda)$ is a nonnegative integrable weight function. More importantly, the test statistics of the form given by Eq. 15 have shown to be consistent against general alternatives [10]. The weight function in Eq. 15 is chosen so that the integral remains bounded and yields a closed form expression. A convenient choice for such a weight function is the pdf of a standard Gaussian distribution i.e.

$$\beta^G(\lambda) = \exp(-\lambda^2/2). \qquad (16)$$

With this weight function, we can write the test statistics given by Eq. 15 as (see [5])

$$T_N^{\text{G}}(u) = \sqrt{2\pi}\sum_{k=1}^{N}\sum_{l=1}^{N} \exp(-(u(k)-u(l))^2/2) - 2\sqrt{\pi}\sum_{k=1}^{N} \exp(-u(k)^2/2) + \sqrt{2\pi/3}. \qquad (17)$$

The above statistics can now be employed as a non-Gaussianity measure in order to perform ICA. More specifically, assuming that the observed vector is already pre-whitened, we can obtain a demixing matrix by maximizing the following objective function with respect to $\mathbf{W}$

$$\Phi_N^G(\mathbf{y}) = \sum_{i=1}^{n} T_N^{\mathrm{G}}(y_i). \tag{18}$$

A simple gradient algorithm can be used for this purpose. However, it may be noted that an optimization of the above objective function will be computationally complex since an evaluation of the test statistics $T_N^{\mathrm{G}}(y_i)$ requires $o(N^2)$ computations in each iteration. We can overcome this drawback by employing the following simple weight function

$$\beta^S(\lambda) = \sum_{i=1}^{M} \delta(\lambda - \lambda_i), \tag{19}$$

where $\delta(.)$ is the delta function and $\{\lambda_i\}_{i=1}^M$ are some appropriately chosen M sample points. By substituting the above weight function in Eq. 15, we get the following sum of squares test statistics

$$T_N^S(u, \boldsymbol{\lambda}) = \sum_{i=1}^{M} \mid c_N(u, \lambda_i) - c(u_{\mathrm{gauss}}, \lambda_i) \mid^2, \tag{20}$$

where $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_M]^T$ is the vector of sample points. The above equation can also be written as

$$\begin{aligned} T_N^S(u, \boldsymbol{\lambda}) &= \sum_{i=1}^{M} (c_N^{\Re}(u, \lambda_i) - \exp(-\lambda_i^2/2))^2 \\ &+ \sum_{i=1}^{M} c_N^{\Im}(u, \lambda_i)^2, \end{aligned} \tag{21}$$

where $c_N^{\Re}(u, \lambda_i)$ and $c_N^{\Im}(u, \lambda_i)$ are the real and imaginary parts of $c_N(u, \lambda_i)$, respectively, i.e.

$$c_N^{\Re}(u, \lambda_i) = \frac{1}{N} \sum_{i=1}^{N} \cos(\lambda_i u(i)), \tag{22}$$

$$c_N^{\Im}(u, \lambda_i) = \frac{1}{N} \sum_{i=1}^{N} \sin(\lambda_i u(i)). \tag{23}$$

The criterion given by Eq. 20 can be considered as a special case of weighted sum of squares test statistics [11] with weighting matrix equal to Identity. We can also utilize a similar statistics by obtaining the weighting matrix from an estimate of the covariance structure given by Eq. 14. However, for simplicity we avoid the estimate of such a matrix. This approach is justified if N is sufficiently large or the sample points are chosen adequately. In Section 5.3, we consider the effect of choosing different sample points on the (separation) performance for some distributions.

Next, we slightly modify the distance measure given by Eq. 20 in an attempt to remove the effect of additive Gaussian noise. In order to do so, let us consider that the observed random variable contains a noise component η i.e. $u = u_1 + \eta$. Assuming that u_1 and η are statistically independent, the cf of u is given by

$$c(u, \lambda) = c(u_1, \lambda) c(\eta, \lambda) = c(u_1, \lambda) \exp(-\sigma_\eta^2 \lambda^2/2) \tag{24}$$

where $\sigma_\eta^2 = \sigma_u^2 - 1$ is the variance of η. Consequently, the effect of noise can be reduced by replacing $c_N(u, \lambda_i)$ in Eq. 20 by $\exp(\sigma_\eta^2 \lambda_i^2/2) c_N(u, \lambda_i)$. Accordingly, the objective function in Eq. 21 modifies to

$$\begin{aligned} T_N^{S_1}(u, \boldsymbol{\lambda}) = &\sum_{i=1}^{M} \exp(-\lambda_i^2)(c_N^{\Re}(u, \lambda_i) \exp(\sigma_u^2 \lambda_i^2/2) - 1)^2 \\ &+ \sum_{i=1}^{M} \exp(-\lambda_i^2) \exp(\sigma_u^2 \lambda_i^2) c_N^{\Im}(u, \lambda_i)^2. \end{aligned} \tag{25}$$

5.2 Optimization Method

Based on the above formulation, we employ the non-Gaussianity measure $T_N^{S_1}(y_i, \boldsymbol{\lambda})$ as a one-unit contrast function in order to obtain a single independent component. In particular, we derive a simple fixed-point iterative algorithm for maximizing $T_N^{S_1}(y_i, \boldsymbol{\lambda})$ under the constraint $\|\mathbf{w}_i\| = 1$.

Such a constraint optimization problem can easily be solved by using the method of Lagrange multipliers.

To begin with, we define the Lagrangian as

$$L(\mathbf{w}_i, \boldsymbol{\lambda}) = \frac{1}{2} T_N^{S_1}(y_i, \boldsymbol{\lambda}) - \frac{1}{2}\tau(\|\mathbf{w}_i\|^2 - 1), \quad (26)$$

where τ is the Lagrange multiplier that is computed so as to satisfy the constraint equation. According to the Lagrange conditions, the optima of the constraint optimization problem are obtained at the points where the gradient of $L(\mathbf{w}_i, \boldsymbol{\lambda})$ is zero. After some simple calculations, we can write the gradient of $L(\mathbf{w}_i, \boldsymbol{\lambda})$ as

$$\begin{aligned} \nabla_{\mathbf{w}_i} L(\mathbf{w}_i, \boldsymbol{\lambda}) &= \sum_{p=1}^{M} \alpha_1(y_i, \lambda_p) E g_1(y_i, \lambda_p)\mathbf{x} + \\ &\sum_{p=1}^{M} \alpha_2(y_i, \lambda_p) E g_2(y_i, \lambda_p)\mathbf{x} - \tau \mathbf{w}_i, \quad (27) \end{aligned}$$

where the four quantities α_1, α_2, g_1, and g_2 are defined as

$$\begin{aligned} \alpha_1(y_i, \lambda_p) &= \exp(-\lambda_p^2)\left[c_N^{\Re}(y_i, \lambda_p)\exp(\sigma_{y_i}^2 \lambda_p^2)\right. \\ &- \left. \exp(\sigma_{y_i}^2 \lambda_p^2/2)\right], \quad (28) \\ \alpha_2(y_i, \lambda_p) &= \exp(-\lambda_p^2)\exp(\sigma_{y_i}^2 \lambda_p^2) c_N^{\Im}(y_i, \lambda_p), \quad (29) \\ g_1(y_i, \lambda_p) &= -\lambda_p \sin(\lambda_p y_i) + \lambda_p^2 c_N^{\Re}(y_i, \lambda_p) y_i, \quad (30) \\ g_2(y_i, \lambda_p) &= \lambda_p \cos(\lambda_p y_i) + \lambda_p^2 c_N^{\Im}(y_i, \lambda_p) y_i. \quad (31) \end{aligned}$$

The Lagrange multiplier τ can easily be computed from Eq. 27 and is given by

$$\begin{aligned} \tau &= \sum_{p=1}^{M} \alpha_1(y_i, \lambda_p) E g_1(y_i, \lambda_p) y_i \\ &+ \sum_{p=1}^{M} \alpha_2(y_i, \lambda_p) E g_2(y_i, \lambda_p) y_i. \quad (32) \end{aligned}$$

In order to solve the problem by an approximate Newton method, we evaluate the Jacobian of $L(\mathbf{w}_i, \boldsymbol{\lambda})$ at the optimum $\mathbf{w}_i^*$. After some straight forward calculations, and using the approximations $E\mathbf{x}\mathbf{x}^T \cos(\lambda_p y_i) \approx E\cos(\lambda_p y_i)\mathbf{I}_n$ and $E\mathbf{x}\mathbf{x}^T \sin(\lambda_p y_i)$ $\approx E\sin(\lambda_p y_i)\mathbf{I}_n$ [3], the approximate Newton iteration reduces to the following fixed-point algorithm

$$\begin{aligned} \mathbf{w}_i^+ &= \sum_{p=1}^{M} \alpha_1(y_i, \lambda_p) E g_1(y_i, \lambda_p)\mathbf{x} \\ &+ \sum_{p=1}^{M} \alpha_2(y_i, \lambda_p) E g_2(y_i, \lambda_p)\mathbf{x}, \quad (33) \\ \mathbf{w}_i &= \frac{\mathbf{w}_i^+}{\|\mathbf{w}_i^+\|}, \quad (34) \end{aligned}$$

where an explicit normalization of the weight vector avoids the computation of τ. In order to implement the above algorithm, we replace the expectations in Eq. 33 by their sample estimates.

The one-unit fixed-point algorithm described by Eq. 33-34 is meant to obtain a single independent component. More than one independent components (or rows of the de-mixing matrix) can be estimated by running the same one-unit algorithm several times but ensuring that different weight vectors are orthogonal to each other. A deflation approach or a symmetric orthogonalization approach can be employed for this purpose [3].

5.3 Choice of Sample Points

It may be noted that the non-Gaussianity measure given by Eq. 25 is only defined at some finite set of sample points. An adequate choice of these sample points can increase the efficiency of the corresponding optimization algorithm. To illustrate this, let us consider a special case when $M = 1$, and the sources are symmetric so that we only utilize the first term in Eq. 25. The distance measure given by Eq. 25 then becomes equivalent to a one-unit contrast function of the form given by Eq. 11 with

$$G(u) = \cos(\lambda u). \quad (35)$$

This family of contrast functions is analyzed mathematically in [12]. In particular, the trace of the

asymptotic covariance matrix of the estimator $\mathbf{w}_i$ can be written as [12]

$$V_G = C_1(\overline{\mathbf{A}}) \frac{E\{g^2(s_i)\} - (E\{s_i g(s_i)\})^2}{(E\{s_i g(s_i) - g'(s_i)\})^2}, \quad (36)$$

where $g(.)$ is the derivative of G, and $C_1(\overline{\mathbf{A}})$ is an irrelevant constant depending on the mixing matrix $\overline{\mathbf{A}}$. With $G(s_i) = \cos(\lambda s_i)$, we have

$$V_G(\lambda) = C_1(\overline{\mathbf{A}}) \frac{\left[0.5 - 0.5c^{\Re}(s_i, 2\lambda) - (\frac{\partial c^{\Re}(s_i,\lambda)}{\partial \lambda})^2\right]}{\left[\frac{\partial c^{\Re}(s_i,\lambda)}{\partial \lambda} + \lambda c^{\Re}(s_i, \lambda)\right]^2}. \quad (37)$$

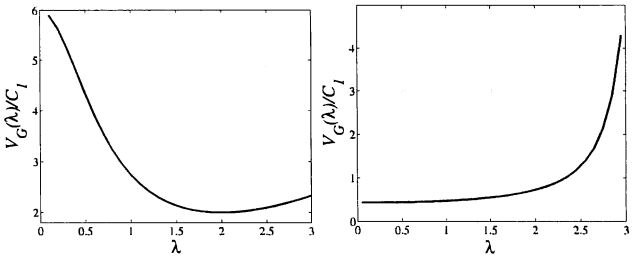
In Figure 1, we plot $V_G(\lambda)/C_1(\overline{\mathbf{A}})$ for two well known distributions including the Laplace distribution and the Uniform distribution. The cfs of these distributions (with zero means and unit variances) are given by

$$c(u_{\text{Laplace}}, \lambda) = \frac{2}{2+\lambda^2}, \quad (38)$$

$$c(u_{\text{Uniform}}, \lambda) = \frac{\sin(\sqrt{3}\lambda)}{\sqrt{3}\lambda}. \quad (39)$$

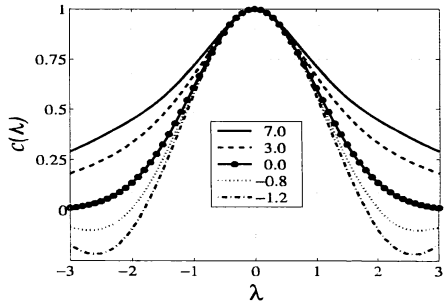
From Figure 1, we note that in case of a Laplace distribution, the asymptotic variance decreases sharply and reaches its minimum value as λ is increased from 0 to 2. Although, for $\lambda > 2$ the asymptotic variance increases again, it remains close to its minimum value in the neighborhood of $\lambda = 2$. On the other hand, for a Uniform distribution, the asymptotic variance is minimized by choosing a very small value of λ near the origin. In addition to this, the asymptotic variance now increases as λ is made large.

From the above examples, it is evident that a good choice of λ depends on the pdfs of sources. In general, a relatively large λ is preferable for super-Gaussian signals as compared to sub-Gaussian signals. Pursuing this further, we can improve the separation performance by selecting appropriate sample points from some initially chosen sample vector.



(a) (b)

Fig. 1 The plot of $V_G(\lambda)/C_1(\overline{\mathbf{A}})$ as a function of λ for (a) the Laplace distribution, and (b) the Uniform distribution.



(a)

Fig. 2 Characteristic functions of some distributions with kurtosis values shown in the legend.

Such a selection can be (roughly) made by considering the cf of the output y_i. As shown in Figure 2, the cf of a super-Gaussian (kurtosis> 0) / sub-Gaussian (kurtosis< 0) random variable usually decays slower / faster than a Gaussian cf. Therefore, in general, we will use sample points in a limited range around zero i.e. $0 < \lambda_i \leqq 1$. However, if $c_N^{\Re}(y_i, \lambda_p)|_{1<\lambda_p \leqq 2}$ is sufficiently larger than $\exp(-\lambda_p^2/2)$, we also add λ_p to the sample vector.

In case of asymmetric distributions, the situation is somewhat more complicated since sine and cosine functions measure two different characteristics of a distribution. Nevertheless, we have found through computer simulations that in case of asym-

metric distributions, sample points in a small range around zero yields good performance.

In the light of above discussion, we initially select sample points such that $0 < \lambda_p \leqq 1$. Subsequently, more sample points are added provided $|c_N^{\Im}(y_i, \lambda_p)|$ is sufficiently small and $c_N^{\Re}(y_i, \lambda_p)$ is larger than $\exp(-\lambda_p^2/2)$ with $1 < \lambda_p \leqq 2$.

6. Simulation Results

In this section, we give some simulation results in order to demonstrate the effectiveness of the proposed method. As described previously, we consider the noisy mixture model given by Eq. 4. This is equivalent to assuming that an estimate of the noise covariance matrix is available so that we can perform the robust pre-whitening as described in Section 2.

The separation performance is evaluated by the following performance index

$$PI = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \sum_{l=1}^{n} \left(\frac{h_{il}^2}{\max_j h_{ij}^2} - 1 \right), \qquad (40)$$

where h_{ij} is the ijth element of the global transformation matrix $\mathbf{H}$. The above performance index effectively measures the inverse of the average output signal to noise ratio. Therefore, a small value of PI implies a good separation performance.

Experiment 1: In the first computer simulation, we evaluate the performance of the proposed approach in the presence of Gaussian noise. Five sources are mixed together using a randomly chosen mixing matrix $\overline{\mathbf{A}} \in \mathbf{R}^{5\times 5}$. The covariance matrix of the noise vector is given by

$$C_{\overline{\nu}} = 0.1\mathbf{I}_5. \qquad (41)$$

A comparison of the proposed method is performed with the modified FastICA algorithm [8] in which Gaussian moments (defined as the expectations of Gaussian functions or their derivatives / integrals) are used as one-unit contrast functions in order to obtain an unbiased estimate of $\mathbf{W}$ in the presence of Gaussian noise. The modified one-unit algorithm (with bias removal) can be written as

$$\mathbf{w}_i^+ = Eh_k(y_i)\mathbf{x} - (\mathbf{I}_n + \mathbf{C}_{\boldsymbol{\nu}})Eh_k'(y_i)\mathbf{w}_i, \qquad (42)$$

$$\mathbf{w}_i \leftarrow \frac{\mathbf{w}_i^+}{\|\mathbf{w}_i^+\|}, \qquad (43)$$

where some choices for the nonlinearities h_k include $h_1(u) = \tanh(u)$, $h_2(u) = u\exp(-u^2/2)$, and $h_3(u) = u^3$ [8]. For ease of reference, we call this algorithm as M-FastICA (h_i) i.e. Modified FastICA algorithm with nonlinearity h_i.

The results of this simulation are shown in Figure 3. In particular, Figure 3 (a) plots the evolution of performance index (averaged over 100 realizations) for various methods when sources are generated from the Uniform distribution. Figure 3(b) shows the same plot for Laplace distributed sources. From these figures, we see that the proposed approach gives a slightly better performance as compared to M-FastICA with different nonlinear functions.

Experiment 2: In the next simulation, we would like to separate sources consisting of signals with asymmetric distributions. In this computer experiment, we compare our approach with the standard FastICA algorithm (with different nonlinear functions). Again, we mix five sources by using a randomly chosen mixing matrix. Figure 4 (a) shows the evolution of performance index for various methods when sources are distributed according to Rayleigh(1) distribution. It is a skewed distribution, which in employed for instance in communications to model the envelope of the fading

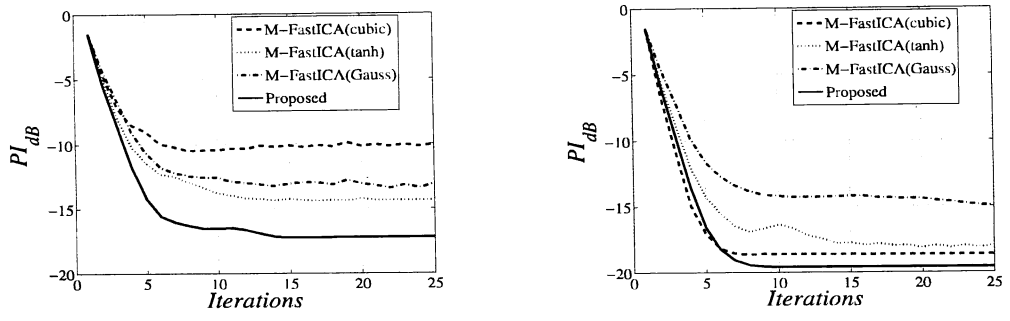


Fig. 3 Adaptation dynamics of average performance index for various methods in Experiment 1 when sources are (a) Uniformly distributed, and (b) Laplace distributed.

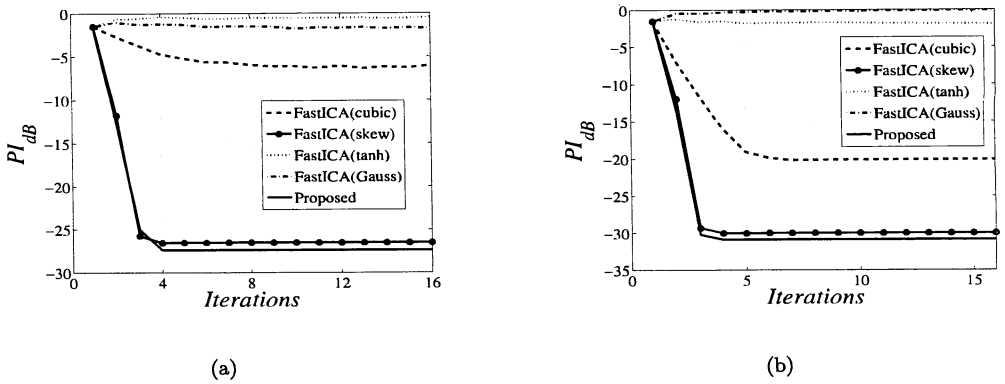


Fig. 4 Adaptation dynamics of average performance index for various methods in Experiment 2 when sources are (a) Rayleigh(1) distributed, and (b) Poisson(0.5) distributed.

channel. Similarly, Figure 4 (b) displays the evolution of performance index when sources are generated from Poisson(0.5) distribution. The Poisson distribution, in addition to be skewed, is also not absolutely continuous.

From Figure 4, we see that odd nonlinearities may perform poorly for the above mentioned sources. Although, in case of Poisson distribution, the cubic nonlinearity gives satisfactory results, its performance is much inferior to that of the ecf based non-Gaussianity measure and the skew nonlinearity. From this simulation, we observe that only a single nonlinearity may not work well for all the pdfs. Even though, the skew function performed very well for the asymmetric sources, we do not expect it to work for symmetric distributions. On the other hand, the ecf based non-Gaussianity measure has the potential to work for many different pdfs.

7. Conclusion

In this paper, we have employed an ecf based non-Gaussianity measure as a one-unit contrast function in order to extract statistically independent signals from their linear mixtures. This objective function is optimized by employing a simple fixed-point algorithm. Some computer simulations are presented in order to compare the separation per-

formance of different non-Gaussianity measures. These simulation results show that provided the sample points are chosen properly, the ecf based contrast function has the potential to work well for wide class of probability distributions. The selection of adequate sample points can be avoided if we utilize the non-Gaussianity measure described by Eq. 15, since it considers the whole frequency range. Obtaining an efficient algorithm for the optimization of this measure is a task of future work.

参考文献

1) A. Cichoki, and S. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, John Wiley and Sons, UK, 2002.

2) H. Attias, "Independent factor analysis," Neural Computation, vol. 11, no. 4, pp. 803-851, 1999.

3) A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," IEEE Transactions on Neural Networks, vol. 10, no. 3, pp. 626-634, 1999.

4) A. Hyvarinen, "New approximations of differential entropy for independent component analysis and projection pursuit," Advances in Neural Information Processing Systems, vol. 10, pp. 273-279, MIT Press, 1998.

5) J. Yu, "Empirical characteristic function estimation and its applications," Econometric Reviews, vol. 22, no. 2, pp. 93-123, 2004.

6) J.Eriksson, and K. Koivunen, "Characteristic-function-based independent component analysis," Signal Processing, vol. 83, no. 10, pp. 2195-2208, 2003.

7) J. Cardoso, and A. Souloumiac, "Blind beamforming for non-Gaussian signals," IEE Proceedings F, vol. 140, no. 6, pp. 362-370, 1993.

8) A. Hyvarinen, "Gaussian moments for noisy independent component analysis," IEEE Signal Processing Letters, vol. 6, no. 6, pp. 145-147, 1999.

9) A. Feuerverger, and R. A. Mureika, "The empirical characteristic function and its applications," The Annals of Statistics, vol. 5, no. 1, pp. 88-97, 1977.

10) B. Klar, and S. G. Meintanis, "Tests for normal mixtures based on the empirical characteristic function," Computational Statistics and Data Analysis, vol. 49, pp. 227-242, 2005.

11) I. A. Koutrouvelis, "A goodness-of-fit test of simple hypotheses based on the empirical characteristic function," Biometrika, vol. 67, no. 1, pp. 238-240, 1980.

12) A. Hyvarinen, "One unit contrast functions for independent component analysis: A statistical analysis," Neural Networks for Signal Processing VII, Proc. IEEE NNSP Workshop, pp. 388-397, 1997.