INDEPENDENT COMPONENT ANALYSIS WITH SEVERAL MIXING MATRICES

Wakako Hashimoto

Lab. for Advanced Brain Signal Processing, Brain Science Institute, RIKEN
2-1 Hirosawa, Wako, Saitama 351-0198, Japan

ABSTRACT

A new algorithm is proposed for the variation of independent component analysis (ICA) in which there are several mixing matrices and, for each set of independent components, one of the matrices is randomly chosen to mix the components. The algorithm utilizes high-order moments and can obtain consistent estimators even if the true probability density function of independent components is not obtained. The effectiveness of our algorithm is verified by a numerical experiment. This method can be used to analyze a class of data generated overcompletely, and to classify data in an unsupervised manner.

1. INTRODUCTION

The problem treated in the well-known original independent component analysis (ICA) is simple. Recently, many extensions of ICA have been studied to relax the assumptions of ICA, and to analyze more data, such as assuming a form of dependence and considering the effect of additive noise on data. By these extensions, we can treat data against which the original ICA is not effective. Here we treat the problem in which the mixing matrix is not a single matrix. Instead, there are several mixing matrices, one of which is randomly chosen for each set of independent components to mix the components. This problem is a special case of the ICA mixture model[1, 2, 3].

The algorithm of Lee et al.[3] for the ICA mixture model takes the form of a maximum likelihood estimation. However, unlike the original ICA, this algorithm cannot obtain consistent estimators of mixing matrices for ICA with several mixing matrices when we do not know the probability density function of independent components. Furthermore, estimation of mixing matrices does not lead directly to the reconstruction of independent components. We must determine which matrix was used to mix the components among the candidates for the mixing matrix, in order to reconstruct the independent components.

In this paper, we propose a new algorithm for ICA with several mixing matrices. This algorithm utilizes high-order moments and can obtain consistent estimators even if we do not know the distribution of independent components. We propose a method to reconstruct independent components by estimating their distribution and by calculating the conditional probability of each mixing matrix given data. We verify the effectiveness of our algorithm by a numerical experiment.

2. INDEPENDENT COMPONENT ANALYSIS WITH SEVERAL MIXING MATRICES

2.1. Model for generating data

We formulate the data-generating process and the goals of ICA with several mixing matrices. As with the original ICA, independent components are mutually independent and we assume that we know they are independent. The probability density function of each independent component, $\kappa_i^*(s_i)$, is not known.

$$\boldsymbol{s}(t) \sim \kappa^*(\boldsymbol{s}) = \kappa_1^*(s_1)\kappa_2^*(s_2)\cdots\kappa_n^*(s_n). \tag{1}$$

For simplicity, we assume that the mean of s_i is 0. There are L different candidates for the mixing matrix, $A_1, A_2, \cdots, A_\lambda, \cdots, A_L$. For each set of independent components, that is, for every t, one of the candidate matrices is randomly chosen and it mixes the components. In other words, there is a sequence of indices $\lambda(t)$ made of $\{1, 2, \cdots, L\}$ randomly, and a sequence of data $\boldsymbol{x}(t)$ is obtained by a sequence of independent components $\boldsymbol{s}(t)$ as follows.

$$\boldsymbol{x}(t) = A_{\lambda(t)}\boldsymbol{s}(t). \tag{2}$$

The probability that A_λ is chosen is q_λ, and these values are obtained. Even if we cannot obtain values of q_λ, in some cases, we can appropriately assume that they are the same for all λ. Only mixed data $\boldsymbol{x}$ are observed and we cannot observe which mixing matrix is used. In this paper, we deal with tractable cases, and assume that mixing matrices A_λ are all square and regular. Their inverse matrices are $W_1, W_2, \cdots, W_L$.

In summary, we assume that we know values of q_λ, the number of independent components n, and the number of candidates for the mixing matrix L. Utilizing observed data series $\boldsymbol{x}(t)$, the first goal is to estimate all candidates for the mixing matrix $A_1, A_2, \cdots, A_L$, and the second goal is to separate independent components from data. In the ICA mixture model, the probability density function of independent components and their means are assumed to be different, linked to the used mixing matrix A_λ. In the abovementioned model, they are assumed to be the same for all mixing matrices and the model is the special case of the ICA mixture model. Our algorithm can be modified to a model with different means, but this is not described in this paper.

2.2. Applications of ICA with several mixing matrices

As an example of the ICA mixture model, Lee et al.[1, 3] separated mixtures of a conversation of two people and music in the background. In this example, ICA with several mixing matrices is effective for data satisfying the following two conditions: (1) The data are generated by overcomplete basis functions, that is, the number of independent signal sources is larger than the dimension of data. (2) The number of signal sources generating a signal simultaneously is limited.

Lee et al.[1, 2, 3] also used the ICA mixture model for data classification. In this context, independent components are latent variables, and data are assumed to reflect them linearly.

We introduce another application. We consider linear mixtures of independent components, but the mixing matrix is not fixed, nor is it chosen from several candidates. Instead, it fluctuates slightly stochastically. In some cases, by applying an algorithm for ICA with several mixing matrices, we can decrease the reconstruction error of independent components compared with the original ICA. A numerical experiment of this application will be reported elsewhere.

3. MAXIMUM LIKELIHOOD ESTIMATION

An algorithm taking the form of a maximum likelihood estimation [3] is summarized and discussed.

We define $W = \{W_1, W_2, \cdots, W_L\}$ and $\boldsymbol{y}_\lambda = W_\lambda \boldsymbol{x}$. If we know the true probability density function of independent components $\kappa^*(\boldsymbol{s})$, W_λ can be estimated by maximum likelihood estimation. However, in most actual situations, we do not know the true distribution except that it is factorable. In the original ICA, we can obtain estimators that converge to true values asymptotically, by a maximum likelihood estimation with an arbitrarily assumed distribution $\kappa(\boldsymbol{s})$ [4, 5]. For ICA with several mixing matrices, the probability that the chosen mixing matrix is A_λ and the obtained data are $\boldsymbol{x}$ is

$$P(\boldsymbol{x}, \lambda | W) = q_\lambda \kappa(W_\lambda \boldsymbol{x}) |\det W_\lambda|. \tag{3}$$

From this probability, the likelihood of W can be calculated:

$$P(\boldsymbol{x}|W) = \sum_\lambda q_\lambda \kappa(W_\lambda \boldsymbol{x}) |\det W_\lambda|, \tag{4}$$

$$P(\lambda|\boldsymbol{x}, W) = \frac{q_\lambda \kappa(W_\lambda \boldsymbol{x}) |\det W_\lambda|}{P(\boldsymbol{x}|W)}, \tag{5}$$

$$\frac{\mathrm{d}}{\mathrm{d}W_\lambda} \log P(\boldsymbol{x}|W) = (-\varphi(\boldsymbol{y}_\lambda)\boldsymbol{x}^T + A_\lambda^T) P(\lambda|\boldsymbol{x}, W), \tag{6}$$

where

$$\varphi(\boldsymbol{y}) = -\left(\frac{\kappa_1'(y_1)}{\kappa_1(y_1)}, \frac{\kappa_2'(y_2)}{\kappa_2(y_2)}, \cdots, \frac{\kappa_n'(y_n)}{\kappa_n(y_n)}\right)^T. \tag{7}$$

κ^* is the true distribution function of $\boldsymbol{s}$, and κ is the distribution that is assumed since we do not know the true distribution function. W^* is the true separating matrix, and W is an estimate of W^*. If the number of data is sufficiently large, the estimate of W_λ obtained by the maximum likelihood estimation is, approximately, the solution of the following equation:

$$\begin{aligned} &\mathrm{E}_{\kappa^*, W^*}\left[(-\varphi(\boldsymbol{y}_\lambda)\boldsymbol{x}^T + A_\lambda^T) P(\lambda|\boldsymbol{x}, W)\right] \\ &= \int (-\varphi(\boldsymbol{y}_\lambda)\boldsymbol{y}_\lambda^T + I) A_\lambda^T Q(\boldsymbol{y}_\lambda) \mathrm{d}\boldsymbol{y}_\lambda = 0, \end{aligned} \tag{8}$$

$$Q(\boldsymbol{y}_\lambda) = q_\lambda \kappa(\boldsymbol{y}_\lambda) \frac{\sum_{\lambda'} q_{\lambda'} \kappa^*(W_{\lambda'}^* A_\lambda \boldsymbol{y}_\lambda) |\det W_{\lambda'}^*|}{\sum_{\lambda'} q_{\lambda'} \kappa(W_{\lambda'} A_\lambda \boldsymbol{y}_\lambda) |\det W_{\lambda'}|}. \tag{9}$$

If the assumed $\kappa(\boldsymbol{s})$ is the same as $\kappa^*(\boldsymbol{s})$, $W = W^*$ is one of the solutions of eq. (8). From these two equations, we obtain $Q(\boldsymbol{y}_\lambda) = q_\lambda \kappa(\boldsymbol{y}_\lambda)$, where $Q(\boldsymbol{y}_\lambda)$ is factorable like eq. (1) and eq. (8) holds. However, if $\kappa(\boldsymbol{s})$ is different from $\kappa^*(\boldsymbol{s})$, under the condition $W = W^*$, $Q(\boldsymbol{y}_\lambda)$ is not factorable and eq. (8) does not hold. Therefore, contrary to the original ICA, maximum likelihood estimation cannot obtain a consistent estimator for this problem.

4. AN ALGORITHM UTILIZING HIGH-ORDER MOMENTS OF INDEPENDENT COMPONENTS

4.1. Calculation of high-order moments and estimation of mixing matrices

We propose a new algorithm for ICA with several mixing matrices. For the original ICA, Cardoso (1999)[6] proposed a method utilizing high-order cumulants.

The relation between high-order moments of data $\boldsymbol{x}$ and independent components $\boldsymbol{s}$ is represented in the following equation with the true mixing matrix A^*_λ.

$$
\begin{aligned}
&\mathrm{E}[x_{i_1} x_{i_2} \cdots x_{i_m}] \\
&= \sum_\lambda q_\lambda \mathrm{E}[x_{i_1} x_{i_2} \cdots x_{i_m} | \text{the mixing matrix is} A^*_\lambda] \\
&= \sum_\lambda q_\lambda \mathrm{E}\left[\left(\sum_{p_1} a^*_{\lambda, i_1 p_1} s_{p_1} \right) \left(\sum_{p_2} a^*_{\lambda, i_2 p_2} s_{p_2} \right) \cdots \right. \\
&\qquad \left. \times \left(\sum_{p_m} a^*_{\lambda, i_m p_m} s_{p_m} \right) \right] \\
&= \sum_\lambda q_\lambda \sum_{p_1} \sum_{p_2} \cdots \sum_{p_m} a^*_{\lambda, i_1 p_1} a^*_{\lambda, i_2 p_2} \cdots a^*_{\lambda, i_m p_m} \\
&\qquad \times \mathrm{E}\left[s_{p_1} s_{p_2} \cdots s_{p_m} \right]. \qquad (10)
\end{aligned}
$$

We define $\tilde{\boldsymbol{x}}$ as the vector that is composed of m-th-order terms of x_i, $x_{i_1} x_{i_2} \cdots x_{i_m}$, arranged in lexicographic order of $I = (i_1, i_2, \cdots, i_m)$. The dimension of $\tilde{\boldsymbol{x}}$ is n^m. In other words, if the following relation holds,

$$
\begin{aligned}
j = &n^{m-1}(i_1 - 1) + n^{m-2}(i_2 - 1) + \cdots \\
&+ n(i_{m-1} - 1) + i_m, \qquad (11)
\end{aligned}
$$

the j-th element of $\tilde{\boldsymbol{x}}$ is $x_{i_1} x_{i_2} \cdots x_{i_m}$. We define $\tilde{\boldsymbol{s}}$ from $\boldsymbol{s}$ in the same way. We define a $n^m \times n^m$ matrix B as follows:

$$
\begin{aligned}
b_{P,I} &= b_{p_1 p_2 \cdots p_m, i_1 i_2 \cdots i_m} \\
&= \sum_\lambda q_\lambda a_{\lambda, i_1 p_1} a_{\lambda, i_2 p_2} \cdots a_{\lambda, i_m p_m}. \qquad (12)
\end{aligned}
$$

The row index of B is the lexicographic order of $P = (p_1, p_2, \cdots, p_m)$, and the column index is the lexicographic order of $I = (i_1, i_2, \cdots, i_m)$. Equation (10) is rewritten as follows, with B calculated from A^*_λ,

$$
\mathrm{E}[\tilde{\boldsymbol{x}}]^T = \mathrm{E}[\tilde{\boldsymbol{s}}]^T B. \qquad (13)
$$

Here, we add another assumption to A_λ, that B is regular. We define G as the inverse matrix of B. From eq. (13), we obtain the following equation:

$$
\mathrm{E}[\tilde{\boldsymbol{s}}]^T = \mathrm{E}[\tilde{\boldsymbol{x}}]^T B^{-1} = \mathrm{E}[\tilde{\boldsymbol{x}}]^T G. \qquad (14)
$$

Using this equation, if G is calculated from true mixing matrices A^*_λ, we can estimate high-order moments of independent components from high-order moments of data.

Let us consider an estimate A_λ of the mixing matrix. The column vector of G is represented by $\boldsymbol{g}_{i_1 i_2 \cdots i_m}$ and we define functions f_I by the following equation.

$$
f_{i_1 i_2 \cdots i_m}(\boldsymbol{x}) = \tilde{\boldsymbol{x}}^T \boldsymbol{g}_{i_1 i_2 \cdots i_m}. \qquad (15)
$$

Note that if G is calculated from true mixing matrices,

$$
\mathrm{E}[f_{i_1 i_2 \cdots i_m}(\boldsymbol{x})] = \mathrm{E}[s_{i_1} s_{i_2} \cdots s_{i_m}]. \qquad (16)
$$

We define $\mathrm{E}_t[f]$ as a sample mean of $f(\boldsymbol{x}(t))$.

$$
\mathrm{E}_t[f] = \frac{1}{T} \sum_{t=1}^{T} f(\boldsymbol{x}(t)). \qquad (17)
$$

T is the number of data sets. We can obtain an estimate of A_λ by bringing values of $\mathrm{E}_t[f_{i_1 i_2 \cdots i_m}(\boldsymbol{x})]$ close to the values of $\mathrm{E}[s_{i_1} s_{i_2} \cdots s_{i_m}]$, which have been decided by assumptions of independence and scale, as described below. If a high-order moment of $\boldsymbol{s}$ contains one degree variable, that is, there is at least one number that appears only once in $I = (i_1, i_2 \cdots i_m)$, the expectation of the moment is 0, based on independence and the assumption $\mathrm{E}[s_i] = 0$. We assume the scale of independent components as $\mathrm{E}[s_i^2] = 1.0$ since their scale cannot be decided from data $\boldsymbol{x}$. Values of some other moments can be decided from this assumption. From these restrictions, we can construct a cost function. For example, using moments up to the 4th order, we obtain the following cost function:

$$
\begin{aligned}
E = &\frac{1}{2} \left\{ \sum_i (\mathrm{E}_t[f_{ii}] - 1.0)^2 + \sum_{i \neq j} \mathrm{E}_t[f_{ij}]^2 \right\} \\
&+ \frac{1}{6} \left\{ \sum_{i \neq j} \mathrm{E}_t[f_{iij}]^2 + \sum_{i \neq j, j \neq k, k \neq i} \mathrm{E}_t[f_{ijk}]^2 \right\} \\
&+ \frac{1}{24} \left\{ \sum \mathrm{E}_t[f_{ijjj}]^2 + \sum \mathrm{E}_t[f_{ijkk}]^2 + \sum \mathrm{E}_t[f_{ijkl}]^2 \right. \\
&\qquad \left. + \sum (\mathrm{E}_t[f_{iijj}] - 1.0)^2 \right\} \qquad (18)
\end{aligned}
$$

Updating the values of A_λ to decrease the cost function, we can obtain an estimate of A_λ. The true values of A_λ minimize this cost function regardless of the true distribution of independent components.

The residual problem is which set of moments is sufficient for consistent estimation. At least, we need $n^2 L$ moments, that is, as many as the unknown variables, and in practice, we need more for the stability of calculations. Further analysis of sufficient conditions will be treated in a forthcoming paper.

4.2. Reconstruction of independent components

In ICA with several mixing matrices, the estimation of mixing matrices is not sufficient for the reconstruction of independent components. For each set of data, one must determine which matrix was used for mixing. If the true probability density function of independent components $\kappa^*(\boldsymbol{s})$ were obtained, the conditional probability of each mixing matrix given a set of data $P(\lambda | \boldsymbol{x}, W)$ would be calculated. It is an appropriate way to employ, as an answer, the matrix with the largest posterior probability.

Since we can consistently estimate mixing matrices, we can consistently estimate all auto moments of independent components by $\mathrm{E}_t[f_I(\boldsymbol{x})]$. One way to estimate the probability density function $\kappa^*(\boldsymbol{s})$ is to calculate estimates of m-th-order auto cumulants of i-th component $\mu_{m,i}$, from auto moments, and substitute these estimates into the Gram-Charlier expansion:

$$\kappa(s_i) \sim \phi(s_i)\left(1+\frac{\mu_{3,i}}{3!}H_3(s_i)+\frac{\mu_{4,i}}{4!}H_4(s_i) + \frac{\mu_{5,i}}{5!}H_5(s_i)+\frac{\mu_{6,i}+10\mu_{3,i}^2}{6!}H_6(s_i)\ldots\right). \quad (19)$$

$H_m(s)$ is the m-th Hermite polynomial, and $\phi(s)$ is a density function of the standard normal distribution. From the estimated distribution, we can estimate $P(\lambda|\boldsymbol{x}, W)$, and can employ λ with the largest posterior probability. Even if we know the true distribution of independent components and the values of the true mixing matrices, independent components are not always reconstructed exactly. The accuracy of reconstruction depends on the problem itself, for example, the combination of mixing matrices.

In many actual problems, independent components are smooth or the same mixing matrix tends to be used successively. These assumptions can be utilized to refine the reconstruction.

In summary, the algorithm we propose to reconstruct independent components is as follows.

Step 1 Estimate mixing matrices by means of the algorithm utilizing high-order moments, as described in section 4.1.

Step 2 Estimate auto moments of independent components $\mathrm{E}[s_i^3]$, $\mathrm{E}[s_i^4]$... by $\mathrm{E}_t[f_{iii}]$, $\mathrm{E}_t[f_{iiii}]$ .... From these estimated moments, calculate auto cumulants $\mu_{3,i}$, $\mu_{4,i}$....

Step 3 Estimate $\kappa^*(\boldsymbol{s})$ by substituting estimated auto cumulants into the Gram-Charlier expansion.

Step 4 From the estimated probability density function $\kappa(\boldsymbol{s})$, calculate conditional probabilities of mixing matrices, given each set of data.

$$P(\lambda|\boldsymbol{x}, A) = \frac{q_\lambda \kappa(W_\lambda \boldsymbol{x})|\det W_\lambda|}{\sum_{\lambda'} q_{\lambda'}\kappa(W_{\lambda'}\boldsymbol{x})|\det W_{\lambda'}|}.$$

Step 5 Choose a mixing matrix with the maximum conditional probability for each set of data, and by means of the matrix, separate independent components from data.

$$\lambda(t) = \arg\max_\lambda P(\lambda|\boldsymbol{x}(t), A), \qquad \boldsymbol{y}(t) = A_{\lambda(t)}^{-1}\boldsymbol{x}(t).$$

Step 6 If some prior knowledge is available for the data, for example, the smoothness of independent components, or the continuity of the used mixing matrix, utilize them to refine choices of the mixing matrix. We describe a simple example of these methods for image data in section 5.

5. SIMULATION

To verify the effectiveness of our algorithm, we carried out a simple demonstration using image data. We used 3 images: a building, fabric and tiles as independent components, shown in the top row of Figure 1, taken from MIT's VisTex database. [1] We simulated data from these 3 images according to the model described in eq. (2) with $L = 2$ and $q_1 = q_2 = 0.5$. Mixing matrices were made randomly. The mixtures of images appear in the second row of Figure 1.

Table 1. Values of auto cumulants of images used as independent components. For each image, true values calculated from original images are in the upper row, and estimates are in the bottom row. s_1:a building, s_2:fabric, s_3:tiles.

order of cumulant		3rd	4th	5th	6th
s_1	true value	0.714	0.093	-1.944	-5.590
s_1	estimate	0.748	0.187	-1.871	-6.468
s_2	true value	-0.245	-0.444	0.874	0.405
s_2	estimate	-0.189	-0.436	0.374	0.614
s_3	true value	-0.375	-1.135	2.114	5.312
s_3	estimate	-0.377	-1.158	2.132	5.557

We used the algorithm described in section 4 with the cost function in eq. (18) to reconstruct independent components. Since we used the gradient method to decrease the cost function, local minima could not be avoided. To obtain good estimates, we made 40 initial values. From the best one of the 40 estimates, we made 48 new initial matrices by permuting columns of A_2, and by changing the signs of columns of A_2. Then, we employed the matrices with the smallest cost function. The estimation was good, and $A_i^{-1}A_i^*$ is close to an identity matrix, as shown in the following equations:

$$A_1^{-1}A_1^* = \begin{pmatrix} 1.009 & -0.010 & 0.019 \\ -0.036 & 1.015 & -0.010 \\ -0.023 & 0.006 & 1.011 \end{pmatrix}, \quad (20)$$

$$A_2^{-1}A_2^* = \begin{pmatrix} 0.987 & -0.034 & 0.063 \\ 0.001 & 0.984 & -0.017 \\ 0.018 & -0.021 & 1.001 \end{pmatrix}. \quad (21)$$

After mixing matrices were estimated, we reconstructed independent components using the algorithm described in section 4.2. The cumulants of original image data and their estimates by $\mathrm{E}_t[f_I]$ appear in Table 1, and the estimation is good. Histograms of original image data appear in Figure 2. Graphs in Figure 2 are four approximations of distributions by Gram-Charlier expansions. One approximation utilizes the expansion up to 6-th order terms, and the used cumu-

Table 2. Error rates of choices of the mixing matrix used for each data set. "True value" means that values of cumulants used to approximate $\kappa^*(\boldsymbol{s})$ were calculated from original images. "Estimate" means that used cumulants values were estimates. "Smoothness" means that after we determined the used mixing matrix by estimated probability density functions, we altered the determination by utilizing the smoothness of images. The orders are those of the Gram-Charlier expansion.

cumulant	error rate(%)
true value 6th	20.8
estimate 6th	20.9
estimate 4th	22.5
estimate 2nd	27.1
estimate 6th + smoothness	6.2

lants were calculated from original image data. Three approximations utilize expansions up to 2nd-order terms, 4th-order terms and 6th-order terms, and the cumulants were estimated by $\mathrm{E}_t[f_I]$. From each approximation, we determined the most plausible mixing matrix for each data set and reconstructed the independent components, according to steps 4 and 5 in section 4.2. Error rates of choices of a mixing matrix are shown in Table 2. Since estimation of cumulants was sufficiently good, we could determine the true mixing matrix as accurately as the decision using true values of cumulants. The error rate decreased considerably when we used higher order terms of the Gram-Charlier expansion. Reconstructions of image data by approximated distributions from estimated cumulants up to the 6th order appear in the third row of Fig. 1.

From the reconstructed images in the third row of Figure 1, we applied a simple method that utilizes smoothness. Suppose we obtain reconstructed images whose (p, q) pixels are $\boldsymbol{y}_{p,q}$. We refine $\boldsymbol{y}_{p,q}$ into $A^{-1}_{\lambda_{\mathrm{new}}}\boldsymbol{x}_{p,q}$ by the following equations:

$$\lambda_{\mathrm{new}} = \arg\max_{\lambda}(\log P(\lambda|\boldsymbol{x}_{p,q}, A) - \sum_{(p',q')\in N} (\boldsymbol{y}_{p',q'} - A^{-1}_{\lambda}\boldsymbol{x}_{p,q})^T C(\boldsymbol{y}_{p',q'} - A^{-1}_{\lambda}\boldsymbol{x}_{p,q})), \quad (22)$$

with the following neighborhood and constants,

$$N = \{(p, q-1), (p-1, q), (p+1, q), (p, q+1)\}, \quad (23)$$
$$C = \mathrm{diag}(c_1, c_2, \cdots, c_n). \quad (24)$$

c_i is defined as $1/2\sigma_i^2$ where σ_i^2 is the mean of squares of differences between adjacent pixels of the reconstruction of the i-th image. Even by this simple method, the error rate decreased to 6.2 %. Reconstructed images are shown in the bottom row of Fig. 1. The reconstruction is satisfactory.

6. CONCLUSION

Our method can retain consistency of estimation even if we do not know the true distribution of independent components. However, it is not clear which set of moments is sufficient for consistent estimation. We can obtain sufficient conditions by formulating the ICA with several mixing matrices as a semiparametric estimation problem and examining the condition under which estimating function[7] can be constructed. We carried out this analysis with 4th-order moments only and the result will be reported in a forthcoming paper. For the original ICA, Amari et al. (1997)[4] treated a similar problem.

We did not compare our algorithm with the maximum likelihood estimation[3], because we anticipated that results depended heavily on whether the presumed distribution of independent components was close to the true one.

What condition decides the maximum number of mixing matrices that can be estimated consistently ? Can we estimate any number of matrices using higher-order moments? Is the maximum number of mixing matrices L related to the dimension of independent components n? These questions remain to be answered in future research.

7. REFERENCES

[1] T-W. Lee, M. S. Lewicki, and T. J. Sejnowski, "ICA mixture models for unsupervised classification and automatic context switching," *International Workshop on Independent Component Analysis and Blind Signal Separation*, pp. 209–214, 1999.

[2] T-W. Lee, M. S. Lewicki, and T. J. Sejnowski, "Unsupervised classification with non-gaussian mixture models using ICA," in *Advances in Neural Information Processing Systems*. 1999, vol. 11, pp. 508–514, MIT Press, Cambridge MA.

[3] T-W. Lee, M. S. Lewicki, and T. J. Sejnowski, "ICA mixture models for unsupervised classification of non-gaussian classes and automatic context switching in blind signal separation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1078–1089, 2000.

[4] S. Amari, T. Chen, and A. Cichocki, "Stability analysis of learning algorithms for blind source separation," *Neural Networks*, vol. 10, pp. 1345–1351, 1997.

[5] D. T. Pham and P. Garat, "Blind separation of mixtures of independent sources through a quasi-maximum likelihood approach," *IEEE Transactions on Signal Processing*, vol. 45, pp. 1712–1725, 1997.

[6] J-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, pp. 157–192, 1999.

[7] V. P. Godambe, "Conditional likelihood and unconditional optimum estimating equations," *Biometrika*, vol. 63, pp. 277–284, 1976.

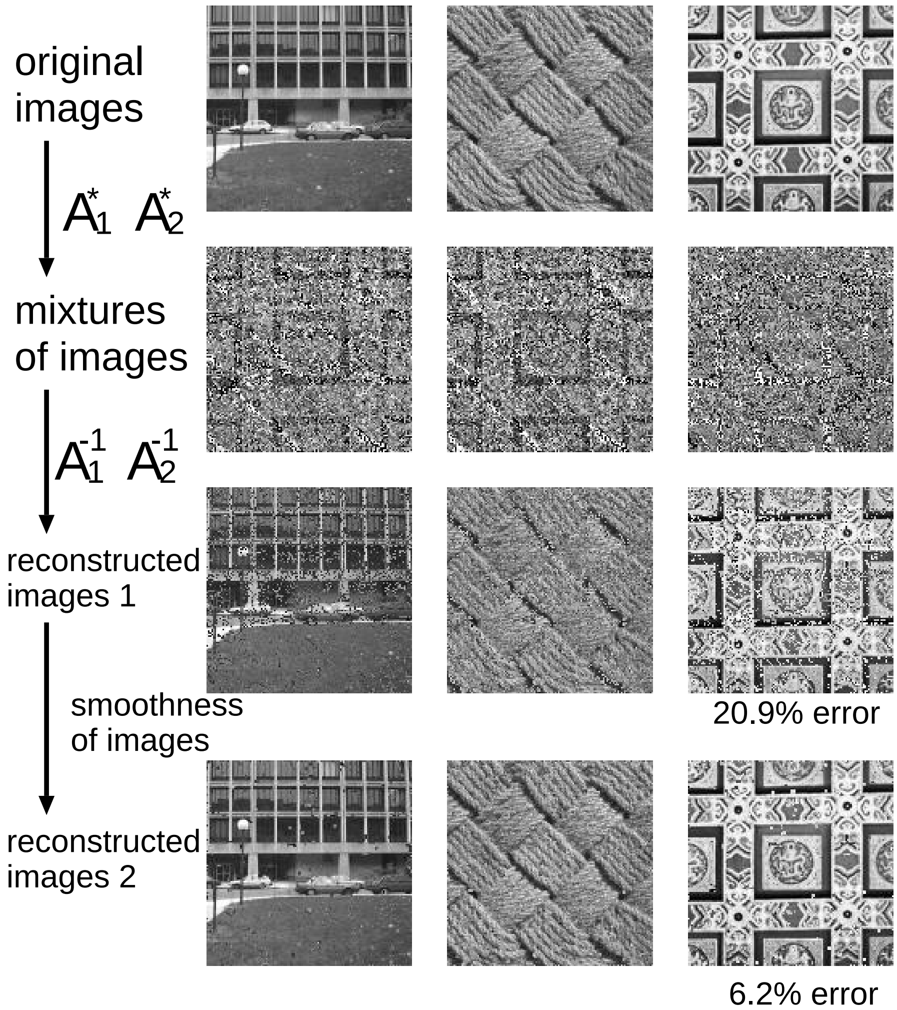

Fig. 1. Results of simulation to verify the effectiveness of the proposed algorithm. Top row: original images, that is, independent components. 2nd row: images mixed by two mixing matrices, that is, data. 3rd row: images reconstructed by approximating the distribution of independent components from estimates of cumulants up to the 6th order. Bottom row: images refined from images in the 3rd row, by utilizing smoothness of images. The error rates are those of the mixing matrices.

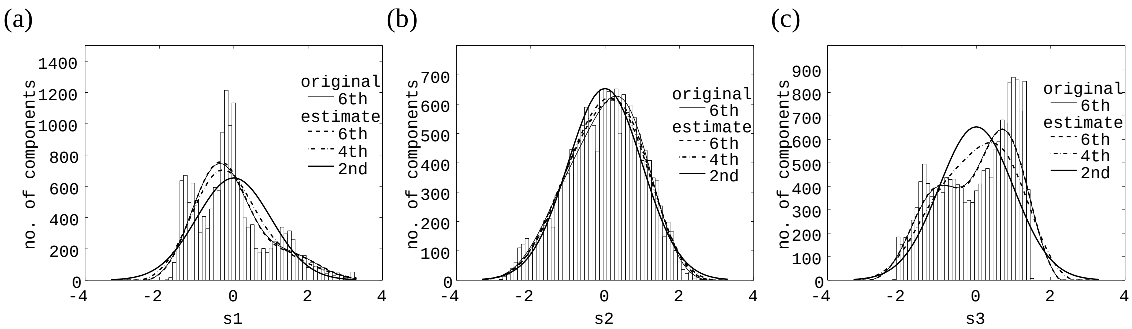

Fig. 2. Histograms represent true distributions of original images. The thin solid line represents an approximation of distribution from cumulants of original image data up to the 6th order. The thick solid line represents an approximation from estimates of cumulants by $\mathrm{E}_t[f_I]$ up to the 2nd order. The dash-dot line represents an approximation from estimates of cumulants up to the 4th order, and the dashed line, up to the 6th order. The three graphs correspond to the three original images. a: a building, b: fabric, c: tiles.