

Density based ICA for non-square mixing matrix

Anonymous Authors¹

Abstract

In its basic form Independent Component Analysis (ICA) aims to find an invertible linear transformation of the data so that the resulting data has independent components. In the case when number of sources is less than the number of sensors, the task has an additional step which consists of filtering out the possible noise components. In such situation we are looking for so-called non-square mixing matrix.

Due to computational constraints, principal component analysis is often used for dimension reduction prior to ICA (PCA+ICA). However, such approach often removes important information. In this paper we present a density based method $\text{ICA}_{\mathcal{SN}}$ based on split-gaussians which is dedicated for determining non-square mixing matrix in ICA maximum likelihood framework.

1. Introduction

Independent component analysis (ICA) is similar in many aspects to principal component analysis (PCA). In PCA we look for an orthonormal base in which the data components are not linearly dependent (uncorrelated), while in ICA we search for the coordinate system in which the components are independent. More precisely the aim of ICA is to transform the observed data \mathbf{X} into maximally independent components \mathbf{S} with use of an invertible linear transformation \mathbf{W} , called the *transformation matrix*:

$$\mathbf{S} = \mathbf{W}^T \mathbf{X}.$$

Popular ICA methodology does not directly attempt to find components that are independent but rather components that are as non-Gaussian as possible. This follows from the fact that one of the theoretical foundations of ICA is given by the dual view at the Central Limit Theorem (Hyvärinen & Oja, 2000), which states that the distribution of the sum (average or linear combination) of N independent random variables approaches Gaussian as $N \rightarrow \infty$. Obviously if all source variables are Gaussian, the ICA method will not work.

Another common approach to ICA based on the maximum likelihood estimation (Pham & Garat, 1997) is recently gaining popularity (Hyvärinen et al., 2004; Samworth et al., 2012; Spurek et al., 2017). Then we search for the optimally fitted to data coordinate system B and marginal densities f_i such that the data density factors in base B as the product of marginal densities. To obtain an efficient method and avoid overfitting we have to restrict the marginal densities f_i to a class \mathcal{F} of densities which has not too many parameters which can be easily estimated (clearly from obvious reasons this class has to be different from gaussians). As \mathcal{F} we typically choose the super-Gaussian logistic density or other heavy tails distributions.

In many applications of ICA we deal with the case when several sensors measure the latent variables and the rest of them record only the noise. This happens when the number of sources is unknown and may be less than the number of sensors (then we are looking for so-called *non-square mixing matrix* \mathbf{W}). Such a case is common for example in the identification of brain networks in functional magnetic resonance imaging (fMRI) (Beckmann, 2012; Green et al., 2002). In practice, most approaches deal with this problem by first applying PCA to the observations prior to classic ICA (PCA+ICA) to meet the assumption of square mixing and to reduce computational costs (Hyvärinen et al., 2004). Although numerically effective, this approach may fail as it is not invariant with respect to linear transformation, since PCA will find a “noise” component if it is sufficiently large.

The aim of this paper is to propose a new density based approach to deal with this case which does not have the above mentioned disadvantage. Our idea is to join the two earlier mentioned approaches to solving ICA - one based on the search for non-gaussian components and the other based on density estimation - to deal with the case when the number of sources is smaller than that of sensors. Observe that the noise typically occurs as a sum of many independent factors, and consequently thanks to the central limit theorem it typically has approximately gaussian distribution. This is why one typically makes the assumption (Cover & Thomas, 2012) that

noise components are coming from a gaussian noise.

Following the density approach to filter them out we fit the

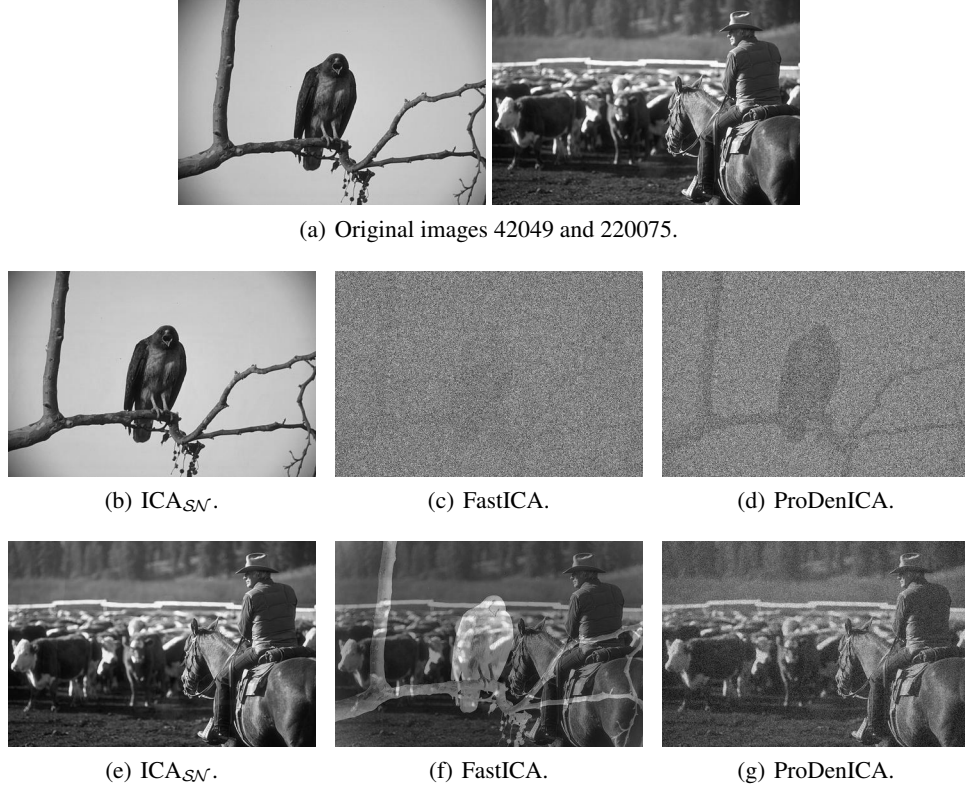


Figure 1. Comparison of images separation by our method (ICA_{SV}), with FastICA and ProDenICA. Before mixing by a linear matrix, we added to the first two components given in (a) the third component given by random normal noise. As we see ICA_{SV} was able to perfectly recover the two first components.

first d -components from a class of \mathcal{F} of densities which is broader than Gaussians, while the rest from the Gaussians \mathcal{N} (the final choice of the value of parameter $d \in \{1, \dots, D\}$ can be decided by applying either AIC or BIC criterion). Following (Spurek et al., 2017) as \mathcal{F} we take the class \mathcal{SN} of split-normal densities.

Our experiments show, which is illustrated by Figure 1, that ICA_{SV} works as desired and effectively removes the components which contain Gaussian noise. However, we cannot objectively conclude that it is better as compared to other state-of-the-art approaches, since the experiment was conducted in the setting optimal to our method as we assumed that the noise was Gaussian.

2. Related works

In literature exist a few approaches dedicated for non-square ICA problem. Most of such methods are dedicated for individual methods. Attias and Schreiner (Attias & Schreiner, 1997) derived a likelihood based algorithm for separation of general sequences with a frequency domain implementation. Belouchrani and Cardoso (Belouchrani et al., 1995) presented a general likelihood approach al-

lowing for additive noise and for non-square mixing matrices. They applied the method to separation of sources taking discrete values and estimated the mixing matrix using an Expectation-Maximization (EM) approach with both a deterministic and a stochastic formulation. In (Moulines et al., 1997) authors used the EM approach for separation of autocorrelated sequences in presence of noise and explored a family of flexible source signal priors based on Gaussian Mixtures.

The assumption is that of square mixing is mostly unrealistic in the case of EEG and FMRI where the number of sources is less than the number of electrodes (Beckmann & Smith, 2004; Samarov et al., 2004; Shi et al., 2017). Therefore, many of algorithms dedicated for this task use a probabilistic ICA (Tipping & Bishop, 1999). The noisy ICA model can be approximated using a variant of PCA+ICA (Beckmann & Smith, 2004), where probabilistic PCA is used to estimate the number of components and achieve dimension reduction (Tipping & Bishop, 1999). In (Allassonniere & Yunes, 2012) authors developed stochastic EM algorithms to estimate the noisy model and proposed parametric methods.

Other methods exploring non-Gaussian structure in multivariate data include non-Gaussian component analysis (NGCA) and projection pursuit (Blanchard et al., 2006; Kawanabe et al., 2007). NGCA is a more general case of linear non-Gaussian component analysis (LNGCA) (Risk et al., 2015) that allows non-linear dependence between the non-Gaussian components.

In the paper (Miettinen et al., 2014) authors propose a novel adaptive twostage deflation-based FastICA algorithm that allows one to use different nonlinearities for different components and optimizes the order in which the components are extracted.

3. Theoretical foundations of ICA_{SN}

In this section we present theoretical foundations of the method we present. We begin with the statement of the problem, next we focus our attention on the presentation of the class of densities we discuss. Last we show the gradient of the method which is needed in the optimization procedure.

3.1. Statement of the problem

Since this general idea of the search for ICA with the use of maximum likelihood is essential in our further considerations, for the convenience of the reader we first describe it briefly. Assume that the random vector \mathbf{X} in \mathbb{R}^D has the density function $F(\mathbf{x})$. Suppose that the components of \mathbf{X} are not independent, but that we know (or suspect) that there is a basis B (we put $W^T = B^{-1}$) such that in that base the components of \mathbf{X} become independent. Observe that where $w_i^T \mathbf{x}$ is the i -th coefficient of \mathbf{x} in the basis B (w_i denotes the i -th column of W), and therefore there exist densities f_1, \dots, f_D such that

$$F(\mathbf{x}) = \det(W) \cdot f_1(w_1^T \mathbf{x}) \cdot \dots \cdot f_D(w_D^T \mathbf{x}). \quad (1)$$

Given W and densities $(f_i)_{i=1}^D$ we introduce notation to represent RHS of the above equation:

$$F_W(f_1, \dots, f_D)(\mathbf{x}) = \det(W) \cdot f_1(w_1^T \mathbf{x}) \cdot \dots \cdot f_D(w_D^T \mathbf{x}).$$

Thus we may state the density based formulation of ICA in the case we have only a sample X from random vector \mathbf{X} .

GENERAL ICA PROBLEM (maximum likelihood formulation).

Find densities f_i and matrix W , so that F given by (1) optimally fits the data $X = (x_i)$ with respect to the likelihood, that is that the value

$$\sum_i \log F_W(f_1, \dots, f_D)(x_i)$$

is maximized.

Since the search over the space of all densities is not feasible, and could lead to overfitting, we naturally have to reduce to a subclass of all densities on \mathbb{R} parametrized by a finite amount of parameters. Clearly, since ICA does not work if the data are gaussian, we have to choose a family \mathcal{F} of densities which is distant from Gaussian ones.

ICA FOR DENSITY CLASS \mathcal{F} .

Find densities matrix W and densities

$$f_1, \dots, f_D \in \mathcal{F},$$

such that the value of

$$\sum_i \log F_W(f_1, \dots, f_D)(x_i)$$

is maximized.

Similarly to (Spurek et al., 2017) as a class \mathcal{F} we are going to take the class of split-gaussians, as as it is easy to deal with (small number of parameters) and is resistant to outliers¹.

As mentioned in the introduction, we assume that components which we would like to filter-out, are coming from a gaussian noise, and the aim it to fit the first d -components from a larger class of densities, while the rest from the gaussians \mathcal{N} . Thus our final problem can be stated as follows.

ICA FOR DENSITY CLASS \mathcal{F} WITH d SOURCES.

Find matrix W , densities

$$f_1, \dots, f_d \in \mathcal{F} \text{ and normal densities } f_{d+1}, \dots, f_D \in \mathcal{N},$$

so that the value of

$$\sum_i \ln F_W(f_1, \dots, f_D)(x_i)$$

is maximized.

Observe that the solution to the above problem is linearly invariant, that is if W is optimal for X and A is linear, then W_A is optimal for AX , where $W_A = (A^{-1})^T W$.

The continuous version of the condition we maximize in the case we know the density f of the random variable \mathbf{X} limits to

$$\begin{aligned} & \int \ln F_W(f_1, \dots, f_D)(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ & = -H(f, F_W(f_1, \dots, f_D)), \end{aligned}$$

where the cross entropy $H(f, g)$ is given by the sum of entropy $H(f)$ and Kullback-Leibler divergence $D_{KL}(f, g)$.

¹The reason is that split gaussians, instead at fitting the distribution with respect to heavy tails, fits the asymmetry of the data.

Thus the continuous version of the ICA problem with d sources reduces to the minimization of

$$D_{KL}(f, F_W(f_1, \dots, f_D))$$

over all matrices W and densities $f_1, \dots, f_D \in \mathcal{F}$. Since for fixed f Kullback-Leibler divergence is minimized for $g = f$, we arrive at the following result, which says that in the ideal case by the discussed approach we restore the unmixing matrix if it exists.

Theorem 3.1. *Let F be a density such that there exist matrix \bar{W} and densities*

$$\hat{f}_1, \dots, \hat{f}_d \in \mathcal{F} \text{ and } \hat{f}_{d+1}, \dots, \hat{f}_D \in \mathcal{N}$$

such that

$$F = F_{\bar{W}}(\hat{f}_1, \dots, \hat{f}_D).$$

Then

$$\begin{aligned} & \bar{W}, \hat{f}_1, \dots, \hat{f}_D \\ &= \operatorname{argmin}_{W, f_1, \dots, f_d \in \mathcal{F}, f_{d+1}, \dots, f_D \in \mathcal{N}} \{F_W(f_1, \dots, f_D)\} \end{aligned}$$

3.2. Split normal distribution

In this section we discuss the class \mathcal{F} we will use in our final algorithm of ICA_{SN}. The density of \mathcal{SN} , the one-dimensional split normal distribution (Villani & Larsson, 2006), is given by the formula

$$\mathcal{SN}(x; m, \sigma^2, \tau^2) = \begin{cases} c \cdot \exp[-\frac{1}{2\sigma^2}(x-m)^2], & x \leq m \\ c \cdot \exp[-\frac{1}{2\tau^2}(x-m)^2], & x > m \end{cases}$$

where $c = \sqrt{\frac{2}{\pi}} \sigma^{-1} (1 + \tau)^{-1}$.

As we see the split normal distribution comes from merging two opposite halves of two normal distributions in their common mode. The main advantage of split normal distributions over normal one is that it allows data asymmetry. In 1982 John (John, 1982) showed that the likelihood function can be expressed in a form in which the scale parameters σ and τ are an explicit function of the location parameter m . In the case when $\mathcal{F} = \mathcal{SN}$ the density class considered in the previous subsection is given in the explicit form by the following observation.

Observation 3.1. *A density of the multivariate split normal d and normal $D - d$ distribution is given by*

$$\begin{aligned} & \mathcal{SN}_d \mathcal{N}_{D-d}(x; m, W, \sigma^2, \tau^2) = \\ & \det(W) \prod_{j=1}^d \mathcal{SN}(w_j^T(x-m); 0, \sigma_j^2, \tau_j^2) \cdot \\ & \cdot \prod_{j=d+1}^D \mathcal{N}(w_j^T(x-m); 0, \sigma_j^2), \end{aligned}$$

where w_j is the j -th column of non-singular matrix W , $m = (m_1, \dots, m_d)^T$, $\sigma = (\sigma_1, \dots, \sigma_d)$ and $\tau = (\tau_1, \dots, \tau_{D-d})$.

Observe that the above density probability function has mode in m . As a consequence of result of John (John, 1982) we can maximize the likelihood of the above function on data X with respect to σ and τ .

Theorem 3.2. *Let x_1, \dots, x_n be given, and let $m \in \mathbb{R}^D$ and matrix W be fixed. Then the likelihood maximized w.r.t. σ and τ is*

$$\begin{aligned} \hat{L}(X; m, W) &= \frac{2^{(d-D/2)n} n^{dn/2}}{(\pi e)^{Dn/2}} \cdot \\ & \cdot \left(\frac{1}{|\det(W)|^{\frac{D}{2}}} \prod_{j=1}^d g_j(m, W) \right)^{-3n/2} \left(\prod_{j=d+1}^D \frac{(s_1 + s_2)}{n} \right)^{-n/2}, \end{aligned} \quad (2)$$

where

$$\begin{aligned} g_j(m, W) &= s_{1j}^{1/3} + s_{2j}^{1/3}, \\ s_{1j} &= \sum_{i \in I_j} [w_j^T(x_i - m)]^2, \quad I_j = \{i: w_j^T(x_i - m) \leq 0\}, \\ s_{2j} &= \sum_{i \in I_j^c} [w_j^T(x_i - m)]^2, \quad I_j^c = \{i: w_j^T(x_i - m) > 0\}, \end{aligned}$$

and the maximum likelihood estimators of σ_j^2 and τ_j are

$$\begin{aligned} \hat{\tau}_j(m, W) &= \left(\frac{s_{2j}}{s_{1j}} \right)^{1/3}, \quad 1 \leq j \leq d \\ \hat{\sigma}_j^2(m, W) &= \begin{cases} \frac{1}{n} s_{1j}^{2/3} g_j(m, W), & 1 \leq j \leq d \\ \frac{1}{n} (s_{1j} + s_{2j}), & d < j \leq D \end{cases} \end{aligned} \quad (3)$$

Proof. See Section 5 (Appendix A). \square

Thanks to the above theorem we can reduce the search for the maximum of the log-likelihood function for two parameters $m \in \mathbb{R}^d$ and $W \in \mathcal{M}(\mathbb{R}^d)$.

$$l(X; m, W) = \frac{1}{|\det(W)|^{\frac{D}{2}}} \prod_{j=1}^d g_j(m, W) \prod_{j=d+1}^D (s_{1j} + s_{2j})^{\frac{1}{3}} \quad (4)$$

where w_j stands for the j -th column of matrix W . Consequently, maximization of likelihood function is equivalent to minimization of $\ln l$.

Corollary 3.1. *Let $X \subset \mathbb{R}^d$, $m \in \mathbb{R}^d$, $W \in \mathcal{M}(\mathbb{R}^d)$ be given, then*

$$\operatorname{argmax}_{m, W} \hat{L}(X; m, W) = \operatorname{argmin}_{m, W} \ln l(X; m, W).$$

To minimize $\ln l$ with the use classical gradient descent method, we need the formula for $\nabla \ln l$ (gradient of the cost function).

Theorem 3.3. *Let $X \subset \mathbb{R}^d$, $m = (m_1, \dots, m_d)^T \in \mathbb{R}^d$, $W = (w_{ij})_{1 \leq i, j \leq d}$ non-singular be given. Then*

$$\nabla_{\mathbf{m}} \ln l(X; \mathbf{m}, W) = \left(\frac{\partial \ln l(X; \mathbf{m}, W)}{\partial m_1}, \dots, \frac{\partial \ln l(X; \mathbf{m}, W)}{\partial m_d} \right)^T,$$

where

$$\begin{aligned} \frac{\partial \ln l(X; \mathbf{m}, W)}{\partial m_k} = & \sum_{j=1}^d \frac{-2}{3(s_{1j}^{\frac{1}{3}} + s_{2j}^{\frac{1}{3}})} \left(\frac{1}{s_{1j}^{\frac{2}{3}}} \sum_{i \in I_j} w_j^T (x_i - \mathbf{m}) w_{jk} + \right. \\ & \left. \frac{1}{s_{2j}^{\frac{2}{3}}} \sum_{i \in I_j^c} w_j^T (x_i - \mathbf{m}) w_{jk} \right) + \sum_{j=d+1}^D \frac{-2}{3(s_{1j} + s_{2j})} \cdot \\ & \left(\sum_{i \in I_j} w_j^T (x_i - \mathbf{m}) w_{jk} + \sum_{i \in I_j^c} w_j^T (x_i - \mathbf{m}) w_{jk} \right). \end{aligned}$$

Moreover, $\nabla_W \ln l(X; \mathbf{m}, W) = \left[\frac{\partial \ln l(X; \mathbf{m}, W)}{\partial w_{pk}} \right]_{1 \leq p, k \leq d},$

where $\frac{\partial \ln l(X; \mathbf{m}, W)}{\partial w_{pk}} =$

$$\begin{aligned} & -\frac{2}{3} (w^{-1})_{pk}^T + \frac{2}{3(s_{1p}^{\frac{1}{3}} + s_{2p}^{\frac{1}{3}})} \left(s_{1p}^{-\frac{2}{3}} \sum_{i \in I_p} w_p^T (x_i - \mathbf{m}) (x_{ik} - m_k) \right. \\ & \left. + s_{2p}^{-\frac{2}{3}} \sum_{i \in I_p^c} w_p^T (x_i - \mathbf{m}) (x_{ik} - m_k) \right) + \frac{2}{3(s_{1p} + s_{2p})} \left(\sum_{i \in I_p} w_p^T (x_i - \mathbf{m}) (x_{ik} - m_k) + \sum_{i \in I_p^c} w_p^T (x_i - \mathbf{m}) (x_{ik} - m_k) \right) \end{aligned}$$

and

$$\begin{aligned} s_{1j} &= \sum_{i \in I_j} [w_j^T (x_i - \mathbf{m})]^2, I_j = \{1 \leq i \leq n: w_j^T (x_i - \mathbf{m}) \leq 0\}, \\ s_{2j} &= \sum_{i \in I_j^c} [w_j^T (x_i - \mathbf{m})]^2, I_j^c = \{1 \leq i \leq n: w_j^T (x_i - \mathbf{m}) > 0\}. \end{aligned}$$

Proof. See Section 6 (Appendix B). \square

Thanks to the above Theorem we are able to use in our experiments the gradient descent for finding the minimum of our cost function.

4. Experiments

To compare ICA_{SN} to other state-of-the-art approaches we use Tucker's congruence coefficient (Lorenzo-Seva & Ten Berge, 2006) which values range between -1 and +1. It can be used to study the similarity of extracted factors across different samples. Generally, a congruence coefficient of 0.9 indicates a high degree of factor similarity, while a coefficient of 0.95 or higher indicates that the factors are virtually identical.

We evaluate our method in the context of 2D and hyperspectral images. For comparison we use R package *ica* (Helwig, 2015), *PearsonICA* (Karvanen, 2008), *ProDenICA* (Hastie & Tibshirani, 2010), *tsBSS* (Matiainen et al., 2016). The most popular method used in practice is FastICA (Hyvärinen, 1999; Helwig & Hong, 2013) algorithm, which uses negentropy. In this context we can use three different functions to estimate neg-entropy: log-cosh, exp and kurtosis. We also compare our method with

algorithm using Information-Maximization (Infomax) approach (Bell & Sejnowski, 1995). Similarly to FastICA we consider three possible non-linear functions: hyperbolic tangent, logistic and extended Infomax.

4.1. Separation of images

One of the most popular application of ICA is the separation of images. In our experiments we use four images from the USC-SIPI Image Database of size 256 × 256 pixels (4.1.01, 4.1.06, 4.1.02, 4.1.03) and eight of size 512 × 512 pixels (4.2.04, 4.2.02, boat.512, elaine.512, 5.2.10, 5.2.08, 5.3.01, 4.2.03). We also use 8 images from the Berkeley Segmentation Dataset of size 482 × 321 with indexes (#119082, #42049, #43074, #38092, #157055, #220075, #295087, #167062).

We make random pairs of above images and one component with noise (random sample from Gaussian distribution $\mathcal{N}(0, 1)$) and use them as a source signal combined by

the mixing matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \end{bmatrix}$. Our goal was to

reconstruct two original images by using only the knowledge about mixed ones. The visualization of this process we present in Fig. 1. The results of this experiment are presented in Tab. 1 where we present Tucker's congruence coefficients which shows that almost in all cases ICA_{SN} obtains best results. This is illustrated in Figure 1, where we can see that ICA_{SN} almost perfectly recovered source signal. Although this is not surprising as the experiments were in fact conducted in the setting which favored our approach, as we chose the noise to be gaussian, this shows that ICA_{SN} works as desired and deals well with removing gaussian components from the data.

4.2. Hyperspectral Unmixing

Independent component analysis has been recently applied into hyperspectral unmixing (Wang et al., 2015; Caiafa et al., 2008) as a result of its low computation time and its ability to perform without prior information. In this subsection we apply simple example which suggests that our method also can be used for spectral data.

Urban data (Zhu et al., 2014a;c;b) is one of the most widely used hyperspectral data-sets used in the hyperspectral unmixing study. Each image has 307 × 307 pixels, each of which corresponds to a 2 × 2 m area. In this image, there are 210 wavelengths ranging from 400 nm to 2500 nm, resulting in a spectral resolution of 10 nm. After the channels 1–4, 76, 87, 101–111, 136–153 and 198–210 are removed (due to dense water vapor and atmospheric effects), there remain 162 channels (this is a common preprocess for hyperspectral unmixing analyses). There is ground truth (Zhu et al., 2014a;c;b), which contains 4 channels: #1 Asphalt,

Table 1. Tucker's congruence coefficients for reconstruction of two images.

	ICA _{SN}	FASTICA LOGCOSH	FASTICA EXP	FASTICA KURTOSIS	INFOMAX TANH	INFOMAX TANGENT	INFOMAX LOGISTIC	JADE	PEARSONICA	PRODENICA	FIXNA
4.1.01	0.6807	0.0325	0.0321	0.0352	0.0327	0.0334	0.0449	0.029	0.0549	0.1184	0.5487
4.1.02	0.5491	0.4588	0.4588	0.459	0.4589	0.4589	0.4593	0.4584	0.4592	0.1781	0.0074
4.1.06	0.4279	0.0143	0.0142	0.0149	0.0142	0.0145	0.0143	0.0147	0.0141	0.0777	0.4295
4.1.03	0.2033	0.4151	0.4151	0.415	0.4151	0.4151	0.4151	0.4151	0.4152	0.0932	0.0107
4.2.04	0.4333	0.0234	0.0233	0.0238	0.0234	0.0236	0.0234	0.0234	0.3681	0.064	0.3647
5.2.10	0.3161	0.3681	0.3681	0.3681	0.3681	0.3681	0.3681	0.3681	0.0234	0.0704	0.2652
4.2.02	0.0235	0.0066	0.0067	0.011	0.0066	0.0066	0.0067	0.0109	0.282	0.0164	0.1244
5.2.08	0.1279	0.282	0.282	0.2818	0.282	0.282	0.282	0.2818	0.0066	0.0099	4E-04
BOAT.512	0.5454	0.0355	0.0357	0.0329	0.0356	0.0339	0.0355	0.035	0.4631	0.0388	0.2753
5.3.01	0.3368	0.463	0.4631	0.463	0.463	0.463	0.463	0.463	0.0367	0.3282	0.4778
ELAINE.512	0.3134	0.0213	0.0283	0.0173	0.0213	0.0175	0.0213	0.0197	0.2507	0.0469	0.2296
4.2.03	0.3216	0.2509	0.2511	0.2506	0.2509	0.2506	0.2509	0.2508	0.0189	0.1894	0.2567
119082	0.5744	0.0428	0.043	0.042	0.0427	0.0423	0.0416	0.0402	0.0412	0.0921	0.371
157055	0.3612	0.482	0.482	0.482	0.482	0.482	0.4819	0.4818	0.4819	0.0599	0.0032
42049	0.5287	0.0398	0.0396	0.039	0.04	0.04	0.0399	0.037	0.0382	0.5014	0.2964
220075	0.3043	0.3668	0.3668	0.3669	0.3668	0.3668	0.3668	0.3671	0.3669	0.061	0.4956
43074	0.3886	0.0501	0.0692	0.0281	0.0502	0.0331	0.0501	0.037	0.0346	0.0461	0.3168
295087	0.401	0.3384	0.3336	0.3407	0.3384	0.3405	0.3384	0.3403	0.3404	0.0676	0.3025
38092	0.3841	0.0504	0.0502	0.0497	0.0504	0.0499	0.0504	0.0497	0.0503	0.2541	0.4965
167062	0.7406	0.6963	0.6962	0.6962	0.6963	0.6962	0.6963	0.6962	0.6963	0.587	0.5496

#2 Grass, #3 Tree and #4 Roof.

A highly mixed area is cut from the original data set in this experiment (similar example was showed in (Wang et al., 2015)), with the size of 200×150 pixels.

In our experiment we compared ICA_{SN} to other two popular ICA methods – ProDenICA and FastICA, see Fig. 2. Observe that ICA_{SN} and ProDenICA give layers which seem to contain more information than FastICA, as the last component in FastICA contains mainly noise.

5. Appendix A

Proof of Theorem 3.2. Let $X = \{x_1, \dots, x_n\}$. We write

$$z_i = W(x_i - m), \quad z_{ij} = w_j^T(x_i - m),$$

for observation i , where $i = 1, \dots, n$ and coordinates $j = 1, \dots, d$.

Let us consider the likelihood function, i.e.

$$\begin{aligned}
 L(X; m, W, \sigma, \tau) &= \prod_{i=1}^n SN_d N_{D-d}(x_i; m, W, \sigma^2, \tau^2) \\
 &= \prod_{i=1}^n |\det(W)| \prod_{j=1}^d SN(w_j^T(x_i - m); 0, \sigma_j^2, \tau_j^2) \cdot \\
 &\quad \prod_{j=d+1}^D N(w_j^T(x_i - m); 0, \sigma_j^2) = \left(c_1 |\det(W)|\right)^n \\
 &\quad \left(\prod_{j=1}^d \sigma_j (1 + \tau_j)\right)^{-n} \prod_{i=1}^n \prod_{j=1}^d \exp\left[-\frac{1}{2\sigma_j^2} z_{ij}^2 (\mathbb{1}_{\{z_{ij} \leq 0\}} + \right. \\
 &\quad \left. \tau_j^{-2} \mathbb{1}_{\{z_{ij} > 0\}})\right] \left(\prod_{j=d+1}^D \sigma_j\right)^{-n} \prod_{i=1}^n \prod_{j=d+1}^D \exp\left[-\frac{1}{2\sigma_j^2} z_{ij}^2\right],
 \end{aligned}$$

where $c_1 = \left(\sqrt{\frac{2}{\pi}}\right)^d \cdot \left(\frac{1}{\sqrt{2\pi}}\right)^{D-d}$. Now we take the log-

likelihood function, i.e. $\ln(L(X; m, W, \sigma, \tau)) =$

$$\begin{aligned}
 &\ln\left(\left(c_1 |\det(W)|\right)^n \left(\prod_{j=1}^d \sigma_j (1 + \tau_j)\right)^{-n} \left(\prod_{j=d+1}^D \sigma_j\right)^{-n}\right) + \\
 &\sum_{i=1}^n \sum_{j=1}^d \left[-\frac{1}{2\sigma_j^2} z_{ij}^2 (\mathbb{1}_{\{z_{ij} \leq 0\}} + \tau_j^{-2} \mathbb{1}_{\{z_{ij} > 0\}})\right] + \\
 &\sum_{i=1}^n \sum_{j=d+1}^D \left[-\frac{1}{2\sigma_j^2} z_{ij}^2\right] \\
 &= \ln\left(\left(c_1 |\det(W)|\right)^n \left(\prod_{j=1}^d (1 + \tau_j)\right)^{-n} \left(\prod_{j=1}^D \sigma_j\right)^{-n}\right) - \\
 &\frac{1}{2} \sum_{j=1}^d \left(\sigma_j^{-2} \sum_{i \in I_j} z_{ij}^2 + \frac{\sigma_j^{-2}}{\tau_j^2} \sum_{i \in I_j^c} z_{ij}^2\right) - \\
 &\frac{1}{2} \sum_{j=d+1}^D \sigma_j^{-2} \left(\sum_{i \in I_j} z_{ij}^2 + \sum_{i \in I_j^c} z_{ij}^2\right) \\
 &= \ln\left(\left(c_1 |\det(W)|\right)^n \left(\prod_{j=1}^d (1 + \tau_j)\right)^{-n} \left(\prod_{j=1}^D \sigma_j\right)^{-n}\right) - \\
 &\sum_{j=1}^d \frac{1}{2\sigma_j^2} \left(s_{1j} + \frac{1}{\tau_j^2} s_{2j}\right) - \sum_{j=d+1}^D \frac{1}{2\sigma_j^2} (s_{1j} + s_{2j}).
 \end{aligned}$$

We fix m , W and maximize the log-likelihood function over τ and σ . In such a case we have to solve the following system of equations

$$\frac{\partial \ln(L(X; m, W, \sigma, \tau))}{\partial \sigma_j} = 0, \quad \frac{\partial \ln(L(X; m, W, \sigma, \tau))}{\partial \tau_j} = 0,$$

for $j = 1, \dots, D$. Hence

$$\begin{aligned}
 -\frac{n}{\sigma_j} + \sigma_j^{-3} (s_{1j} + \tau_j^{-2} s_{2j}) &= 0, \text{ for } j = 1, \dots, d, \\
 -\frac{n}{\sigma_j} + \sigma_j^{-3} (s_{1j} + s_{2j}) &= 0, \text{ for } j > d, \\
 -\frac{n}{1 + \tau_j} + \frac{s_{2j}}{\tau_j^3 \sigma_j^2} &= 0, \text{ for } j = 1, \dots, d.
 \end{aligned}$$

By simple calculations we obtain the expressions for the estimators in 3. Substituting it into the log-likelihood func-



(a) Ground truth layers which contains 4 channels: #1 Asphalt, #2 Grass, #3 Tree and #4 Roof.


 (b) The effect of the ICA_{SN} method.


(c) The effect of the FastICA (logcosh) method.



(d) The effect of the ProDenICA method.

Figure 2. Results of image separation with the uses of various ICA algorithms.

tion, we get $\hat{L}(m, W) =$

$$= \left(\frac{2}{\pi} \right)^{\frac{dn}{2}} \left(\frac{1}{2\pi} \right)^{\frac{(D-d)n}{2}} |\det(W)|^n \left(\prod_{j=1}^d \frac{1}{\sqrt{n}} g_j(m, W)^{\frac{3}{2}} \right)^{-n} \\ e^{-\frac{dn}{2}} \left(\prod_{j=d+1}^D \left(\frac{s_{1j} + s_{2j}}{n} \right)^{\frac{1}{2}} \right)^{-n} = \frac{2^{(d-D/2)n} n^{dn/2}}{(\pi e)^{Dn/2}}. \\ \left(\frac{1}{|\det(W)|^{\frac{2}{3}}} \prod_{j=1}^d g_j(m, W) \right)^{-\frac{3n}{2}} \left(\prod_{j=d+1}^D \left(\frac{s_{1j} + s_{2j}}{n} \right)^{-\frac{n}{2}} \right)$$

6. Appendix B

We will need the following well-known lemma (for the convenience of the reader we provide the proof).

Lemma 6.1. *Let $A = (a_{ij})_{1 \leq i, j \leq d}$ be a differentiable map from real numbers to $d \times d$ matrices then*

$$\frac{\partial \det(A)}{\partial a_{ij}} = \text{adj}^T(A)_{ij}, \quad (5)$$

where $\text{adj}(A)$ stands for the adjugate of A , i.e. the transpose of the cofactor matrix.

□ *Proof.* By the Laplace expansion $\det A =$

$\sum_{j=1}^d (-1)^{i+j} a_{ij} M_{ij}$ where M_{ij} is the minor of the entry in the i -th row and j -th column. Hence

$$\frac{\partial \det A}{\partial a_{ij}} = (-1)^{i+j} M_{ij} = \text{adj}^T(A)_{ij}.$$

□

Proof of Theorem 3.3. Let us start with the partial derivative of $\ln(l)$ with respect to \mathbf{m} . We have

$$\begin{aligned} \frac{\partial \ln l(X; \mathbf{m}, W)}{\partial \mathbf{m}_k} &= \sum_{j=1}^d \frac{\partial \ln(g_j(\mathbf{m}, W))}{\partial \mathbf{m}_k} + \sum_{j=d+1}^D \frac{\partial \ln((s_{1j} + s_{2j})^{\frac{1}{3}})}{\partial \mathbf{m}_k} \\ &= \sum_{j=1}^d \frac{1}{s_{1j} + s_{2j}} \frac{\partial (s_{1j} + s_{2j})}{\partial \mathbf{m}_k} + \sum_{j=d+1}^D \frac{1}{(s_{1j} + s_{2j})^{\frac{2}{3}}} \frac{\partial ((s_{1j} + s_{2j})^{\frac{1}{3}})}{\partial \mathbf{m}_k} \\ &= \sum_{j=1}^d \frac{1}{s_{1j} + s_{2j}} \left(\frac{1}{3s_{1j}} \frac{\partial s_{1j}}{\partial \mathbf{m}_k} + \frac{1}{3s_{2j}} \frac{\partial s_{2j}}{\partial \mathbf{m}_k} \right) \\ &\quad + \sum_{j=d+1}^D \frac{1}{(s_{1j} + s_{2j})^{\frac{2}{3}}} \frac{1}{3} \frac{1}{(s_{1j} + s_{2j})^{\frac{2}{3}}} \left(\frac{\partial s_{1j}}{\partial \mathbf{m}_k} + \frac{\partial s_{2j}}{\partial \mathbf{m}_k} \right). \end{aligned}$$

Now, we need $\frac{\partial s_{1j}}{\partial \mathbf{m}_k}$ and $\frac{\partial s_{2j}}{\partial \mathbf{m}_k}$, therefore

$$\begin{aligned} \frac{\partial s_{1j}}{\partial \mathbf{m}_k} &= \sum_{i \in I_j} \frac{\partial [w_j^T(\mathbf{x}_i - \mathbf{m})]^2}{\partial \mathbf{m}_k} = \\ &\sum_{i \in I_j} 2w_j^T(\mathbf{x}_i - \mathbf{m}) \frac{\partial w_j^T(\mathbf{x}_i - \mathbf{m})}{\partial \mathbf{m}_k} = \sum_{i \in I_j} -2w_j^T(\mathbf{x}_i - \mathbf{m}) w_{jk}. \end{aligned}$$

Analogously we get

$$\frac{\partial s_{2j}}{\partial \mathbf{m}_k} = \sum_{i \in I_j^c} -2w_j^T(\mathbf{x}_i - \mathbf{m}) w_{jk}.$$

Hence

$$\begin{aligned} \frac{\partial \ln l}{\partial \mathbf{m}_k} &= \sum_{j=1}^d \frac{-1}{s_{1j} + s_{2j}} \left(\frac{1}{3s_{1j}} \sum_{i \in I_j} 2w_j^T(\mathbf{x}_i - \mathbf{m}) w_{jk} + \right. \\ &\quad \left. \frac{1}{3s_{2j}} \sum_{i \in I_j^c} 2w_j^T(\mathbf{x}_i - \mathbf{m}) w_{jk} \right) + \sum_{j=d+1}^D \frac{-1}{3(s_{1j} + s_{2j})} \\ &\quad \left(\sum_{i \in I_j} 2w_j^T(\mathbf{x}_i - \mathbf{m}) w_{jk} + \sum_{i \in I_j^c} 2w_j^T(\mathbf{x}_i - \mathbf{m}) w_{jk} \right). \end{aligned}$$

Now we calculate the partial derivative of $\ln l(X; \mathbf{m}, W)$ with respect to the matrix W . We have $\frac{\partial \ln l(X; \mathbf{m}, W)}{\partial w_{pk}} =$

$$\frac{\partial \ln |\det(W)|^{-\frac{2}{3}}}{\partial w_{pk}} + \sum_{j=1}^d \frac{\partial \ln(g_j(\mathbf{m}, W))}{\partial w_{pk}} + \sum_{j=d+1}^D \frac{\partial \ln((s_{1j} + s_{2j})^{\frac{1}{3}})}{\partial w_{pk}}.$$

To calculate the derivative of the determinant we use Jacobi's formula (see Lemma 6.1). Hence

$$\begin{aligned} \frac{\partial \ln(\det(W)^{-\frac{2}{3}})}{\partial w_{pk}} &= \det(W)^{\frac{2}{3}} \left(-\frac{2}{3} \right) \det(W)^{-\frac{5}{3}} \frac{\partial \det(W)}{\partial w_{pk}} \\ &= -\frac{2}{3} \det(W)^{-1} \text{adj}^T(W)_{pk} \\ &= -\frac{2}{3} \frac{1}{\det(W)} [\det(W) (W^{-1})_{pk}^T] = -\frac{2}{3} (w^{-1})_{pk}^T, \end{aligned}$$

where $(w^{-1})_{pk}^T$ is the element in the p -th row and k -th column of the matrix $(W^{-1})^T$. Now we calculate

$$\begin{aligned} \frac{\partial \ln(g_j(\mathbf{m}, W))}{\partial w_{pk}} &= \frac{1}{s_{1j} + s_{2j}} \frac{\partial (s_{1j} + s_{2j})^{\frac{1}{3}}}{\partial w_{pk}} = \\ &\frac{1}{s_{1j} + s_{2j}} \left(\frac{1}{3s_{1j}} \frac{\partial s_{1j}}{\partial w_{pk}} + \frac{1}{3s_{2j}} \frac{\partial s_{2j}}{\partial w_{pk}} \right), \end{aligned}$$

where

$$\begin{aligned} \frac{\partial s_{1j}}{\partial w_{pk}} &= \sum_{i \in I_j} \frac{\partial [w_j^T(\mathbf{x}_i - \mathbf{m})]^2}{\partial w_{pk}} = \sum_{i \in I_j} 2w_j^T(\mathbf{x}_i - \mathbf{m}) \frac{\partial w_j^T(\mathbf{x}_i - \mathbf{m})}{\partial w_{pk}} \\ &= \begin{cases} 0, & \text{if } j \neq p \\ \sum_{i \in I_p} 2w_p^T(\mathbf{x}_i - \mathbf{m})(x_{ik} - m_k), & \text{if } j = p \end{cases} \end{aligned}$$

and x_{ik} is the k -th element of the vector \mathbf{x}_i . Analogously we get

$$\frac{\partial s_{2j}}{\partial w_{pk}} = \begin{cases} 0, & \text{if } j \neq p \\ \sum_{i \in I_p^c} 2w_p^T(\mathbf{x}_i - \mathbf{m})(x_{ik} - m_k), & \text{if } j = p. \end{cases}$$

Moreover,

$$\begin{aligned} \frac{\partial \ln((s_{1j} + s_{2j})^{\frac{1}{3}})}{\partial w_{pk}} &= \frac{1}{(s_{1j} + s_{2j})^{\frac{2}{3}}} \frac{\partial ((s_{1j} + s_{2j})^{\frac{1}{3}})}{\partial w_{pk}} = \\ &\frac{1}{(s_{1j} + s_{2j})^{\frac{2}{3}}} \frac{1}{3} \frac{1}{(s_{1j} + s_{2j})^{\frac{2}{3}}} \left(\frac{\partial s_{1j}}{\partial w_{pk}} + \frac{\partial s_{2j}}{\partial w_{pk}} \right), \end{aligned}$$

Hence we obtain

$$\begin{aligned} \frac{\partial \ln l}{\partial w_{pk}} &= -\frac{2}{3} (w^{-1})_{pk}^T + \\ &\frac{1}{s_{1p} + s_{2p}} \left(\frac{1}{3s_{1p}} \sum_{i \in I_p} 2w_p^T(\mathbf{x}_i - \mathbf{m})(x_{ik} - m_k) \right. \\ &\quad \left. + \frac{1}{3s_{2p}} \sum_{i \in I_p^c} 2w_p^T(\mathbf{x}_i - \mathbf{m})(x_{ik} - m_k) \right) + \\ &\frac{1}{3(s_{1p} + s_{2p})} \left(\sum_{i \in I_p} 2w_p^T(\mathbf{x}_i - \mathbf{m})(x_{ik} - m_k) + \right. \\ &\quad \left. \sum_{i \in I_p^c} 2w_p^T(\mathbf{x}_i - \mathbf{m})(x_{ik} - m_k) \right). \end{aligned}$$

□

References

- Allasonniere, Stéphanie and Younes, Laurent. A stochastic algorithm for probabilistic independent component analysis. *The Annals of Applied Statistics*, pp. 125–160, 2012.
- Attias, H and Schreiner, CE. Blind source separation and deconvolution by dynamic component analysis. In *Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop*, pp. 456–465. IEEE, 1997.

- Beckmann, Christian F. Modelling with independent components. *Neuroimage*, 62(2):891–901, 2012.
- Beckmann, Christian F and Smith, Stephen M. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, 23(2):137–152, 2004.
- Bell, Anthony J and Sejnowski, Terrence J. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159, 1995.
- Belouchrani, Adel, Cardoso, Jean-François, et al. Maximum likelihood source separation by the expectation-maximization technique: Deterministic and stochastic implementation. In *Proc. Nolta*, volume 95, pp. 49–53. Citeseer, 1995.
- Blanchard, Gilles, Kawanabe, Motoaki, Sugiyama, Masashi, Spokoiny, Vladimir, and Mazller, Klaus-Robert. In search of non-gaussian components of a high-dimensional distribution. *Journal of Machine Learning Research*, 7(Feb):247–282, 2006.
- Caiafa, Cesar F, Salerno, Emanuele, Proto, Araceli N, and Fiumi, L. Blind spectral unmixing by local maximization of non-gaussianity. *Signal Processing*, 88(1):50–68, 2008.
- Cover, Thomas M and Thomas, Joy A. *Elements of information theory*. John Wiley and Sons, 2012.
- Green, Christopher G, Nandy, Rajesh R, and Cordes, Dietmar. Pca-preprocessing of fmri data adversely affects the results of ica. In *Proceedings of international society of magnetic resonance in medicine*, volume 10, 2002.
- Hastie, Trevor and Tibshirani, Rob. *ProDenICA: Product Density Estimation for ICA using tilted Gaussian density estimates*, 2010. URL <http://CRAN.R-project.org/package=ProDenICA>. R package version 1.0.
- Helwig, Nathaniel E. *ica: Independent Component Analysis*, 2015. URL <http://CRAN.R-project.org/package=ica>. R package version 1.0-1.
- Helwig, Nathaniel E and Hong, Sungjin. A critique of tensor probabilistic independent component analysis: implications and recommendations for multi-subject fmri data analysis. *Journal of neuroscience methods*, 213(2):263–273, 2013.
- Hyvärinen, Aapo. Fast and robust fixed-point algorithms for independent component analysis. *Neural Networks, IEEE Transactions on*, 10(3):626–634, 1999.
- Hyvärinen, Aapo and Oja, Erkki. Independent component analysis: algorithms and applications. *Neural networks*, 13(4):411–430, 2000.
- Hyvärinen, Aapo, Karhunen, Juha, and Oja, Erkki. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.
- John, Sreeba. The three-parameter two-piece normal family of distributions and its fitting. *Communications in Statistics-Theory and Methods*, 11(8):879–885, 1982.
- Karvanen, J. *PearsonICA*, 2008. URL <https://CRAN.R-project.org/package=PearsonICA>. R package version 1.2-3.
- Kawanabe, Motoaki, Sugiyama, Masashi, Blanchard, Gilles, and Müller, Klaus-Robert. A new algorithm of non-gaussian component analysis with radial kernel functions. *Annals of the Institute of Statistical Mathematics*, 59(1):57–75, 2007.
- Lorenzo-Seva, Urbano and Ten Berge, Jos MF. Tucker’s congruence coefficient as a meaningful index of factor similarity. *Methodology*, 2(2):57–64, 2006.
- Matilainen, Markus, Miettinen, Jari, Nordhausen, Klaus, Oja, Hannu, and Taskinen, Sara. *tsBSS: Tools for Blind Source Separation for Time Series*, 2016. URL <http://CRAN.R-project.org/package=tsBSS>. R package version 0.2.
- Miettinen, Jari, Nordhausen, Klaus, Oja, Hannu, and Taskinen, Sara. Deflation-based fastica with adaptive choices of nonlinearities. *IEEE Transactions on Signal Processing*, 62(21):5716–5724, 2014.
- Moulines, Eric, Cardoso, J-F, and Gassiat, Elisabeth. Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, volume 5, pp. 3617–3620. IEEE, 1997.
- Pham, Dinh Tuan and Garat, Philippe. Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *Signal Processing, IEEE Transactions on*, 45(7):1712–1725, 1997.
- Risk, Benjamin B, Matteson, David S, and Ruppert, David. Likelihood component analysis. *arXiv preprint arXiv:1511.01609*, 2015.
- Samarov, Alexander, Tsybakov, Alexandre, et al. Nonparametric independent component analysis. *Bernoulli*, 10(4):565–582, 2004.

990	Samworth, Richard J, Yuan, Ming, et al. Independent com-	1045
991	ponent analysis via nonparametric maximum likelihood	1046
992	estimation. <i>The Annals of Statistics</i> , 40(6):2973–3002,	1047
993	2012.	1048
994		1049
995	Shi, Ran, Guo, Ying, et al. Investigating differences in	1050
996	brain functional networks using hierarchical covariate-	1051
997	adjusted independent component analysis. <i>The Annals</i>	1052
998	<i>of Applied Statistics</i> , 10(4):1930–1957, 2017.	1053
999		1054
1000	Spurek, P. et al. Ica based on the data asymmetry. <i>Pattern</i>	1055
1001	<i>Recognition (accepted)</i> , 2017.	1056
1002		1057
1003	Tipping, Michael E and Bishop, Christopher M. Probabilis-	1058
1004	tic principal component analysis. <i>Journal of the Royal</i>	1059
1005	<i>Statistical Society: Series B (Statistical Methodology)</i> ,	1060
1006	61(3):611–622, 1999.	1061
1007		1062
1008	Villani, Mattias and Larsson, Rolf. The multivariate split	1063
1009	normal distribution and asymmetric principal compo-	1064
1010	nents analysis. <i>Communications in StatisticsTheory and</i>	1065
1011	<i>Methods</i> , 35(6):1123–1140, 2006.	1066
1012		1067
1013	Wang, Nan, Du, Bo, Zhang, Liangpei, and Zhang, Lifu.	1068
1014	An abundance characteristic-based independent compo-	1069
1015	nent analysis for hyperspectral unmixing. <i>IEEE Trans-</i>	1070
1016	<i>actions on Geoscience and Remote Sensing</i> , 53(1):416–	1071
1017	428, 2015.	1072
1018		1073
1019	Zhu, Feiyun, Wang, Ying, and Bin Fan, Shiming Xiang, and	1074
1020	Pan, Chunhong. Structured sparse method for hyper-	1075
1021	spectral unmixing. <i>ISPRS Journal of Photogrammetry</i>	1076
1022	<i>and Remote Sensing</i> , 88:101–118, 2014a.	1077
1023		1078
1024	Zhu, Feiyun, Wang, Ying, Fan, Bin, Meng, Gaofeng, and	1079
1025	Pan, Chunhong. Effective spectral unmixing via ro-	1080
1026	bust representation and learning-based sparsity. <i>CoRR</i> ,	1081
1027	abs/1409.0685, 2014b.	1082
1028		1083
1029	Zhu, Feiyun, Wang, Ying, Fan, Bin, Meng, Gaofeng, Xi-	1084
1030	ang, Shiming, and Pan, Chunhong. Spectral unmixing	1085
1031	via data-guided sparsity. <i>CoRR</i> , abs/1403.3155, 2014c.	1086
1032		1087
1033		1088
1034		1089
1035		1090
1036		1091
1037		1092
1038		1093
1039		1094
1040		1095
1041		1096
1042		1097
1043		1098
1044		1099