EcoICA: Skewness-based ICA via Eigenvectors of Cumulant Operator

Liyan Song   LXS189@CS.BHAM.AC.UK
Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China
School of Computer Science, the University of Birmingham, Birmingham, UK

Haiping Lu   HPLU@IEEE.ORG
Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China

Abstract

Independent component analysis (ICA) is an important unsupervised learning method. Most popular ICA methods use *kurtosis* as a metric of non-Gaussianity to maximize, such as FastICA and JADE. However, their assumption of kurtosic sources may not always be satisfied in practice. For weak-kurtosic but skewed sources, kurtosis-based methods could fail while skewness-based methods seem more promising, where *skewness* is another non-Gaussianity metric measuring the non-symmetry of signals. Partly due to the common assumption of signal symmetry, skewness-based ICA has not been systematically studied in spite of some existing works. In this paper, we take a systematic approach to develop EcoICA, a new skewness-based ICA method for weak-kurtosic but skewed sources. Specifically, we design a new *cumulant operator*, define its *eigenvalues* and *eigenvectors*, reveal their connections with the ICA model to formulate the EcoICA problem, and use Jacobi method to solve it. Experiments on both synthetic and real data show the superior performance of EcoICA over existing kurtosis-based and skewness-based methods for skewed sources. In particular, EcoICA is less sensitive to sample size, noise, and outlier than other methods. Studies on face recognition further confirm the usefulness of EcoICA in classification.

Keywords: Independent Component Analysis, Cumulant Operator, Skewness, Eigenvectors

1. Introduction

Independent component analysis (ICA) is a popular unsupervised learning method with many applications (Hyvärinen and Oja, 2000). In order to estimate independent components (ICs), ICA assumes that the sources are mutually independent, where such independence can be measured by *non-Gaussianity*. The mainstream approach for maximizing non-Gaussianity is based on *kurtosis*,[1] which is the *fourth-order univariate cumulant* (Hyvärinen, 1999). FastICA (Hyvärinen, 1999) (with contract function *pow3*) employs kurtosis as a measure of non-Gaussianity and explicitly maximizes a kurtosis-based objective function. The joint approximate diagonalization of eigenmatrices (JADE) (Cardoso and Souloumiac, 1993) jointly diagonalizes the eigenmatrices derived from the *fourth-order cumulant tensor* (the multivariate version of kurtosis) to recover the sources.

Kurtosis-based ICA methods have two implicit assumptions: 1) the non-Gaussianity of random process mainly comes from kurtosis; and 2) many random processes are naturally symmetric so their skewness would decay to zero (Nikias and Mendel, 1993). In other words, they assume that the *probability density functions (PDFs)* of sources are *kurtosic* and highly *symmetric*. The former

---

1. Kurtosis/skewness can be positive or negative. When talking about largeness/smallness, we refer to absolute values.

(*a*) Four weak-kurtosic source images.

(*b*) Mixed images via a 4×4 random matrix.

(*c*) Separated images by FastICA.
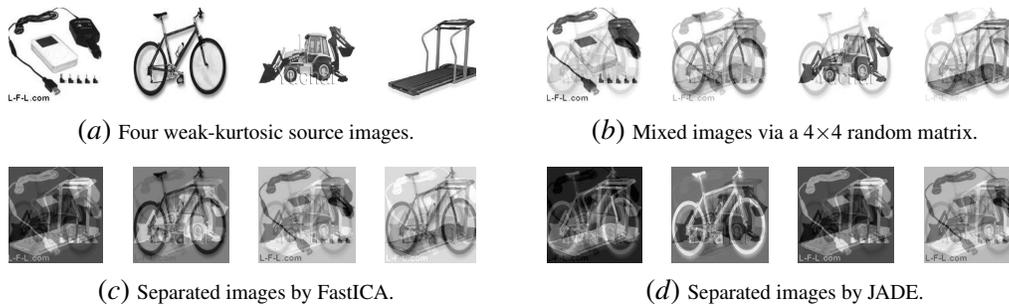
(*d*) Separated images by JADE.

Figure 1: A typical weak-kurtosic image separation case where both FastICA and JADE fail. The source kurtosis are small: 0.44, 0.39, 0.14, and 0.41. In contrast, the source skewness (absolute value) are large: -1.48, -1.34, -1.35, and -1.45. Thus, these sources are *weak-kurtosic but skewed*.
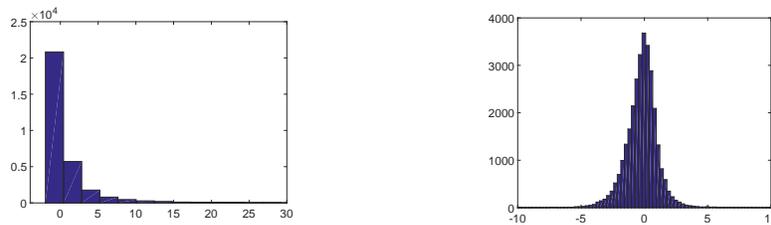

Figure 2: Histograms of kurtosis (left) and skewness (right) of all 30,607 Caltech256 images. About 48.08% have small kurtosis ($|\text{kurtosis}| < 1$), among which 22.70% have large skewness ($|\text{skewness}| > 1$).

assumption leads to the popularity of kurtosis-based approach, while the latter one discourages skewness-based approach.

However, both assumptions can be violated in real applications. On one hand, there are weak-kurtosic sources such as about 48.08% images from the Caltech256 database (Griffin et al., 2006). Intuitively, if the non-Gaussianity of the sources is not mainly from kurtosis, kurtosis-based ICA methods can hardly perform well. For illustration, we show the performance of FastICA and JADE in recovering weak-kurtosic source images of Fig. 1(*a*) from their mixtures in Fig. 1(*b*). As seen from Fig. 1(*c*) and Fig. 1(*d*), neither FastICA nor JADE can separate the mixed images. On the other hand, some real-world data such as the fMRI data (Stone et al., 2002) are shown to be non-symmetric, for which skewness can be exploited to improve the recovery performance.

Therefore, this paper focuses on *ICA for weak-kurtosic but skewed sources*, for which kurtosis-based ICA methods are expected to fail and the skewness-based ones are more promising. To give an example for such weak-kurtosic but skewed data, we depict the histograms of kurtosis and skewness for all Caltech256 images in Fig. 2: Out of the total 30,607 images, around 48.08% (14,717 images) have small kurtosis ($|\text{kurtosis}| < 1$); out of these weak-kurtosic images, 22.70% (3,341 images) have large skewness ($|\text{skewness}| > 1$). Hence, skewness-based ICA could best suit these 3,341 images, which violate the assumptions of kurtosis-based ICA. This motivates us to explore skewness for ICA problem with weak-kurtosic but skewed sources, where kurtosis-based ICA is expected to fail.

Several existing works have studied skewness-based ICA. However, most of them (Choi et al., 1998; Lathauwer et al., 2001; Geng et al., 2014) focused on efficiency improvement, and none of them was shown to outperform FastICA/JADE in accuracy. FastICA$_{skew}$ (Ollila, 2010) (classical FastICA with contrast function *skew*) is another state-of-the-art skewness-based method, but it has a

stability issue and is sensitive to noise. Furthermore, none of existing skewness-based methods provides a systematic treatment in terms of spectral analysis as in JADE (Cardoso, 1999). On the other hand, a few ICA methods study the third-order cumulants, the multivariate version of skewness. However, rather than working with a *linear operator* in a systematic way as we do, existing third-order cumulant-based ICA methods mainly involve diagonalizing the cumulant tensors (Comon, 1994b; Moreau, 2000, 2007; Lathauwer et al., 2001; Moreau, 2001; Blaschke and Wiskott, 2004; Wang and Lu, 2005). In particular, some of them focus on the mixed cumulants of different orders. For instance, Moreau (2001) combines the third-order cumulant tensor with the fourth one and optimizes them simultaneously, while Blaschke and Wiskott (2004) minimize a weighted summation based on the third- and fourth-order tensors.

Based on the above, we propose EcoICA, a new skewness-based ICA method for weak-kurtosic but skewed sources. We develop EcoICA in a systematic way with two main contributions: (1) We design a new *cumulant operator* $\mathscr{T}$ based on the third-order cumulant tensor $\mathcal{Q}_{\mathbf{x}}$. To explore the spectral properties of $\mathscr{T}$, we further define its *eigenvalues* as well as *eigenvectors*, and reveal their connections with the ICA model. This new cumulant operator enables a systematic approach to explore the third-order cumulants and other higher-order cumulants. (2) We compute the **e**igenvectors of **c**umulant **o**perator to tackle the **ICA** problem (**EcoICA**), which can be solved by the Jacobi method. Experimental results show that EcoICA outperforms kurtosis-based and skewness-based methods for skewed sources, and it is less sensitive to sample size, noise, and outlier in particular.

2. Preliminaries

2.1. Cumulants

The r-th order cumulants are conventionally denoted by $\kappa_r(x)$ for random variable x, and $(\mathcal{Q}_{\mathbf{x}})_{i_1,\cdots,i_r}$ or $cum(x_{i_1}, \cdots, x_{i_r})$ for random vector $\mathbf{x}$, where $i_1, \cdots, i_r$ are the mode-wise indices. In particular, the *third/fourth* cumulants of a zero-mean random variable are the *skewness/kurtosis*; $\kappa_r(x_i) = cum(x_i, \cdots, x_i)$ (r times) is the r-th cumulant of x_i.

Cumulants of a random vector $\mathbf{x}$ have the following properties: (1) *symmetry*: $(\mathcal{Q}_{\mathbf{x}})_{i_1,\cdots,i_r} = (\mathcal{Q}_{\mathbf{x}})_{i_{\sigma(1)},\cdots,i_{\sigma(r)}}$ for any permutation σ; (2) *linearity*: $cum(x_1, \cdots, x_i+y, \cdots, x_r) = cum(x_1, \cdots, x_r) + cum(x_1, \cdots, y, \cdots, x_r)$ and $cum(x_1, \cdots, \alpha x_i, \cdots, x_r) = \alpha \cdot cum(x_1, \cdots, x_r)$ for any random variable y and scalar α; (3) *independence*: if $\exists p, q \in \{1, \cdots, r\}$ where x_{i_p} and x_{i_q} are independent, $(\mathcal{Q}_{\mathbf{x}})_{i_1,\cdots,i_r} = 0$; (4) *vanishing Gaussian*: if $\mathbf{x}$ is Gaussian, $(\mathcal{Q}_{\mathbf{x}})_{i_1,\cdots,i_r} = 0$ for any order $r \geq 3$.

2.2. The ICA Model

The standard ICA model is formulated as $\mathbf{x} = \mathbf{As}$, where $\mathbf{x} = [x_1, \cdots, x_N]^T$ denotes the observed mixture variables, $\mathbf{s} = [s_1, \cdots, s_N]^T$ are the independent components (ICs), N is the number of sources or observations (usually assumed to be equal), and $\mathbf{A}$ is the unknown constant matrix, namely *mixing matrix*. The aim of ICA is to estimate sources $\mathbf{s}$ and the mixing matrix $\mathbf{A}$ simultaneously using observations $\mathbf{x}$ only, assuming that the source components are mutually independent.

2.2.1. BASIC ICA PROCEDURES

There are three steps for ICA. (1) *Centering*: remove the first-order statistics from the data by shifting the sample mean to the origin. (2) *Whitening*: remove the second-order statistics from the data. Suppose the observed data in the ICA model is whitened by matrix $\mathbf{W}$ such that $\mathbf{z} = \mathbf{WAs} =$

Us, where $\mathbf{U} = \mathbf{WA}$ is the *whitened mixing matrix*. Since $\mathbf{I} = E\{\mathbf{zz}^T\} = E\{\mathbf{Uss}^T\mathbf{U}^T\} = \mathbf{UU}^T$, the whitened mixing matrix $\mathbf{U}$ is orthogonal. (3) **ICA Estimation**: use higher-order statistics of the data to estimate ICs. It is the core step of ICA, and different ICA methods do it differently.

2.2.2. THE WHITENED ICA MODEL

For simpler notation, we keep using $\mathbf{x}$ for the centred and whitened observations in the remaining of this paper. The whitened ICA model is

$$\mathbf{x} = \mathbf{Us} = \sum_{i=1}^{N} \mathbf{u}_i s_i, \tag{1}$$

where $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_N]$ is the *whitened mixing matrix* and $\mathbf{u}_i$ is the i-th column of $\mathbf{U}$. The whitened ICA model (1) satisfies the following equations for $\forall i, j \in \{1, \cdots, N\}$:

(1) $x_i = \sum_k U_{ik} s_k$;
(2) $\mathbf{u}_i^T \mathbf{u}_j = \sum_k \mathbf{U}_{ki}\mathbf{U}_{kj} = \sum_k \mathbf{U}_{ik}\mathbf{U}_{jk} = \delta_{ij}$ due to the orthogonality of $\mathbf{U}$.

3. Eigenvectors of Cumulant Operator for ICA (EcoICA)

In this section, we first define and systematically study a new cumulant operator $\mathscr{T}$ based on the third-order cumulant tensor $\mathcal{Q}_{\mathbf{x}}$. Then we derive a novel ICA method based on the eigenvectors of this newly designed operator.

3.1. A New Cumulant Operator $\mathscr{T}$

Definition 1 *The cumulant operator $\mathscr{T}$ is defined by the third-order cumulant tensor $\mathcal{Q}_{\boldsymbol{x}}$ of mixture components in the whitened ICA model* (1):

$$\mathscr{T}: \quad \mathbb{R}^N \quad \to \quad \mathbb{R}^{N \times N}$$
$$\boldsymbol{v} \quad \to \quad \mathscr{T}(\boldsymbol{v}), \mathscr{T}_{ij}(\boldsymbol{v}) := \sum_k (\mathcal{Q}_{\boldsymbol{x}})_{ijk} \cdot v_k.$$

Theorem 1 *The cumulant operator $\mathscr{T}$ is linear and symmetric.*

Proof (i) For $\forall \alpha \in \mathbb{R}$ and $\forall i, j \in \{1, \cdots, N\}$, we have

$$\mathscr{T}_{ij}(\alpha\mathbf{v}) = \sum_k (\mathcal{Q}_{\mathbf{x}})_{ijk}(\alpha v_k) = \alpha \sum_k (\mathcal{Q}_{\mathbf{x}})_{ijk}(v_k) = \alpha \mathscr{T}_{ij}(\mathbf{v}).$$

The second equality is due to the constancy of α with respect to the summation over k. So $\mathscr{T}(\alpha\mathbf{v}) = \alpha\mathscr{T}(\mathbf{v})$ holds.

(ii) For $\forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^N$ and $\forall i, j \in \{1, \cdots, N\}$, we have

$$\mathscr{T}_{ij}(\mathbf{v} + \mathbf{w}) = \sum_k (\mathcal{Q}_{\mathbf{x}})_{ijk} \cdot (\mathbf{v} + \mathbf{w})_k = \sum_k (\mathcal{Q}_{\mathbf{x}})_{ijk} \cdot (v_k + w_k) = \mathscr{T}_{ij}(\mathbf{v}) + \mathscr{T}_{ij}(\mathbf{w}).$$

So it is proved that $\mathscr{T}(\mathbf{v} + \mathbf{w}) = \mathscr{T}(\mathbf{v}) + \mathscr{T}(\mathbf{w})$.

(iii) Linear operator $\mathscr{T}$ is symmetric due to the symmetry of the cumulant tensor $(\mathcal{Q}_{\mathbf{x}})_{ijk}$ where the order of the indices makes no difference. ∎

3.2. Eigenvalues and Eigenvectors of $\mathscr{T}$

Definition 2 *If a unit vector $\boldsymbol{v} \in \mathbb{R}^N$ and a scalar λ satisfy the following condition:*

$$\mathscr{T}(\boldsymbol{v}) = \lambda \boldsymbol{v}\boldsymbol{v}^T, \text{ where } ||\boldsymbol{v}|| = 1, \tag{2}$$

*i.e. $\mathscr{T}_{ij}(\boldsymbol{v}) = \lambda(\boldsymbol{v}\boldsymbol{v}^T)_{ij}$ for $\forall i, j \in \{1, \cdots, N\}$, we call $\boldsymbol{v}$ the **eigenvector** and λ the corresponding **eigenvalue** of the cumulant operator $\mathscr{T}$.*

Note that the condition $||\mathbf{v}|| = 1$ is important to avoid the ambiguity of scale. Given an eigenvector $\mathbf{v}$ and its eigenvalue λ, denote $\mathbf{w} = \alpha\mathbf{v}$ ($\forall \alpha \neq 0$ and $\alpha \neq 1$), we have $\mathscr{T}(\mathbf{w}) = \mathscr{T}(\alpha\mathbf{v}) = \alpha\mathscr{T}(\mathbf{v}) = \alpha\lambda\mathbf{v}\mathbf{v}^T = \frac{\lambda}{\alpha}\mathbf{w}\mathbf{w}^T$, which is identical to Eq. (2) in formula. However, $\mathbf{w}$ is not an eigenvector of $\mathscr{T}$ as $||\mathbf{w}|| = |\alpha| \cdot ||\mathbf{v}|| = |\alpha| \neq 1$.

Lemma 1 *Given the whitened ICA model (1), the third-order cumulant tensor $\mathcal{Q}_{\boldsymbol{x}}$ satisfies the following equation:*

$$(\mathcal{Q}_{\boldsymbol{x}})_{ijk} = \sum_p U_{ip}U_{jp}U_{kp}\kappa_3(s_p),$$

where $\kappa_3(s_p)$ represents the skewness of the p-th source component s_p.

Proof This lemma can be easily proved via properties of cumulants in Sec. 2.1. ∎

Theorem 2 *Given the whitened ICA model (1), the columns of the whitened mixing matrix $\mathbf{U}$ consist of the **full** eigenvector space of the cumulant operator $\mathscr{T}$. In other words, (i) any column $\boldsymbol{u}_m$ ($\forall m \in \{1, \cdots, N\}$) of $\boldsymbol{U}$ is an eigenvector of $\mathscr{T}$; and (ii) $\{\boldsymbol{u}_m\}_{m=1}^N$ are $\mathscr{T}$'s full eigenvectors.*

Proof We will prove this theorem in two parts (i) and (ii).

(i) We first prove that any column of $\mathbf{U}$ is an eigenvector of $\mathscr{T}$. From the orthogonality of $\mathbf{U}$, $||\mathbf{u}_m|| = 1$ satisfies the unit length requirement. By Def. 1, we have:

$$\begin{aligned}\mathscr{T}_{ij}(\mathbf{u}_m) &= \sum_k cum(x_i, x_j, x_k)(\mathbf{u}_m)_k = \sum_k cum(\sum_p U_{ip}s_p, \sum_q U_{jq}s_q, \sum_r U_{kr}s_r)U_{km} \\ &= \sum_k U_{km}\sum_{p,q,r} U_{ip}U_{jq}U_{kr}cum(s_p, s_q, s_r).\end{aligned}$$

Due to the independence of $\{s_i\}$, only those cumulants with $p = q = r$ are nonzero. Thus, we have

$$\mathscr{T}_{ij}(\mathbf{u}_m) = \sum_k U_{km}\sum_p U_{ip}U_{jp}U_{kp}\kappa_3(s_p) = \sum_p \kappa_3(s_p)U_{ip}U_{jp}(\sum_k U_{km}U_{kp}).$$

Since $\mathbf{U}$ is orthogonal, $\sum_k U_{km}U_{kp} = \delta_{mp}$ holds and only those with $p = m$ are nonzero. Thus,

$$\mathscr{T}_{ij}(\mathbf{u}_m) = \sum_p \kappa_3(s_p)U_{ip}U_{jp}\delta_{mp} = \kappa_3(s_m)U_{im}U_{jm} = \kappa_3(s_m)(\mathbf{u}_m\mathbf{u}_m^T)_{ij}.$$

According to Def. 2, we have shown that $\{\mathbf{u}_m\}$ are $\mathscr{T}$'s eigenvectors, and their eigenvalues $\{\kappa_3(s_m)\}$ are the skewness of the independent components $\{s_m\}$.

Since columns of $\mathbf{U}$ are orthogonal, they can be a full basis of the $\mathbb{R}^N$ space, i.e. $\mathbb{R}^N = span\{\mathbf{u}_1, \cdots, \mathbf{u}_N\}$. In conventional linear algebra, it indicates $\{\mathbf{u}_m\}$ are the full eigenvectors. We prove the similar property of $\{\mathbf{u}_m\}$ for the cumulant operator $\mathscr{T}$ in the second part of this proof.

(ii) We now prove that $\{\mathbf{u}_m\}$ are $\mathscr{T}$'s full eigenvectors by showing that any eigenvector of $\mathscr{T}$ is indeed one of $\{\mathbf{u}_m\}$. Given $\mathscr{T}$'s any eigenvalue λ and eigenvector $\mathbf{v} \in \mathbb{R}^N = span\{\mathbf{u}_1, ..., \mathbf{u}_N\}$, there are $\alpha_1, \cdots, \alpha_N \in \mathbb{R}$ such that

$$\mathbf{v} = \frac{1}{\gamma}\sum_{k=1}^N \alpha_k\mathbf{u}_k, \tag{3}$$

where $\gamma = \sqrt{\sum_{k=1}^{N} \alpha_k^2} > 0$. Thus, we have $||\mathbf{v}|| = 1$.

Before continuing the proof, we need to derive an alternative form of γ for the following derivation. Since $||\mathbf{v}||^2 = \mathbf{v}^T \mathbf{v} = \gamma^{-2} (\sum_m \alpha_m \mathbf{u}_m^T)(\sum_m \alpha_m \mathbf{u}_m)$, we have

$$\gamma^2 = \sum_{m,n} \alpha_m \alpha_n \mathbf{u}_m^T \mathbf{u}_n = \sum_{m,n} \alpha_m \alpha_n (\sum_k U_{km} U_{kn}) = \sum_k (\sum_m \alpha_m U_{km})^2. \tag{4}$$

Therefore, γ satisfies the following equation:

$$\gamma^2 = \sum_k \alpha_k^2 = \sum_k (\sum_m \alpha_m U_{km})^2 > 0. \tag{5}$$

Let us return the proof. On one hand, we have

$$\mathscr{T}(\mathbf{v}) = \mathscr{T}(\gamma^{-1} \sum_k \alpha_k \mathbf{u}_k) = \gamma^{-1} \sum_k \alpha_k \mathscr{T}(\mathbf{u}_k) = \gamma^{-1} \sum_k \alpha_k \kappa_3(s_k) \mathbf{u}_k \mathbf{u}_k^T.$$

That is to say, for $\forall i, j$, the following equality holds

$$\mathscr{T}_{ij}(\mathbf{v}) = \gamma^{-1} \sum_k \alpha_k \kappa_3(s_k)(\mathbf{u}_k \mathbf{u}_k^T)_{ij} = \gamma^{-1} \sum_k \alpha_k \kappa_3(s_k) U_{ik} U_{jk}. \tag{6}$$

On the other hand, according to Eq. (3),

$$\lambda \mathbf{v} \mathbf{v}^T = \lambda/\gamma^2 (\sum_m \alpha_m \mathbf{u}_m)(\sum_n \alpha_n \mathbf{u}_n)^T = \lambda/\gamma^2 \sum_{m,n} \alpha_m \alpha_n \mathbf{u}_m \mathbf{u}_n^T.$$

In other words, it holds for $\forall i, j$ that

$$(\lambda \mathbf{v} \mathbf{v}^T)_{ij} = \lambda/\gamma^2 \sum_{m,n} \alpha_m \alpha_n (\mathbf{u}_m \mathbf{u}_n^T)_{ij} = \lambda/\gamma^2 (\sum_m \alpha_m U_{im})(\sum_m \alpha_m U_{jm}). \tag{7}$$

Since $\mathbf{v}$ is an eigenvector, we have for $\forall i, j$

$$\mathscr{T}_{ij}(\mathbf{v}) = (\lambda \mathbf{v} \mathbf{v}^T)_{ij}.$$

Substituting (6) and (7) into the above equation, we have

$$\gamma^2 \sum_m \alpha_m \kappa_3(s_m) U_{im} U_{jm} = \lambda\gamma (\sum_m \alpha_m U_{im})(\sum_m \alpha_m U_{jm}). \tag{8}$$

For $i = j$, Eq. (8) can be simplified to

$$\gamma^2 \sum_m \alpha_m \kappa_3(s_m) U_{im}^2 = \lambda\gamma (\sum_m \alpha_m U_{im})^2.$$

Summing across all indices i, based on Eq. (5) and the orthogonality of $\mathbf{U}$, we have

$$\begin{array}{ll} \sum_i \gamma^2 \sum_m \alpha_m \kappa_3(s_m) U_{im}^2 & = \sum_i \lambda\gamma (\sum_m \alpha_m U_{im})^2 \\ \gamma^2 \sum_m \alpha_m \kappa_3(s_m) \sum_i U_{im}^2 & = \lambda\gamma \sum_i (\sum_m \alpha_m U_{im})^2 \\ \gamma^2 \sum_m \alpha_m \kappa_3(s_m) \delta_{mm} & = \lambda\gamma\gamma^2. \end{array}$$

Therefore, λ satisfies

$$\lambda\gamma = \sum_m \alpha_m \kappa_3(s_m). \tag{9}$$

Substituting λ of Eq. (9) into Eq. (8) and using Eq. (5), we get

$$(\sum_m \alpha_m^2)[\sum_m \alpha_m \kappa_3(s_m)U_{im}U_{jm}] = [\sum_m \alpha_m \kappa_3(s_m)](\sum_m \alpha_m U_{im})(\sum_m \alpha_m U_{jm}). \qquad (10)$$

To guarantee the equality of (10), the coefficient of the LHS of the entry $\kappa_3(s_m)U_{im}U_{jm}$ should equal to the RHS of this entry for $\forall i, j, m$. Formally, we have

$$(\sum_k \alpha_k^2)\alpha_m = \alpha_m \alpha_m \alpha_m.$$

Thus, we conclude that either $\alpha_m = 0$ or $\alpha_m^2 = \sum_k \alpha_k^2 = \alpha_m^2 + \sum_{k \neq m} \alpha_k^2$ for $\alpha_m \neq 0$. That is to say, there exists and only exists one **nonzero** coefficient α_k where $k \in \{1, \cdots, N\}$. In this case, eigenvector $\mathbf{v}$ is indeed some $\mathbf{u} \in \{\mathbf{u}_1, ..., \mathbf{u}_N\}$. Therefore, we have proved that $\{\mathbf{u}_m\}$ constitute the full eigenvector space of the cumulant operator $\mathscr{T}$. ∎

3.3. The Objective Function of EcoICA

According to Theorem 2, once the *full eigenvectors* of the *cumulant operator* $\mathscr{T}$ are known, we can recover the whitened mixing matrix $\mathbf{U}$ and the independent components. Thus, computing *eigenvectors* of *cumulant operator* is the core of our **ICA** method, and we name it **EcoICA**. However, their exact computation is not always feasible because the ICA model does not hold exactly in real applications (Hyvärinen et al., 2001). Inspired by JADE, we further convert computing $\mathscr{T}$'s eigenvectors by minimizing the off-diagonal entries of several matrices based on Theorem 3 below.

Theorem 3 *Given the whitened ICA model* (1)*, the matrix* $\mathbf{U}^T \mathscr{T}(\mathbf{v})\mathbf{U}$ *is diagonal for* $\forall \mathbf{v} \in \mathbb{R}^N$.

Proof For $\forall i, j \in \{1, \cdots, N\}$ and $\forall \mathbf{v} \in \mathbb{R}^N$, by the definition of matrix multiplication, we have

$$[\mathbf{U}^T \mathscr{T}(\mathbf{v})\mathbf{U}]_{ij} = \sum_{p,q} U_{ip}^T \mathscr{T}_{pq}(\mathbf{v})U_{qj} = \sum_{p,q} U_{pi}U_{qj}[\sum_k v_k cum(x_p, x_q, x_k)].$$

We reformulate the above based on Lemma 1 as

$$\begin{aligned}[\mathbf{U}^T \mathscr{T}(\mathbf{v})\mathbf{U}]_{ij} &= \sum_{p,q} U_{pi}U_{qj}\{\sum_k v_k[\sum_r U_{pr}U_{qr}U_{kr}\kappa(s_r)]\} \\ &= \sum_r \kappa_3(s_r)\sum_k v_k U_{kr}(\sum_p U_{pi}U_{pr})(\sum_q U_{qj}U_{qr}) = \sum_r \kappa_3(s_r)\sum_k v_k U_{kr}\delta_{ir}\delta_{jr}.\end{aligned}$$

The last equality is due to $\mathbf{U}$'s orthogonality. Only those with $i = j = r$ are nonzero so we have

$$[\mathbf{U}^T \mathscr{T}(\mathbf{v})\mathbf{U}]_{ij} = \begin{cases} 0 & \text{for } i \neq j \\ \kappa_3(s_i)\sum_k v_k U_{ki} & \text{for } i = j. \end{cases}$$

∎

We further notice that the diagonal entries can be neater:

$$[\mathbf{U}^T \mathscr{T}(\mathbf{v})\mathbf{U}]_{ii} = \kappa_3(s_i)[\sum_k v_k(\mathbf{u}_i)_k] = \kappa_3(s_i)\langle \mathbf{u}_i, \mathbf{v}\rangle. \qquad (11)$$

If we view the columns of $\mathbf{U}$ as one full basis of the $\mathbb{R}^N$ space, $\langle \mathbf{u}_i, \mathbf{v}\rangle$ can be interpreted as the i-th *coordinate* of the vector $\mathbf{v}$ under this basis.

Based on Theorem 3, we can take a set of vectors $\{\mathbf{v}_n\}$ and make matrix set $\{\mathbf{U}^T \mathscr{T}(\mathbf{v}_n)\mathbf{U}\}$ as diagonal as possible. In this paper, we use the standard basis $\{\mathbf{e}_n\}_{n=1}^N$ of the $\mathbb{R}^N$ space as $\{\mathbf{v}_n\}$. The diagonality of a symmetric matrix $\mathbf{Q} = \mathbf{U}^T \mathscr{T}(\mathbf{v})\mathbf{U}$ can then be measured by the sum of the

---

Algorithm 1 EcoICA: ICA for skewed sources

1: **Input:** Observation matrix $\mathbf{X}$: rows correspond to mixture components and columns are samplings.
2: **Preprocessing:** Centering $\mathbf{X}$ and whitening $\mathbf{X}$.
3: **Estimation:**
4:    (1) Compute the empirical third-order cumulant tensor $\mathcal{Q}_{\mathbf{x}}$.
5:    (2) Compute $\{\mathscr{T}(\mathbf{e}_n)\}_{n=1}^N$ based on Def. 1, where $\{\mathbf{e}_n\}$ are the full standard basis of $\mathbb{R}^N$.
6:    (3) Align the matrices $\{\mathscr{T}(\mathbf{e}_n)\}$ to form the matrix $\mathbf{M} = [\mathscr{T}(\mathbf{e}_1), ..., \mathscr{T}(\mathbf{e}_N)]$.
7:    (4) Apply the Jacobi method on $\mathbf{M}$ to obtain rotation matrices.
8:    (5) Get $\mathbf{U}^{-1}$ by multiplication of these Jacobi rotation matrices.
9:    (6) Invert $\mathbf{U}^{-1}$ to get $\mathbf{U}$.
10: **Output:** the whitened mixing matrix $\mathbf{U}$.

---

squares of off-diagonal entries: $\sum_{i \neq j} q_{ij}^2$ (Comon, 1994a). Since for a given vector $\mathbf{v}$, the square sum over all elements of the matrix is preserved under an orthogonal transformation, minimizing the sum of squares of off-diagonal elements is equivalent to maximizing the sum of squares of diagonal elements (Deco and Obradovic, 1996). We formulate our objective function based on this property.

The objective function: For the set of standard basis $\{\mathbf{e}_n\}_{n=1}^N$ of the $\mathbb{R}^N$ space, we maximize the following objective with respect to the orthogonal matrix $\mathbf{U}$

$$\mathcal{F}_{EcoICA}(\mathbf{U}) = \sum_{n=1}^N ||diag(\mathbf{U}^T \mathscr{T}(\mathbf{e}_n)\mathbf{U})||^2, \tag{12}$$

where $||diag(\cdot)||^2$ denotes the sum of squares of the diagonal.

3.4. The EcoICA Algorithm

Similar to JADE, we apply the Jacobi method to optimize our objective function (12) and compute the whitened mixing matrix $\mathbf{U}$. Specifically, we first align the set of matrices $\{\mathscr{T}(\mathbf{e}_n)\}_{n=1}^N$ into an extended matrix $\mathbf{M} = [\mathscr{T}(\mathbf{e}_1), \cdots, \mathscr{T}(\mathbf{e}_N)]$. Then, we apply the Jacobi method on $\mathbf{M}$ to conduct a series of Jacobi rotations, each of which handles two rows and two columns at a time (Clarkson, 1988). We then obtain $\mathbf{U}^{-1}$, the inverse of the whitened mixing matrix, by multiplying the rotation matrices. Subsequently, we get $\mathbf{U}$. The EcoICA algorithm is summarized in Algorithm 1.

EcoICA is based on the third-order cumulant tensor, whereas JADE is based on the fourth-order one. Thus, EcoICA has lower computational load than JADE, which is verified in experiments.

3.5. Further Extensions and Application Guidelines

EcoICA is endowed with a new third-order cumulant operator, which is, to our best knowledge, the first known extension of the fourth-order cumulant operator in JADE. This allows the exploitation of spectral structures for ICA and further extensions to higher-order cumulant operators. For instance, we can define the fifth-order cumulant operator as a mapping from a third-order tensor to a matrix, and then develop a new algorithm based on it.

As a skewness-based method, EcoICA is designed for problems with skewed sources. On one hand, when the nature of sources is known to be kurtosic or skewed, we can apply kurtosis-based ICA or EcoICA respectively for good performance. On the other hand, when the nature of sources is unknown, we can determine whether EcoICA is appropriate by examining the kurtosis and skewness of the observed data (the mixtures), which indicate those of the sources due to linear mixing (as to

be shown in the end of Sec. 4.2). Specifically, we suggest to apply EcoICA when the observed data have larger skewness (in magnitude) than kurtosis, and apply kurtosis-based ICA when they are more kurtosic. Furthermore, it may be useful to develop hybrid models considering both kurtosis and skewness (Moreau, 2001; Blaschke and Wiskott, 2004). With a new systematic approach for ICA based on linear operator, EcoICA offers a new approach to combine cumulant operators of different orders, which could be an interesting future direction.

4. Experiments

Since EcoICA is based on skewness, it works best for skewed sources. In this section, we evaluate EcoICA against FastICA (Hyvärinen, 1999), its skewness version FastICA$_{skew}$ (Ollila, 2010), JADE (Cardoso, 1999), and Infomax (Bell and Sejnowski, 1995) using the codes provided by the authors. We mainly study the blind source separation (BSS) problem and compute the Amari error (Amari et al., 1996) of the demixing matrices (i.e., the inverse of the mixing matrices).

We consider three experimental settings: (1) BSS on synthetic data, where source skewness and kurtosis can be fully controlled. Thus, we can set up a **perfect** simulating environment for evaluation. We also study the sensitivity against sample size, noise, and outlier. (2) Image separation, where we can **mildly** control the experiments by calculating the skewness and kurtosis of source images and selecting those of our interest. (3) Face recognition, where the ground-truth of the sources and the mixing matrix are **unknown** so we can only evaluate EcoICA via recognition rate indirectly.

4.1. Blind Source Separation on Synthetic Data

We first design *four source scenarios* to verify the conditions for EcoICA to work: (a) skewed and kurtosic sources, (b) skewed but non-kurtosic sources, (c) non-skewed but kurtosic sources, and (d) non-skewed and non-kurtosic sources. Next, we focus on scenarios (a) and (b) where EcoICA is designed for to study recovery performance versus the number of sources. Third, the effects of sample size, noise, and outlier are investigated.

Take scenario (b) as an example, to simulate skewed but non-kurtosic sources, we set source skewness to one and source kurtosis to zero and randomly generate the sources via *MATLAB* function $pearsrnd()$. To guarantee the invertibility of mixing matrices, they are generated in three steps: (i) uniformly generate a $N \times N$ matrix with each entry between zero and one, where N represents the number of sources; (ii) normalize the generated matrix by column; (iii) add an identity matrix to the one in (ii).

4.1.1. BSS WITH TWO SOURCES

For easy visualization, we investigate BSS with two sources, and each source has 5000 noise-free random samples. We aim to verify the situations where skewness/kurtosis-based methods work.

Figure 3 depicts one example of source recovering processes by Infomax, FastICA, JADE, FastICA$_{skew}$, and EcoICA. Note that ICA estimation is only unique up to sign and permutation. We can clearly see that (1) when sources are both skewed and kurtosic, all ICA methods can successfully recover the source probability density function (PDF); (2) when sources are skewed but non-kurtosic, skewness-based methods, i.e. EcoICA and FastICA$_{skew}$, succeed in recovering the source PDF, but neither kurtosis-based methods nor Infomax can recover; (3) when sources are non-skewed but kurtosic, kurtosis-based methods, i.e. FastICA and JADE, can recover the source

(*a*) Skewed and kurtosic sources. Amari errors of Infomax, FastICA, <u>JADE</u>, FastICA$_{skew}$, and **EcoICA** are 0.055, 0.0665, <u>0.007</u>, 0.017, and **0.0036** respectively.

(*b*) Skewed but non-kurtosic sources. Amari errors of Infomax, FastICA, JADE, <u>FastICA$_{skew}$</u>, and **EcoICA** are 0.6361, 0.1164, 0.1175, <u>0.0117</u>, and **0.0085** respectively.

(*c*) Non-skewed but kurtosic sources. Amari errors for Infomax, <u>FastICA</u>, **JADE**, FastICA$_{skew}$, and EcoICA are 0.9708, <u>0.0214</u>, **0.0128**, 0.195, and 0.2809 respectively.

(*d*) Non-skewed and non-kurtosic sources. Amari errors for Infomax, <u>FastICA</u>, **JADE**, FastICA$_{skew}$, and EcoICA are 0.5901, <u>0.4635</u>, **0.4017**, 0.8282, and 0.9101 respectively.
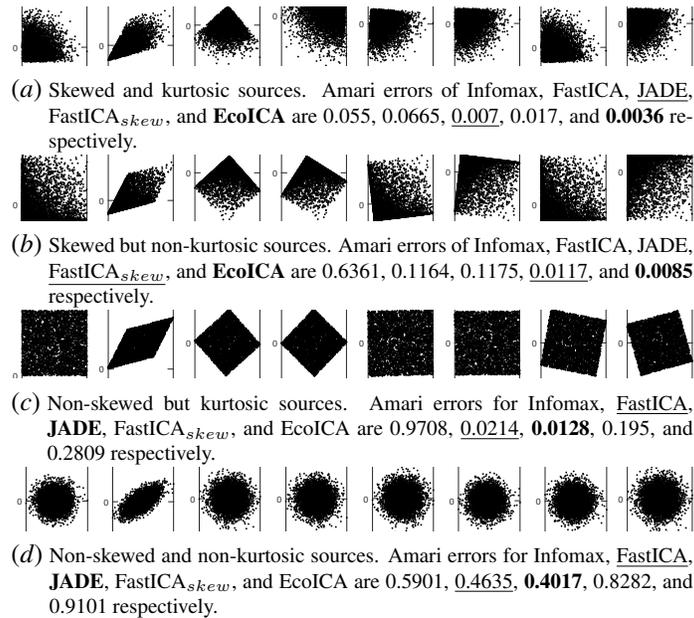
Figure 3: One example of source recovery processes under different source scenarios. The randomly generated mixing matrix is [1.8703, 0.8058; 0.4925, 1.5922]. The first three images of each row are, from left to right: the PDFs of the source, mixed signals, and whitened signals; the rest five of each row are the PDFs of recovered sources via Infomax, FastICA, JADE, FastICA$_{skew}$, and EcoICA. Their Amari errors are listed in the respective subfigure captions with the best (second-best) results in bold (underline).

Table 1: Amari errors (multiplied by 100) for 2-source noise-free BSS with 5000 samples, demonstrating the conditions where skewness/kurtosis-based methods work. Each entry is the mean±std of 100 repetitions. The four scenarios are: (a) skewed and kurtosic sources, (b) skewed but non-kurtosic sources, (c) non-skewed but kurtosic sources, and (d) non-skewed and non-kurtosic sources. The best (second-best) Amari errors are highlighted in bold (underline).

Amari	Infomax	FastICA	JADE	FastICA$_{skew}$	EcoICA
(a)	7.10±5.63	4.66±3.81	3.46±2.72	<u>1.71±1.09</u>	**1.27±0.81**
(b)	80.00±12.55	44.97±28.55	44.51±27.32	<u>1.45±0.91</u>	**1.01±0.55**
(c)	97.57±0.96	<u>1.07±0.67</u>	**0.77±0.46**	44.04±28.39	47.10±28.39
(d)	43.67±28.25	<u>39.72±27.46</u>	**38.64±26.50**	40.43±28.33	42.51±28.96

PDF, but both skewness-based methods and Infomax fail the recovery; (4) when sources are neither skewed nor kurtosic, none of the investigated ICA methods can recover.

Next, we repeat 100 times for each scenario with different mixing matrices. Table 1 reports the average performance. The systematic results are consistent with the previous observation: for skewed sources (a) and (b), skewness-based ICA methods, i.e. EcoICA and FastICA$_{skew}$, can succeed in recovering the sources, but they fail when the sources are non-skewed in (c) and (d) where kurtosis-based methods, i.e. FastICA and JADE, play their role.

Table 2: Amari errors (multiplied by 100) for m-source noise-free BSS with 5000 samples, demonstrating the recovery performance w.r.t. the number of sources. Each entry is the mean±std of 100 repetitions. The two scenarios are the same as Table 1. The best (second-best) Amari errors are highlighted in bold (underline).

Amari	m-source	Infomax	FastICA	JADE	FastICA$_{skew}$	EcoICA
(a) ⇑s, ⇑k	2	7.10±5.63	4.66±3.81	3.46±2.72	<u>1.71±1.09</u>	**1.27±0.81**
	4	23.81±16.37	24.89±18.36	**19.90±15.41**	<u>23.31±17.00</u>	23.70±17.03
	8	33.44±20.27	36.20±20.92	35.18±17.58	<u>33.06±18.04</u>	**31.90±17.95**
	16	39.98±17.22	42.22±16.01	**37.16±17.16**	39.15±18.09	<u>38.63±19.32</u>
(b) ⇑s, ⇓k	2	80.00±12.55	44.97±28.55	44.51±27.32	<u>1.45±0.91</u>	**1.01±0.55**
	4	41.50±19.37	43.51±19.02	44.87±17.21	<u>22.06±18.20</u>	**20.14±14.85**
	8	37.20±17.19	47.43±15.08	44.32±17.09	**34.40±16.36**	<u>34.77±19.96</u>
	16	40.61±15.12	44.27±23.65	44.51±14.96	<u>40.05±19.17</u>	**37.87±18.73**

4.1.2. BSS WITH MORE SOURCES

We tested the algorithms in simulations with more components (i.e. 4, 8, and 16) for scenario (a) and (b) where skewness-based methods are designed for. The other experimental settings are the same to two-source BSS simulation. We report the average Amari errors across 100 repetition with different mixing matrices in Table 2. We can observe that, though all ICA methods deteriorate with more sources, EcoICA achieves the best performance overall. Especially for skewed but non-kurtosic sources, EcoICA can often yield the smallest Amari errors.

4.1.3. SENSITIVITY STUDIES

We further investigate the sensitivity of ICA algorithms with respect to sample size, noise, and outlier. Without loss of generalization, we study two-source BSS where the sources are both skewed and kurtosic so all ICA methods can recover the mixture to some extent. For each study, we generated 100 mixing matrices with 100 corresponding ICA data sets, and report the mean performance.

Effect of sample size: We compare algorithms at various sample sizes for noise-free BSS. The mean performance of each ICA method is shown in Fig. 4(a). We can see that the performance of Infomax, FastICA and JADE drastically deteriorate with small samples, possibly due to their heavier dependence on estimate accuracy of the statistics than skewness-based methods. We also note that EcoICA is superior to FastICA$_{skew}$ constantly especially for smaller sample size.

Effect of noise: Figure 4(b) shows the performance of ICA methods on different noise levels. The observed data are drawn from the noisy ICA model $\mathbf{x} = \mathbf{As} + \boldsymbol{\eta}$, where $cov(\boldsymbol{\eta}) = p\mathbf{I}$, matrix $\mathbf{I}$ is identity, and the parameter p ($p > 0$) controls the noise power (Voss et al., 2015). Each noisy source has 5000 samples to avoid the influence of sample size, and is normalized to have zero mean and unit variance. We can see that EcoICA is much less sensitive to noise compared with FastICA$_{skew}$ and always outperforms others. This superiority is more obvious with small sample size.

Effect of outlier: We simulate outliers by randomly choosing up to 25 source samples to corrupt, where the total sample size is 5000. This is done by adding Gaussian noise with very large power ($p = 10$) to the selected sampling points. Figure 4(c) illustrates the degrading performance of ICA methods with more outliers. We can see that skewness-based methods are significantly less sensitive to outliers than kurtosis-based ones, where EcoICA performs slightly better than FastICA$_{skew}$.

(*a*) Effect of sample size.   (*b*) Effect of noise.   (*c*) Effect of outlier.
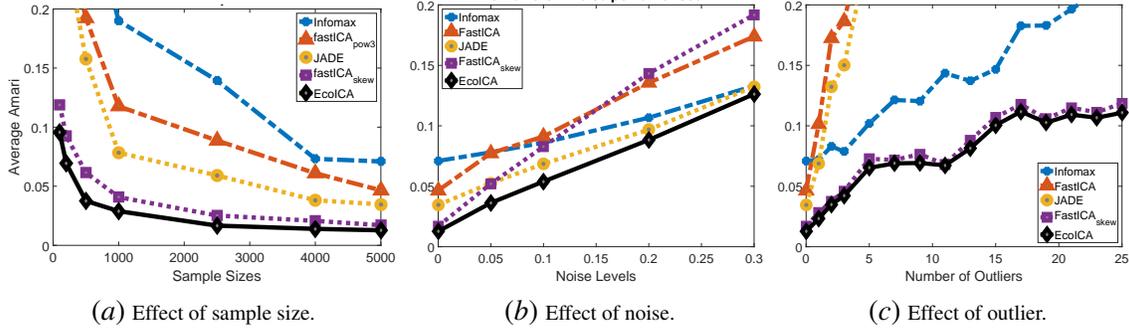
Figure 4: Sensitivity studies for two-source BSS with skewed and kurtosic sources. Each point is the average Amari across 100 repetition. We show errors below 0.2 for more details.
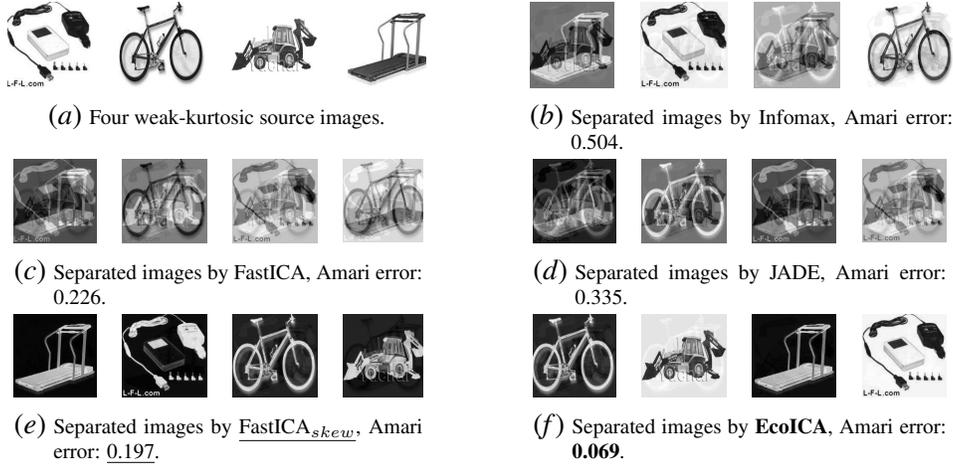

(*a*) Four weak-kurtosic source images.

(*b*) Separated images by Infomax, Amari error: 0.504.

(*c*) Separated images by FastICA, Amari error: 0.226.

(*d*) Separated images by JADE, Amari error: 0.335.

(*e*) Separated images by $\underline{\text{FastICA}_{skew}}$, Amari error: $\underline{0.197}$.

(*f*) Separated images by **EcoICA**, Amari error: **0.069**.

Figure 5: Images separated from mixtures in Fig. 1(*b*) by Infomax, FastICA, JADE, FastICA$_{skew}$, and EcoICA. Only skewness-based methods succeed, and EcoICA yields the best Amari error.

4.2. Blind Image Separation

First, we separate the mixed images in Fig. 1(*b*). Figure 5 shows the separated images by Infomax, FastICA, JADE, FastICA$_{skew}$, and EcoICA. We can see that only skewness-based ICA methods, i.e. EcoICA and FastICA$_{skew}$, can separate the mixed images successfully, for which EcoICA achieves the smallest Amari error.

Four scenarios: For more extensive studies, we investigate the following four scenarios of image separation that differ in source skewness and kurtosis: [s1] skewed and kurtosic images, [s2] skewed but weak-kurtosic images, [s3] weak-skewed but kurtosic images, and [s4] randomly selected images without considering skewness and kurtosis. We do not report the sources that are both weak-skewed and weak-kurtosic, where all investigated ICA methods perform badly.

Image data: The source images are from the Caltech256 repository (Griffin et al., 2006). The images whose skewness (kurtosis) are among top 20% are considered as ***skewed*** (***kurtosic***); the images whose skewness (kurtosis) are among bottom 20% are ***weak-skewed*** (***weak-kurtosic***). We

Table 3: The mean±std Amari errors (multiplied by 100) of 100 repetitions on four image separation. Each four images are mixed and then estimated by 100 mixing matrices. The best (second-best) Amari errors are highlighted in bold (underline).

Amari	Infomax	FastICA	JADE	FastICA$_{skew}$	EcoICA
[s1] $\Uparrow$s $\Uparrow$k	22.95±10.20	23.89±9.16	25.18±8.19	**21.99±7.22**	<u>22.68±7.58</u>
[s2] $\Uparrow$s $\Downarrow$k	31.81±13.03	39.19±16.28	44.70±13.97	<u>25.08±5.46</u>	**24.96±6.30**
[s3] $\Downarrow$s $\Uparrow$k	32.53±13.03	**24.23±6.21**	<u>30.37±10.50</u>	45.86±18.13	44.36±14.77
[s4] $\forall$s $\forall$k	31.20±14.11	25.00±10.47	26.91±10.39	**24.48±10.79**	<u>24.89±7.31</u>

classify the Caltech256 images into three categories, each of which corresponds to one of scenarios [s1]~[s3]. For scenario [s4], we randomly select images from the entire Caltech256.

Experimental settings: For each scenario from [s1]~[s4], we randomly choose four source images, mix them with 100 different mixing matrices, and separate the mixtures via ICA methods. The average Amari errors for these four source images are computed accordingly. We repeat the above calculation with 100 groups of randomly selected source images (i.e. 100×100 mixtures in total), and report the average Amari errors in Table 3.

Image separation results: Table 3 shows that for skewed source images ([s1] and [s2]) and randomly selected source images ([s4]), FastICA$_{skew}$ and EcoICA achieve the best or second-best results. However, FastICA$_{skew}$ is a fixed-point method that is sensitive to *starting points* (Ollila, 2010) so we often require *restarts* to estimate all the four source images; in contrast, EcoICA does not suffer from this problem and is more stable. Moreover, neither FastICA$_{skew}$ nor EcoICA work for weak-skewed source images ([s3]). For kurtosic source images ([s1] and [s3]), FastICA gives good results, while the performance of JADE is much inferior to FastICA for [s3]. Infomax does not explicitly make use of kurtosis or skewness so it has the smallest performance variance over source kurtosis or skewness.

Source statistics prediction from mixtures: Next, we study the prediction of source statistics from the observed data (the mixtures) when we do not know whether the sources are more kurtosic or skewed in practice. We examine whether the dominance of kurtosis or skewness in the mixtures can predict that in the sources accurately. For all the images in [s2] and [s3] above, we compute the mean kurtosis and the mean skewness across all the four mixtures in each BSS experiment. We then examine their magnitudes (absolute values). If the magnitude of the mean skewness (kurtosis) is greater than that of the mean kurtosis (skewness), we predict that the sources are more skewed (kurtosic) and skewness-based (kurtosis-based) ICA is more appropriate and promising. In this prediction task, we achieve 96.9% accuracy for mixtures in [s2] and 98.8% accuracy for mixtures in [s3]. This indicates that we can indeed choose skewness/kurtosis-based methods via comparing the skewness and kurtosis of the mixtures.

4.3. Face Recognition

ICA methods can be used to extract features for face recognition (FR) (Bartlett et al., 2002; Fernandes and Bala, 2013; Lu, 2013). In this subsection, we evaluate the usefulness of EcoICA on FR, where the ground-truth sources (and their kurtosis and skewness) are unknown. FR is not a typical BSS problem, and the performance measure is classification accuracy rather than Amari error.

Data: We use a subset of medium difficulty from the CMU PIE database (Sim et al., 2003), with five frontal poses (C05, C07, C09, C27, C29) under 14 illumination conditions (05 to 14 and 18 to 21). This subset has 4,754 face images from 68 subjects. All the face images were manually cropped, aligned and normalized to 32×32 pixels, with 256 gray levels per pixel. Face recognition performance is tested under varying numbers of training samples per subject, denoted by L.

Algorithms and settings: For all studied ICA methods, we keep $Q\%$ ($Q = 85, 90$) energy in PCA preprocessing. Then, we extract ICs as features and sort them in descending *class discriminability* (Duda et al., 2000). We test various numbers of features and report the best results. For classification, we use the simple nearest neighbour classifier with Euclidean distance. We test the two ICA architectures in (Bartlett et al., 2002). For each subject, $L(= 2, 4, 6, 8, 10, 20)$ samples are randomly selected as training data, and the rest are used for testing. We study the average performance of each Q over 10 repetitions.

Face recognition results: All studied ICA methods achieve very similar performance in recognizing faces. For instance, with Architecture I (II): the average recognition rate gain of EcoICA over FastICA across all values of L is 0.0022 ± 0.0060 (0.0050 ± 0.0046) for $Q = 85$ and 0.0003 ± 0.0043 (-0.0005 ± 0.0035) for $Q = 90$; the average recognition rate gain of EcoICA over JADE across all values of L is 0.0017 ± 0.0058 (-0.0002 ± 0.0050) for $Q = 85$ and -0.0036 ± 0.0060 (-0.0003 ± 0.0033) for $Q = 90$. These results indicate that the EcoICA features, though capturing different statistics of face images, can yield useful information for recognition. Further studies on other type of data (e.g. fMRI) may give more interesting results.

Computing the statistics of the face data shows that they are more kurtosic so kurtosis-based ICA has advantages over skewness-based ICA. Nevertheless, EcoICA obtains similar performance.

5. Conclusion

Kurtosis-based ICA methods have difficulty in recovering weak-kurtosic sources. In this paper, we proposed a new skewness-based ICA in a systematic way to deal with weak-kurtosic but skewed sources. We first designed a new *cumulant operator* $\mathscr{T}$ based on the *third-order cumulant tensor* $\mathcal{Q}_{\mathbf{x}}$. We further defined its *eigenvalues* and *eigenvectors*, and revealed their connections with the columns of the whitened mixing matrix. Next, we proved that we can compute $\mathscr{T}$'s eigenvectors via joint diagonalization of the images of this operator on a set of basis vectors. Based on this proof, we constructed the objective function and solved it with the Jacobi method. Experimental results on both synthetic BSS and blind image separation confirmed EcoICA's overall superiority on skewed sources over existing kurtosis-based and skewness-based ICA methods. In particular, EcoICA showed less sensitivity to sample size, noise, and outlier over FastICA$_{skew}$. Studies on face recognition demonstrated competitive performance of EcoICA features in classification. We hope these positive outcomes could encourage further studies into skewness-based ICA and higher-order cumulant operators.

Acknowledgments

References

S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Neural Information Processing Systems (NIPS)*, pages 757–763, 1996.

M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE Trans. on Neural Networks*, 13:1450–1464, 2002.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

T. Blaschke and L. Wiskott. CuBICA: Independent component analysis by simultaneous third- and fourth-order cumulant diagonalization. *IEEE Trans. on Signal Processing*, 52:1250–1256, 2004.

J. F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11: 157–192, 1999.

J. F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. *IEE Processings F (Radar and Signal Processing)*, 140(6):362–370, 1993.

S. Choi, R. Liu, and A. Cichocki. A spusious equilibria-free learning algorithm for the blind separation of non-zero skewness signals. *Neural Processing Letters*, 7:61–68, 1998.

B. D. Clarkson. A least squares version of algorithm as 211: The F-G diagonalization algorithm. *Applied Statistics*, 37:317–321, 1988.

P. Comon. Tensor diagonalization, a useful tool in signal processing. *IFAC Symposium on System Identification*, 1:77–82, 1994a.

P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36:287–314, 1994b.

G. Deco and D. Obradovic. *An Information-Theoretic Approach to Neural Computing (1st Edition)*. Springer Series in Perspectives in Neural Computing, 1996. ISBN 0387946667.

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000. ISBN 978-0-471-05669-0.

S.L. Fernandes and G.J. Bala. A comparative study on ICA and LPP based face recognition under varying illuminations and facial expressions. In *Int. Conf. on Signal Processing Image Processing Pattern Recognition*, pages 122–126, 2013.

X. Geng, L. Ji, and K. Sun. Principal skewness analysis: Algorithm and its application for multi-spectral/hyperspectral images indexing. *IEEE Geoscience and Remote Sensing Letters*, 11(10): 1821–1825, 2014.

G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. http://www.vision.caltech.edu/Image_Datasets/Caltech256, 2006.

A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 10:626–634, 1999.

A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *IEEE Trans. on Neural Networks*, 13(4-5):411–430, 2000.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley-Interscience, 2001. ISBN 0-471-22131-7.

L. Lathauwer, B. Moor, and Vandewalle J. Independent component analysis and (simultaneous) third-order tensor diagonalization. *IEEE Trans. on Signal Processing*, 49:2262–2271, 2001.

H. Lu. Learning modewise independent components from tensor data using multilinear mixing model. In *European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pages 288–303, 2013.

E. Moreau. Joint-diagonalization of cumulant tensors and source separation. In *IEEE Workshop on Statistical Signal and Array Processing*, pages 339–343, 2000.

E. Moreau. A generalization of joint-diagonalization criteria for source separation. *IEEE Trans. on Signal Processing*, 49:530–541, 2001.

E. Moreau. Joint diagonalization of third order complex symmetric tensors and application to blind separation of non-circular sources. In *Asilomar Conf. on Signals, Systems and Computers*, pages 417–421, 2007.

C. L. Nikias and J. M. Mendel. Signal processing with higher-order spectra. *IEEE Signal Processing Magazine*, 10:10–37, 1993.

E. Ollila. The deflation-based FastICA estimator: Statistical analysis revisited. *IEEE Trans. on Signal Processing*, 58:1527–1541, 2010.

T. Sim, S. Baker, and M. Bsat. The cmu pose illumination, and expression database. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25:1615–1618, 2003.

J. V. Stone, J. Porrill, N. R. Porter, and I. D. Wilkinson. Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions. *NeuroImage*, 15: 407–421, 2002.

J. Voss, M. Belkin, and L. Rademacher. A pseudo-euclidean iteration for optimal recovery in noisy ICA. In *Neural Information Processing Systems (NIPS)*, pages 2872–2880, 2015.

B. Wang and W. Lu. An in-depth comparasion on FastICA, CuBICA and IC-FastICA. In *Advances in Natural Computation, LNCS*, volume 3611, pages 410–414. 2005.