

## COMPARISON OF PRINCIPAL COMPONENT ANALYSIS AND INDEPENDENT COMPONENT ANALYSIS FOR BLIND SOURCE SEPARATION

R. MUTIHAC<sup>1</sup>, MARC M. VAN HULLE<sup>2</sup>

<sup>1</sup> University of Bucharest, Department of Electricity and Biophysics, RO-76900, Romania,  
<sup>2</sup> K.U. Leuven, Labo voor Neuro- en Psychofysiologie, B-3000 Leuven, Belgium,  
{radu.mutihac, marc.vanhulle}@med.kuleuven.ac.be

*Abstract.* Our contribution briefly outlines the basics of the well-established technique in data mining, namely the principal component analysis (PCA), and a rapidly emerging novel method, that is, the independent component analysis (ICA). The performance of PCA singular value decomposition-based and stationary linear ICA in blind separation of artificially generated data out of linear mixtures was critically evaluated and compared. All our results outlined the superiority of ICA relative to PCA in faithfully retrieval of the original independent source components.

*Key words:* principal component analysis (PCA), independent component analysis (ICA), blind source separation (BSS), and higher-order statistics.

### 1. INTRODUCTION

In data analysis, a common task consists in finding an adequate representation of multivariate data, which is expected to provide the underlying factors describing their essential structure. Linear transformations are often envisaged to perform such a task due to their computational and conceptual simplicity. Some common linear transformation methods are principal component analysis (PCA), factor analysis (FA), projection pursuit (PP), and more recently the independent component analysis (ICA). ICA has emerged as a useful extension of PCA and developed in context with blind source separation (BSS) and digital signal processing (DSP) [1], [2]. ICA is also related to recent theories of the visual brain, which assume that consecutive processing steps lead to a progressive reduction in the redundancy of the representation [3], [4]. It is also related to work on sparse and low entropy coding [5]. A major area of application for PCA and ICA is biomedical time series analysis, with benefits in both fundamental research on the human brain and in medical diagnosis and treatment [6], [7], [8].

## 2. THE PCA AND ICA MODELS

Mathematically, the stationary linear PCA and ICA models can be defined on the basis of a common data model. Suppose that some stochastic processes are represented by three random (column) vectors  $\mathbf{x}(t)$ ,  $\mathbf{n}(t) \in \mathfrak{R}^N$ , and  $\mathbf{s}(t) \in \mathfrak{R}^M$  with zero mean and finite covariance, with the components of  $\mathbf{s}(t)$  being statistically independent and at most one Gaussian. Let  $\mathbf{A}$  be a rectangular constant full column rank  $N \times M$  matrix with at least as many rows as columns ( $N \geq M$ ), and denote by  $t$  the sample index (e.g. time or point) taking discrete values  $t = 1, 2, \dots, T$ , then we suppose:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) = \sum_{i=1}^M s_i(t) \mathbf{a}_i + \mathbf{n}(t) \quad (1)$$

Here  $\mathbf{s}(t)$ ,  $\mathbf{x}(t)$ ,  $\mathbf{n}(t)$ , and  $\mathbf{A}$  are the sources, the observed data, the (unknown) noise in data, and the (unknown) mixing matrix, respectively, whereas the columns of  $\mathbf{A}$  denoted by  $\mathbf{a}_i$ ,  $i = 1, 2, \dots, M$  are the ICA basis vectors. Mixing is supposed to be instantaneous, so there is no time delay between the (latent) source variable  $s_i(t)$  mixing into an observable (data) variable  $x_j(t)$ , with  $i = 1, 2, \dots, M$  and  $j = 1, 2, \dots, N$ .

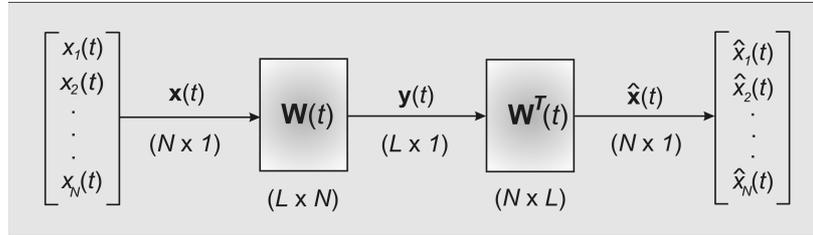


Fig. 1. – Decorrelation of data  $\mathbf{x}(t)$  by an orthogonal matrix  $\mathbf{W}(t)$  yielding an approximation  $\hat{\mathbf{x}}(t)$  of  $\mathbf{x}(t)$ .

Consider that the stochastic vector process  $\{\mathbf{x}(t)\} \in \mathfrak{R}^N$  has the mean  $E\{\mathbf{x}(t)\} = 0$  and its covariance matrix  $\mathbf{C}_x = E\{\mathbf{x}(t)\mathbf{x}(t)^T\}$ . The goal of PCA (Fig. 1) is to identify the dependence structure in each dimension and to come out with an orthogonal transform matrix  $\mathbf{W}$  of size  $L \times N$  from  $\mathfrak{R}^N$  to  $\mathfrak{R}^L$ ,  $L < N$ , such that the  $L$ -dimensional output vector  $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$  sufficiently represents the intrinsic features of the input data, and where the covariance matrix  $\mathbf{C}_y$  of  $\{\mathbf{y}(t)\}$  is a diagonal matrix  $\mathbf{D}$  with the diagonal elements arranged in a descending order, that is  $d_{i,i} \geq d_{i+1,i+1}$ . The reconstruction of  $\{\mathbf{x}(t)\}$  from  $\{\mathbf{y}(t)\}$  denoted by  $\{\hat{\mathbf{x}}(t)\}$  is consequently given by  $\hat{\mathbf{x}}(t) = \mathbf{W}^T \mathbf{W} \mathbf{x}(t)$ . For  $L$  fixed,

PCA aims to find an optimal value (denoted by  $\tilde{\mathbf{W}}$ ) of the orthogonal matrix  $\mathbf{W}$ , such as to minimize the reconstruction error  $J = E\{\|\mathbf{x}(t) - \hat{\mathbf{x}}(t)\|^2\}$ . The rows of the transform matrix  $\tilde{\mathbf{W}}$  are the *principal components* (PCs) of the stochastic process  $\{\mathbf{x}(t)\}$ , which actually are the eigenvectors  $\mathbf{c}_j$ ,  $j = 1, 2, \dots, L$  of the input covariance matrix  $\mathbf{C}_x$ . The subspace spanned by the principal eigenvectors  $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_L\}$  with  $L < N$ , is called the PCA subspace (of dimensionality  $L$ ).

The stationary linear ICA problem [1], [9] can be formulated as given  $T$  realizations of  $\mathbf{x}(t)$ , estimate both the (constant) matrix  $\mathbf{A}$  and the corresponding realizations of  $\mathbf{s}(t)$ . In BSS the task is somewhat relaxed to finding the waveforms  $\{s_i(t)\}$  of the sources knowing only the (observed) mixtures  $\{x_j(t)\}$ .

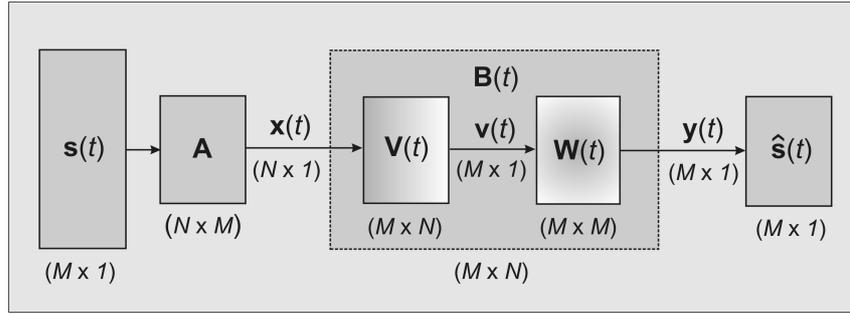


Fig. 2. – Linear mixing ( $\mathbf{A}$ ) of the latent source signals  $\mathbf{s}(t)$ , demixing by whitening ( $\mathbf{V}$ ) and rotation ( $\mathbf{W}$ ) of the observed data  $\mathbf{x}(t)$ .

There are several limitations to solving the ICA problem. If no suppositions are made about the noise, then the additive noise term is omitted from (1). A practical strategy is to include it in the signals as a supplementary term in the sum (1), thus the ICA model becomes:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) = \sum_{i=1}^M \mathbf{a}_i s_i(t) \quad (2)$$

Next limitation refers to the sizes of the vectors involved. The size of  $\mathbf{s}(t)$  (usually unknown) should not be greater than the size of data  $\mathbf{x}(t)$ , otherwise the problem becomes under-determined. If the size of  $\mathbf{x}(t)$  is greater than the size of  $\mathbf{s}(t)$  (i.e. there are more receptors than sources), the problem is over-determined and the extra data can be used for reducing the noise [3].

The blind source separation consists in updating a demixing matrix  $\mathbf{B}(t)$  of size  $M \times N$ , without resorting to any information about the spatial mixing matrix  $\mathbf{A}$ , so that the output vector  $\mathbf{y}(t) = \mathbf{B}(t)\mathbf{x}(t)$  becomes an estimate  $\mathbf{y}(t) = \hat{\mathbf{s}}(t)$  of the original independent

(latent) source signals  $\mathbf{s}(t)$ . Since ICA deals with higher-order statistics it is justified to normalize in some sense the first- and second-order moments. The effect is that the separating matrix  $\mathbf{B}(t)$  is divided in two parts: the first is dealing with dependencies in the first two moments – the *whitening* matrix  $\mathbf{V}(t)$ , and the second comprises the dependencies in higher-order statistics – the *orthogonal separating matrix*  $\mathbf{W}(t)$  in the whitened space (Fig. 2). If we assume zero-mean observed data  $\mathbf{x}(t)$ , then by whitening we get a vector  $\mathbf{v}(t) = \mathbf{V}(t)\mathbf{x}(t)$  with decorrelated components. The subsequent linear transform  $\mathbf{W}(t)$  seeks the solution by an adequate *rotation* in the space of component densities and yields  $\mathbf{y}(t) = \mathbf{W}(t)\mathbf{v}(t)$ . In the standard stationary case, the whitening and the orthogonal separating matrices converge to some constant values after a finite number of iterations, so that the *total separation matrix* becomes constant:

$$\mathbf{B}(t) = \mathbf{W}(t)\mathbf{V}(t) \rightarrow \mathbf{W}\mathbf{V} \quad (3)$$

Standard PCA is often used for whitening because information can be optimally compressed in the mean-square error sense and some possible noise is filtered out. The PCA whitening matrix can be expressed in the form:

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^T \quad (4)$$

where  $\mathbf{E}\mathbf{D}\mathbf{E}^T = E\{\mathbf{x}\mathbf{x}^T\}$  is the eigenvector decomposition of the covariance matrix of the (zero mean) data  $\mathbf{x}$ , implying that  $\mathbf{D} = \text{diag}[d_1, d_2, \dots, d_M]$  is a  $M \times M$  diagonal matrix containing the eigenvalues, and  $\mathbf{E} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_M]$  is an orthogonal  $N \times M$  matrix having the eigenvectors as columns.

The appropriate orthogonal transform  $\mathbf{W}$  can be sought invoking: (i) heuristic conditions for independence, (ii) optimizing some information-theoretic criterion, or (iii) optimizing some suitable contrast functions, in such a way as to yield independent outputs. Once the proper rotation was found, the best approximation of the (unknown) mixing matrix  $\mathbf{A}$  is given by  $\hat{\mathbf{A}} = \mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^{-1}$ , or simpler  $\hat{\mathbf{A}} = \mathbf{B}^{-1}$ , if the source signals equal the number of recorded data ( $N = M$ ). The columns of  $\hat{\mathbf{A}}$  yield the ICA basis vectors, whereas the rows of  $\mathbf{B}$  provide the filters (weight vectors) in the original, not whitened space. Finally, the estimated independent components (ICs) of the stationary noiseless linear ICA model at any sample index  $t$  are given by:

$$\hat{\mathbf{s}}(t) = \mathbf{y}(t) = \mathbf{B}\mathbf{x}(t) \quad (5)$$

The basis vectors  $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M)$  of the linear ICA model (1) are the counterparts of the principal eigenvectors  $(\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_L)$  in the PCA approach. Therefore, they are useful at least in similar applications, providing in most cases a more meaningful characterization of the data (e.g. in biomedical time series analysis).

### 3. EXPERIMENTAL

The artificially generated data, which consisted of 6 different time series of 512 samples each, were mixed by a square matrix  $\mathbf{A}$ , whose elements were randomly created and checked to avoid poor scaling and nonsingularity by having  $|\det(\mathbf{A})| > 1$ .

Table 1. – The analytical form of the original source signals.

<i>Original source signals</i>	
Modulated sinusoid:	$S(1) = 2 * \sin(t/149) * \cos(t/8) + 0.2 * \text{rand}(1, T)$
Square waves:	$S(2) = \text{sign}(\sin(12 * t + 9 * \cos(2/29))) + 0.1 * \text{rand}(1, T)$
Saw-tooth:	$S(3) = (\text{rem}(t, 79) - 17) / 23 + 0.1 * \text{rand}(1, T)$
Impulsive curve:	$S(4) = ((\text{rem}(t, 23) - 11) / 9)^5 + 0.1 * \text{rand}(1, T)$
Exponential decay:	$S(5) = 5 * \exp(-t/121) * \cos(37 * t) + 0.1 * \text{rand}(1, T)$
Spiky noise:	$S(6) = ((\text{rand}(1, T) < .5) * 2 - 1) * \log(\text{rand}(1, T))$

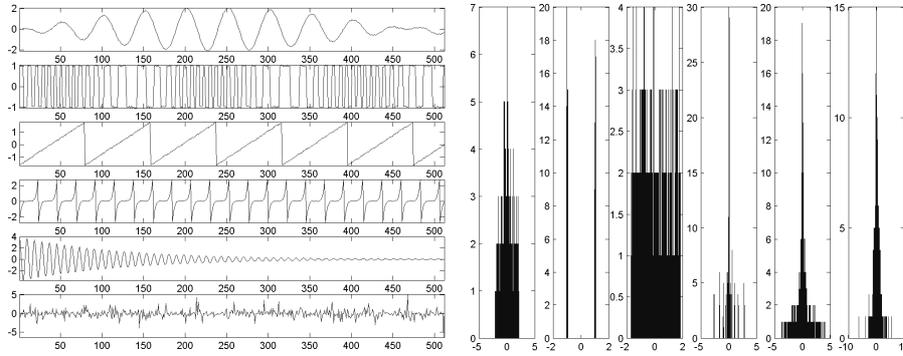


Fig.3. – Standardized source signals (left) and their histograms (right).

We specifically built up time series having both super-Gaussian and sub-Gaussian distributions (Fig. 3) in order to encompass as large a number of practical cases as possible [10]. All deterministic source signals were corrupted with zero-mean additive white noise in order to alleviate their pure deterministic character (Table 1).

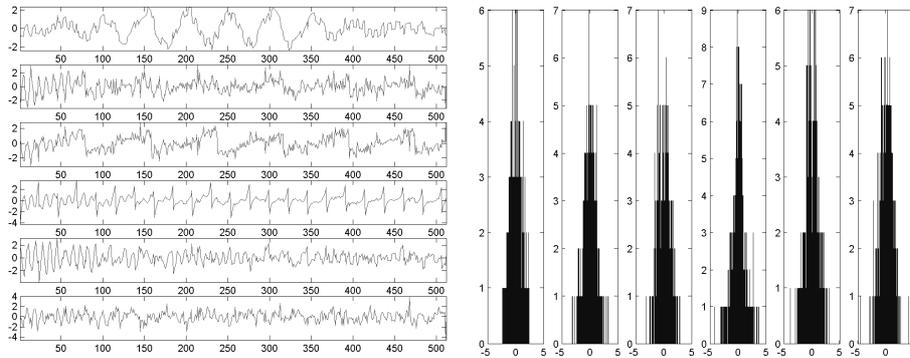


Fig. 4. – PCA eigenvectors (left) and their histograms (right).

The source mixtures were analyzed using both a neural implementation of a PCA singular value decomposition-based algorithm, and our own semi-adaptive batch-type implementation of a deflation fixed-point fast ICA algorithm proposed by Hyvärinen [11]. We based the choice of the optimal contrast (objective) function on the entropy and negentropy of the estimated ICs because of their ability to measure both the asymmetry and the bimodality/sparsity of the time series [12]. Essentially, the ranking of the contrast functions followed the strict monotonicity of entropy and negentropy as function of the Gaussian character of the estimated components, that is, the closer to a Gaussian distribution the probability density of a component was, the higher its entropy and the lower its negentropy [13]. The full dimension PCA eigenvector decomposition is presented in Fig. 4 and the ICA estimated independent components are shown in Fig. 5. Both figures include the entropy histograms of their corresponding components. Both PCA and ICA separated components were permuted in such a way that the best fit with the original source signals to be accomplished. One may easily note a high similarity between the source signals and the estimated independent components produced by ICA, whereas a clear discrepancy exists between the source signals and the principal components yielded by PCA. Basically, the point is that ICA is much better suited than the PCA to blind source separation, blind deconvolution, and equalization.

All codes were written in Matlab 6.5 and run on a PC with Pentium processor at 1.5 GHz. Full data processing was within seconds, heavily depending on the number of seeds used in the ICA algorithm for randomly generating the initial form of the orthogonal demixing (separating) matrix  $\mathbf{W}$ .

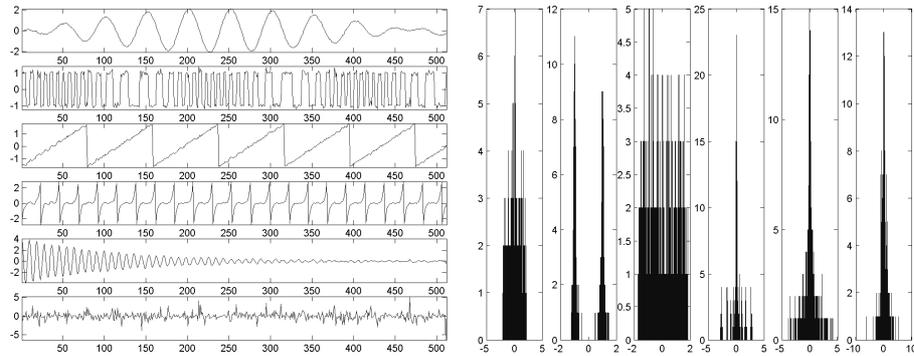


Fig 5. – ICA decomposition (left) and the histograms of the estimated ICs (right).

#### 4. RESULTS AND DISCUSSION

In order of a better comparison and for computational needs, the waveforms presented in Figs. 3,4, and 5 were normalized to zero mean and unit variance. Then the waveforms were paired in compliance with their maximum absolute correlation with respect to the original source signals, and their phases were reversed where necessary for best match. The matching process indicated a much better fit of the permuted ICs than of the PCs in terms of correlations with the source components (Fig. 6).

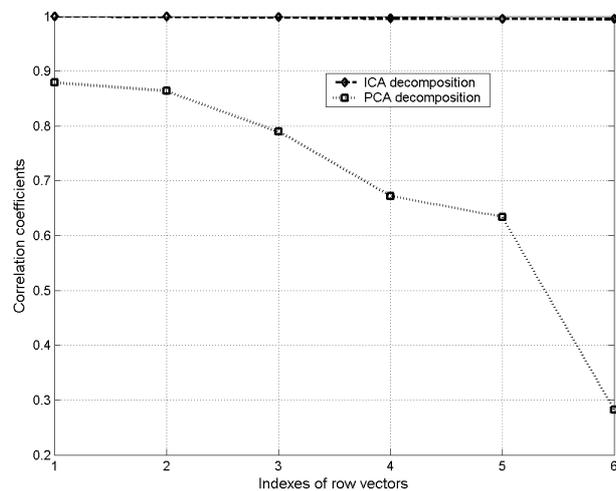


Fig. 6. – The correlation coefficients between the source components and the best matching PCs and ICs, respectively.

Figs. 3, 4, and 5 clearly put into evidence the similarity of the original source signals to the estimated ICs, and a relatively large dissimilarity to the PCs. In quantitative terms, we calculated the Euclidian distance between the PCA and ICA components relatively to the original source signals. The results, which are illustrated graphically in Fig. 7, fully support a much faithful separation in the case of ICA relative to PCA decomposition.

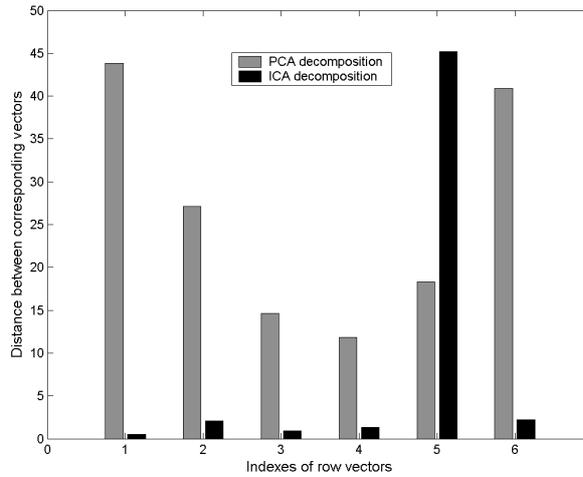


Fig. 7. – The Euclidian distance (arbitrary units) between the PCs and the estimated ICs with respect to the original source signals.

In practical cases when the source signals are unknown, ranking the separated components constitutes one reasonable criterion used for assessing the decomposition algorithms. Friedman [14] proposed a robust *structural* measure to arrange the vectors. The idea is to first sphere the data and then to map them into the interval  $[0, 1]$  with the gaussian cumulative density function  $\Phi(v)$ . For  $T$  realizations of a (scalar) random variable  $y_k(t)$  the proposed scheme leads to the index for the  $k$ -th component:

$$E_I(y_k) = \sum_{t=1}^T \left[ y_{\sigma_t} - \left( \frac{2t}{T} - 1 \right) \right]^2 \quad (6)$$

where  $\{\sigma_t\}$  are the indexes of the ordered  $\{y_k(t)\}$  in such a way that  $y_k(i) \leq y_k(j)$  iff  $\sigma_i \leq \sigma_j$ . The higher  $E_I(y_k)$  is, the more structural information contains the  $k$ -th estimated independent component. In Fig. 8 the structural information contained in the PCs and ICs, respectively, is presented comparatively with the corresponding original source components (signals). The structural information retrieved by the ICs closely follows its distribution among the source signals, whereas the structural information contained by the PCs strongly deviates from the original components.

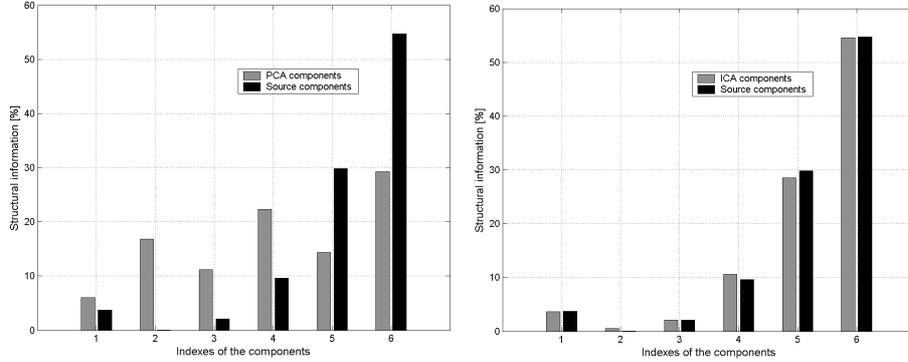


Fig. 8. – Comparison between structural information contained by the PCs (left) and the ICs (right) with respect to original structural information contained in the source distributions.

Another criterion in ranking the separated components is the amount of contribution of each component to the variance of the mixtures. In the case of PCA decomposition, the *contribution* of an eigenvector to the observed data is given by its corresponding eigenvalue. For ICA, the contribution of an estimated component  $y_k(t)$ ,  $k = 1, 2, \dots, M$  to the mixtures can be evaluated by the mean projected variance accounted for in the observable data set retrieved solely on the basis of this component  $\hat{\mathbf{x}}_0 = \hat{\mathbf{A}}\mathbf{y}_0$ , where  $\hat{\mathbf{A}}$  is the estimated mixing matrix (generally  $\hat{\mathbf{A}} = \mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^{-1}$  or  $\hat{\mathbf{A}} = \mathbf{B}^{-1}$ ) and  $\mathbf{y}_0$  has only one nonzero row corresponding to  $y_k(t)$ .

An alternative evaluation of the retrieval quality of the (latent) source signals is graphically presented in Fig. 9, which displays the variance accounted for in the mixtures by the PCs and the mean projected variances of the ICs, as compared with the variance of the observed data. The source signals were standardized to zero mean and unit variance prior to mixing. Consequently, a perfect retrieval of the (latent) source signals out of the mixtures should lead to estimates with constant variance. As expected, the plots clearly indicate that the first PCs are oriented along the mixtures with maximum variance and rapidly decrease with the increasing of PC index. Contrarily, the variances of the ICs are much slower decreasing, so that the estimated ICs provide a more realistic (blind) decomposition of the original (latent) source signals.

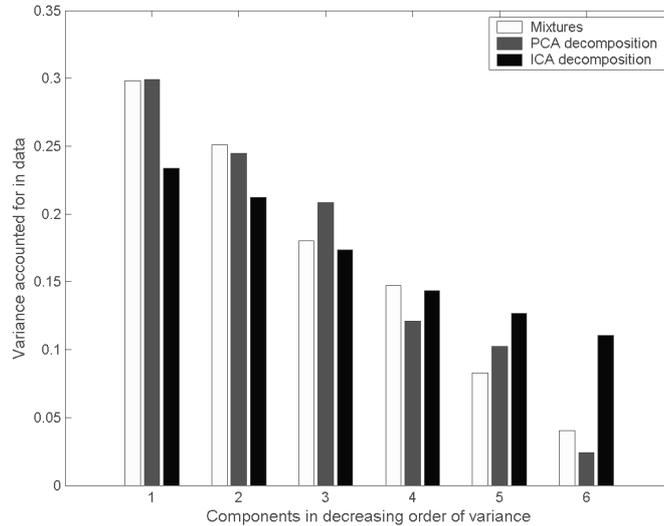


Fig. 9. – Comparison among the variances accounted for in the mixtures by the PCs and ICs, respectively, with respect to the true variance distribution in the source components.

## 5. CONCLUSION

PCA is a way of encoding second-order dependencies in the data by rotating the axis to correspond to the directions of maximum covariance. PCA decorrelates the input data, but does not address the high-order dependencies. *Decorrelation* means that variables cannot be predicted from each other using a *linear* predictor, though *nonlinear* dependencies may still exist. ICA is a generalization of PCA that separates the higher-order dependencies in the input, in addition to second order dependencies. ICA does not constrain the axes to be orthogonal as PCA, rather it attempts to place them in the directions of maximal statistical dependencies in the data. Each estimate in ICA attempts to encode a portion of the dependencies in the input, so that the dependencies are removed from between the output components. The projection of the distributions on the ICA axes would have less overlap and the output distribution of the estimates would be kurtotic [15]. Therefore ICA, contrarily to other statistical methods searching for underlying factors from multivariate data, looks for components that are simultaneously *independent* and *nongaussian*.

Both PCA and ICA share some common features like: (i) aiming at building generative models that are likely to have produced the observed data (in a neuromorphic approach, the model parameters are treated as network weights in unsupervised learning) and (ii) performing information preservation and redundancy reduction. PCA and ICA have major applications in data mining and exploratory data analysis, such as signal characterization, optimal feature extraction, and data compression, as well as the basis of subspace classifiers in pattern recognition.

In many cases the goal is to determine a subset of independent components containing as much *interesting* structural information as possible. The nongaussianity of a distribution is perceived as a measure of its *interestingness* [14], [16], [17]. Why should be a *sparse*, or more general, a supergaussian distribution interesting? Friston [18] gives a simple and extremely compelling answer: because measurements of biological systems receive contributions from many sources (e.g. dipoles generated by neural activity), the observations usually represent a roughly linear mixture of interesting events. By the Central Limit Theorem in statistics, this mixture conforms to a Gaussian distribution. But mixtures themselves are rather uninteresting, so that the only interesting things must be non-Gaussian (assuming that Gaussian distributions arise only from mixing). This perspective of ICA motivates why ICA is so pertinent to biological time-series and appealing to data analysis in imaging neuroscience.

The classical measure of non-Gaussianity is *kurtosis* or the fourth order cumulant since most non-Gaussian random variables have nonzero kurtosis. Kurtosis exhibits nevertheless some shortcomings in practice, particularly when its value has to be estimated from a measured sample: (i) finite-sample estimators of higher-order cumulants are highly sensitive to outliers [16], (ii) perfectly estimated cumulants measure mainly the tail of the distributions and are largely unaffected by structure around the center of distributions [14], and (iii) as noted by Marriot quoted in Jones and Sibson [17], clustered projections that are approximately symmetric and mesokurtic can sometimes be difficult to identify with indices based on third and fourth moments. A natural criterion for non-Gaussianity is (differential) entropy [16], [17], which attains its maximum (for constant variance) when the distribution is Gaussian and all other distributions have smaller entropies. However, the estimation of differential entropy and, consequently of negentropy, proved to be too sensitive to noise and to the tails of the distributions of independent components, whereas the interesting structure lies mainly in the central part of the distribution.

The principle of maximum non-Gaussianity shows the very close connection between ICA and an independently developed technique called *projection pursuit* (PS). In PS, we are actually looking for maximally non-Gaussian linear combinations, which are used for visualization and other purposes. Thus, the independent components can be interpreted as projection pursuit directions.

When ICA is used to extract features, the principle of maximum non-Gaussianity reflects the connection to *sparse coding* that has been used in neuroscientific theories of feature extraction [19]. The idea in sparse coding is to represent data with components so that only a small number of them are *active* at the same time. It turns out that this is equivalent, in some situations, to finding components that are maximally non-Gaussian. PS and sparse coding connections are related to a fundamental result saying that ICA gives a linear representation that is *as structured as possible*. This statement can be given a rigorous meaning by information-theoretic concepts formulated by Bell and Sejnowski [2], and shows that the ICs are in many ways easier to process than the original random variables. In particular, ICs are easier to code (compress) than the original source signals.

In practice, the independence of components achieved by ICA is only approximate, which raises the question of utility and interpretation of the decomposition yielded by ICA. The estimates provided by ICA are useful even when they are far from mutual independency because ICA also serves some other useful purposes than dependence reduction, such as projection pursuit and sparse coding. Furthermore, models may be formulated in which the assumption of independence is explicitly relaxed, like in two recently developed methods, namely *independent subspace analysis* (ISA) and *topographic ICA* (TICA) [20].

Since PCA and ICA have different goals, they naturally yield quite different results in nearly all cases. In PCA the data are represented in an orthonormal basis determined merely by the second order statistics (covariances) of the data matrix. Such representation is adequate for Gaussian data only. However, non-Gaussian data may contain a lot of additional information in its higher order statistics. Contrarily, ICA captures all statistical dependencies in data. The flexibility of ICA approach by incorporating higher-order statistical information resides in transforming the PCA ill-posed problem associated with decorrelated decompositions into a well-posed problem of independent decompositions, that is, ICA avoids the non-uniqueness associated with PCA. By all means, when reasons are to believe that the latent source signals are independent, their estimation by ICA techniques is much more faithful than the PCA counterpart and, consequently, ICA decomposition is recommended for realistic blind source separation, signal deconvolution and equalization on the basis of minimum statistical prior assumptions.

*Acknowledgments:* R.M. is supported by a postdoc grant from the European Community, FP5 (QLG3-CT-2000-30161). M.M.V.H. is supported by research grants received from the Fund for Scientific Research (G.0185.96N), the National Lottery (Belgium) (9.0185.96), the Flemish Regional Ministry of Education (Belgium) (GOA 95/99-06; 2000/11), the Flemish Ministry for Science and Technology (VIS/98/012), and the European Community, FP5 (QLG3-CT-2000-30161 and IST-2001-32114).

#### REFERENCES

1. Comon, P., Independent component analysis, A new concept? *Signal Processing* **36**, 3 (1994) pp. 287-314.
2. Bell, A., and Sejnowski, T., An information-maximization approach to blind separation and blind deconvolution, *Neural Computation*, **7** (1995) pp. 1129-1159.
3. Barlow, H.B., Possible principles underlying the transformation of sensory messages. In Rosenblith, W.A. (ed.): *Sensory Communication*. MIT Press, Cambridge, MA (1961) pp. 217-234.
4. Olshausen, B.A., Field, D.J., Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381** (1996) pp. 607-609.
5. Atick, J.J (1992) could information theory provide an ecological theory of sensory processing? *Network* **3** (1992) pp. 213-251.
6. Jung, T.-P., Makeig, S., McKeown, M.J., Bell, A.J., Lee, T.-W., Sejnowski, T.J., Imaging brain dynamics using independent component analysis. *Proc. IEEE* **89**, 7 (2001) pp. 1107-1122.
7. Vigário, R., Sarela, J., Jousmäki, V., Hämäläinen, M., Oja, E., Independent component approaches to the analysis of EEG and MEG recordings, *IEEE Trans. Biomed. Eng.* **47** (2000) pp. 589-593.
8. Pekar, J.J., Calhoun, V., Adali, T., Pearlson, G.D., "Spatial and temporal independent component analysis of fMRI data with two task-related waveforms." *Proc. Intl. Soc. Mag. Reson. Med.* **9** (2001) p. 24.

9. Cardoso, J.-F., Blind signal separation: Statistical principles. *Proc. IEEE* **9**, 10 (1998) pp. 2009-2025.
10. Mutihac, R., van Hulle, M.M., *A Comparative Survey on Adaptive Neural Network Algorithms for Independent Component Analysis*, to appear in *Romanian Reports in Physics*, Vol. **55**, Nos. 1-2, 2003.
11. Hyvärinen, A., Oja, E., A fast fixed-point algorithm for independent component analysis. *Neural Computation* **9** (1997) pp. 1483-1492 (code at <http://www.cis.hut.fi/projects/ica/fastica/>).
12. Hyvärinen, A., New approximations of differential entropy for ICA and projection pursuit. In *Advances in Neural Information Processing Systems (NIPS'97)* MIT Press, 1998, pp.273-279.
13. Mutihac, R., van Hulle, M.M., Neural network implementations of independent component analysis. *Proc. 2002 IEEE Workshop on Neural Networks for Signal Processing XII (NNSP 2002)*, Boulard, H., Adali, T., Bengio S., Larsen, J., Douglas, S. (Eds.) Martigny, Switzerland, 2002, pp. 505-515.
14. Friedman, J.H., Exploratory projection pursuit. *J. of the American Statistical Association* **82**, 397 (1987) pp. 249-266.
15. Field, D.J., What is the goal of sensory coding? *Neural Comput.* **6** (1994) pp. 559-601.
16. Huber, P.J., Projection pursuit. *The Annals of Statistics* **13**, 2 (1985) pp. 435-475.
17. Jones, M.C., Sibson R., What is projection pursuit? *J. of the Royal Statistical Society, ser. A* **150** (1987) pp.1-36.
18. Friston, K.J., Modes or models: A critique on independent component analysis for fMRI. *Trends in Cognitive Sciences* **2**, 10 (1998) pp. 373-375.
19. Olshausen, B.A., Field, D.J. (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381** (1996) pp. 607-609.
20. Hyvärinen, A., beyond independent components. *Proc. Int. Conf on Artificial Neural Networks*, Edinburgh, UK, 1999, pp. 809-814.