A FAST ALGORITHM FOR ESTIMATING OVERCOMPLETE ICA BASES FOR IMAGE WINDOWS

Aapo Hyvärinen, Razvan Cristescu and Erkki Oja

Helsinki University of Technology
Laboratory of Computer and Information Science
P.O. Box 5400, FIN-02015 Espoo, Finland
`aapo.hyvarinen@hut.fi`
`http://www.cis.hut.fi/projects/ica/`

ABSTRACT

We introduce a very fast method for estimating overcomplete bases of independent components from image data. This is based on the concept of quasi-orthogonality, which means that in a very high-dimensional space, there can be a large, overcomplete set of vectors that are almost orthogonal to each other. Thus we may estimate an overcomplete basis by using one-unit ICA algorithms and forcing only partial decorrelation between the different independent components. The method can be implemented using a modification of the FastICA algorithm, which leads to a computationally highly efficient method.

1. INTRODUCTION

Indendent component analysis (ICA) [4, 10] is a statistical model where the observed data is expressed as a linear transformation of latent variables that are nongaussian and mutually independent. We may express the model as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{x} = (x_1, x_2, ..., x_m)$ is the vector of observed random variables, $\mathbf{s} = (s_1, s_2, ..., s_n)$ is the vector of the latent variables called the independent components or source signals, and $\mathbf{A}$ is an unknown constant matrix, called the mixing matrix.

In the classic case, we assume that the number of independent components equals the number of the observed variables, i.e. $n = m$. Exact conditions for the identifiability of the model were given in [4], and several methods for estimation of the classic ICA model have been proposed in the literature [1, 2, 3, 4, 7, 6].

Recently, a non-classic modification of the model, where it is assumed that $n > m$, has attracted the attention of a number of researchers [12, 13, 14, 15]. Such a model could be especially interesting when using ICA for feature extraction of image or other sensorial data. This is because it leads to decomposition of image windows that is closely related to overcomplete wavelet bases (see [14]), which are in certain ways superior to ordinary wavelet basis. Basically, the larger number of independent components in the model means that we have a larger 'dictionary' from which to construct the representation. The dictionary consists of the basis vectors that are given as columns $\mathbf{a}_i$ of the mixing matrix $\mathbf{A}$. Using an overcomplete basis may also allow for some invariances (e.g. with respect to translation) in the representation.

Some methods have already been proposed for estimating the mixing matrix in the ICA model with $n > m$ [12, 13, 14, 15]; this is often called the problem of estimating an overcomplete ICA basis. A drawback with all the proposed methods is that they are computationally very demanding. This is basically because the model then becomes a model with missing data: the computation of the likelihood is not straightforward as in the classic case. In fact, the evaluation of the likelihood contains an integral and even reasonable approximations of that integral are hard to compute [13]. On the other hand, since these method are usually applied to data of very high dimensions, it would be very useful to have an estimation method that can cope with very large dimensions in a moderate computational load.

In this paper, we propose a computationally very efficient method for estimating overcomplete ICA bases from image data. The method is a rather heuristic one, and utilizes the concept of quasi-orthogonality. Experiments with image data show that the obtained basis vectors are at least qualitatively similar to those obtained by statistically more principled estimation methods [12, 14]. The computational complexity of the method is essentially of the same order as that of the estimation of the classic ICA model. In fact, the resulting algorithm is a simple modification of the FastICA algorithm [7, 6].

2. QUASI-UNCORRELATED SPARSE DECOMPOSITIONS

Instead of using maximum likelihood estimation, we take here a more heuristic approach which is based on sparse coding. Our heuristic approach is justified by the fact that for image data, the ICA model is only a rather coarse approximation. In particular, the number of potential "independent components" seems to be infinite: The set of such components is closer to a continuous manifold that a discrete set. One evidence for this is that classic ICA estimation methods give different basis vectors when started with different initial values, and the number of components thus produced does not seem to be limited. Any classic ICA estimation method gives a rather arbitrary collection of components which are somewhat independent, and have sparse marginal distributions [9].

We can also assume, for simplicity, that the data is prewhitened as a preprocessing step. The utility of such a preprocessing is that in the classic case, the mixing matrix can be considered to be orthogonal, and the independent components are simply given by the dot-products of the data vector with the basis vectors.

Due to the above considerations, we assume in our approach that what is usually needed, is a collection of basis vectors which has the following two properties.

1. The dot-products of the observed data with the basis vectors have sparse (super-Gaussian) marginal distributions.

2. The dot-products should be approximately uncorrelated ('quasi-uncorrelated')

A decomposition with these two properties seems to capture the essential properties of the decomposition obtained by estimation of the ICA model. Such decompositions could be called 'quasi-uncorrelated sparse decompositions'.

It is clear that it is possible to find very overcomplete basis sets that have the first property of these two. Classic ICA estimation is usually based on maximizing the sparseness (or, in general, nongaussianity) of the dot-products, so the existence of several different classic ICA decompositions for a given image data set shows the existence of decompositions with the first property.

What is not obvious, however, is that it is possible to find strongly overcomplete decompositions such that the dot-products are approximately uncorrelated. The main point here is that this is possible because of the phenomenon of quasi-orthogonality.

Quasi-orthogonality [11] is a somewhat counter-intuitive phenomenon encountered in very high-dimensional spaces. In a certain sense, there is much more room for vectors in high-dimensional spaces. The point is that in an n-dimensional space, where n is large, it is possible to have (say) $2n$ vectors that are practically orthogonal, i.e. their angles are close to 90 degrees. In fact, when n grows, the angles can be made arbitrarily close to 90 degrees. This must be contrasted with small-dimensional spaces: If, for example, $n = 2$, the even the maximally separated $2n = 4$ vectors exhibit angles of 45 degrees. In image decomposition, we are usually dealing with spaces whose dimensions are of the order of 100. Therefore, we can easily find decompositions of, say, 400 basis vectors, such that the vectors are quite orthogonal, with practically all the angles between basis vectors staying above 80 degrees. Recall that orthogonality in the whitened space is equivalent to uncorrelation of the dot-products.

3. THE ALGORITHM

To obtain a quasi-uncorrelated sparse decomposition as defined in Sec. 2, we need two things. First, a method for finding vectors that have maximally sparse dot-products, and second, a method of quasidecorrelating such vectors. Actually, most classic ICA algorithms can be considered as maximizing the nongaussianity of the dot-products with the basis vectors, provided that the data is prewhitened. This is especially true for the one-unit algorithms as in [5, 8, 7, 6]. Thus the main problem here is constructing a proper method for quasi-decorrelation.

We have developed two methods for quasi-decorrelation: one on them is symmetric and the other one is deflationary. This dichotomy is the same as in ordinary decorrelation methods used in ICA [7, 8, 6]. As above, it is here assumed that the data is whitened.

A simple way of achieving quasi-decorrelation is to modify the ordinary deflation scheme based on a Gram-Schmidt-like decorrelation [6]. This means that we estimate the basis vectors one by one. When we have estimated p basis vectors $\mathbf{a}_1, ..., \mathbf{a}_p$, we run the one-unit fixed-point algorithm for $\mathbf{a}_{p+1}$, and after every iteration step subtract from $\mathbf{a}_{p+1}$ a certain proportion of the 'projections' $\mathbf{a}_{p+1}^T \mathbf{a}_j \mathbf{a}_j, j = 1, ..., p$ of the previously estimated p vectors, and then renormalize $\mathbf{a}_{p+1}$:

$$\begin{aligned} &1.\ \text{Let } \mathbf{a}_{p+1} = \mathbf{a}_{p+1} - \alpha \textstyle\sum_{j=1}^{p} \mathbf{a}_{p+1}^T \mathbf{a}_j \mathbf{a}_j \\ &2.\ \text{Let } \mathbf{a}_{p+1} = \mathbf{a}_{p+1} / \|\mathbf{a}_{p+1}\| \end{aligned} \quad (2)$$

where α is a constant determining the force of the quasi-decorrelation. If $\alpha = 1$, we have ordinary decorrelation. We have found in our experiments that an α in the range [0.1 ... 0.3] is sufficient in spaces where the dimension is 64.

In certain applications it may be desired to use a symmetric quasi-decorrelation, in which no vectors are 'privileged' over others [7, 6]. This can be accomplished, e.g., by the following formula:

$$\begin{aligned} &1.\ \text{Let } \mathbf{A} = \tfrac{3}{2}\mathbf{A} - \tfrac{1}{2}\mathbf{A}\mathbf{A}^T\mathbf{A}. \\ &2.\ \text{Normalize each column of } \mathbf{A} \text{ to unit norm.} \end{aligned} \quad (3)$$

which is closely related to the symmetric decorrelation method in [6].

Thus, the algorithm that we propose is similar to the FastICA algorithm [7, 6] in all other respects than the decorrelation, which is replaced by one of the quasi-decorrelation method above.

4. EXPERIMENTS

We applied our algorithm on image windows (patches) of 8×8 pixels taken from natural images. The mean of the image window (DC component) was removed as a preprocessing step, so the dimension of the data was 63. Both deflationary and symmetric quasi-decorrelation were used. The nonlinearity used in the FastICA algorithm was the hyperbolic tangent. Figure 1 shows an estimated approximately 2 times overcomplete basis (with 120 independent components). The sample size was 12000. Figure 2 shows an estimated approximately 4 times overcomplete basis (with 240 components). The sample size was 14000. The results shown here were obtained using the symmetric approach; the deflationary approach yielded similar results, with the parameter α fixed at 0.3 and 0.1 in for 2 and 4 times overcomplete bases, respectively.

The results show that the estimated basis vectors are qualitatively quite similar to those obtained by other methods [12]. This validates our heuristic approach.

To make sure that the basis vectors estimated are unique, i.e. that the algorithms didn't estimate several times the same components due to failure of the quasi-decorrelation methods, we computed the angles between all possible couples of basis vectors in a given overcomplete basis (in the whitened space). Figure 3 a) shows the histogram for the basis in Figure 1, and Figure 3 b) shows the histogram for the basis in Figure 2. One can see that the practically all the angles are quite large, i.e. near orthogonality or uncorrelation. Duplicates would show up in these results as degrees that are near to zero. No such angles were found, in fact less that 0.2 % of the angles were smaller than 30 degrees. This shows that the 'soft' symmetric orthogonalization method introduced above works correctly, forcing quasi-decorrelation. For the deflationary approach, essentially identical results were obtained.

5. CONCLUSION

We introduced a new algorithm for estimating overcomplete bases from image and other sensorial data. The algorithm is based a modification of the FastICA algorithm and gives a basis for quasi-decorrelating sparse decomposition that can be considered as an approximate ICA decomposition. The algorithm is computationally very efficient, reducing the complexity of overcomplete basis estimation to that of classic ICA estimation.

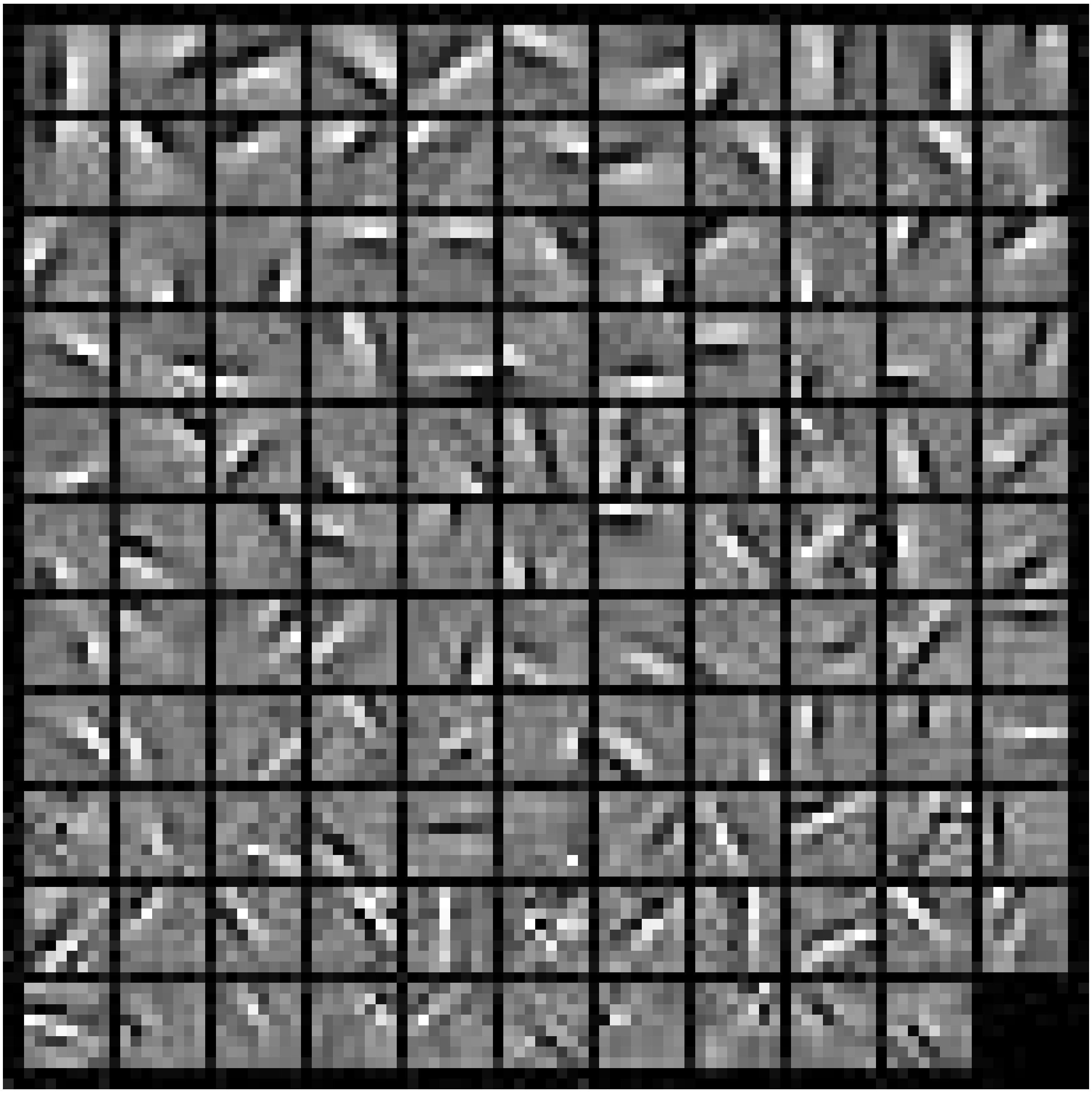

Figure 1: The basis vectors of an 2 times overcomplete basis. The dimension of the data is 63 (excluding the DC component, i.e. the local mean) and the number of basis vectors is 120. The results are shown in the original space, i.e. the inverse of the preprocessing (whitening) was performed. The symmetric approach was used. The basis vectors are very similar to Gabor functions or wavelets, as in the classic (not overcomplete) case.

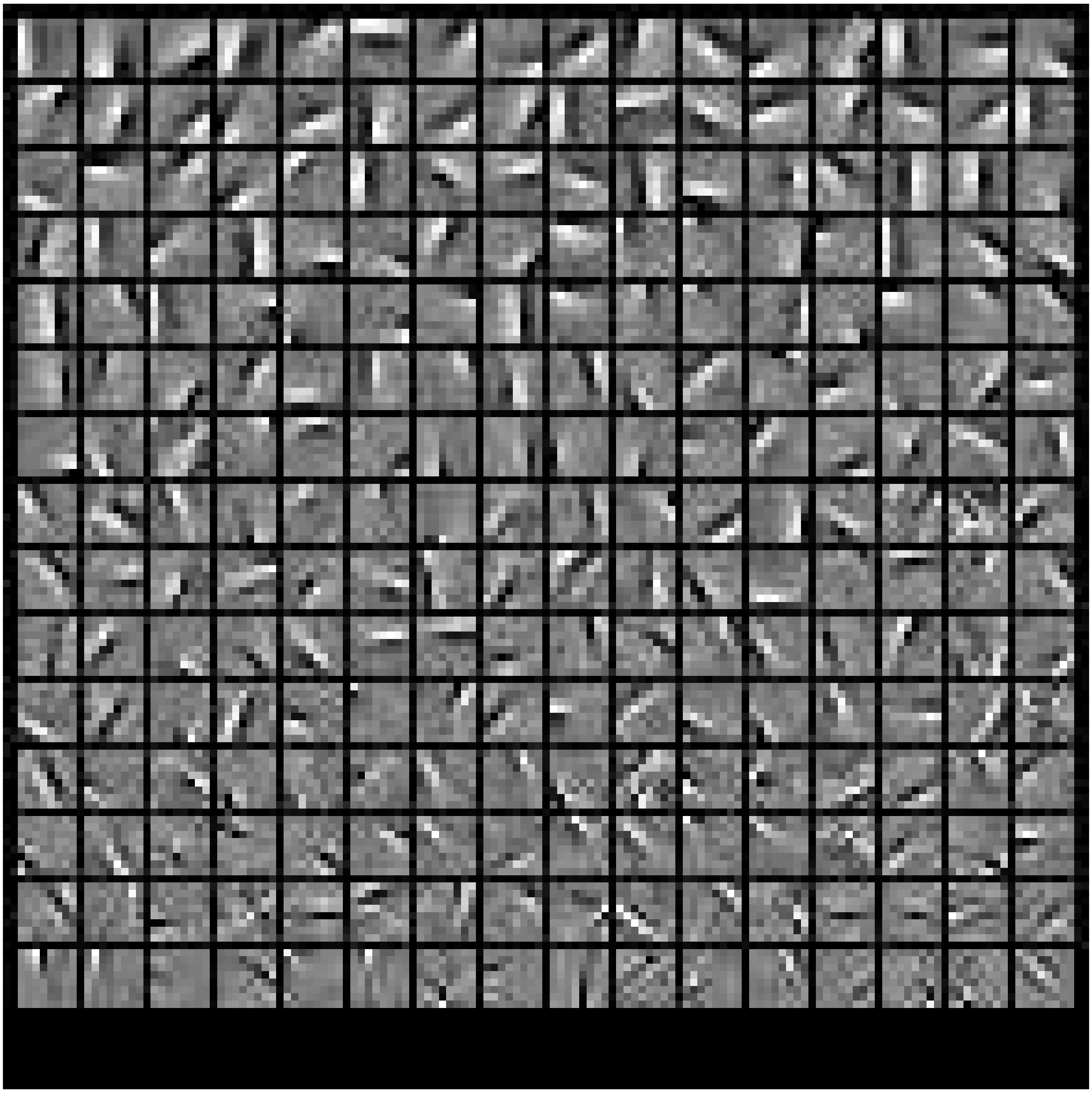

Figure 2: The basis vectors of an 4 times overcomplete basis. The dimension of the data is 63 (excluding the DC component, i.e. the local mean) and the number of basis vectors is 240. The results are shown in the original space, i.e. the inverse of the preprocessing (whitening) was performed. The symmetric approach was used. Again, the basis vectors are very similar to Gabor functions or wavelets.

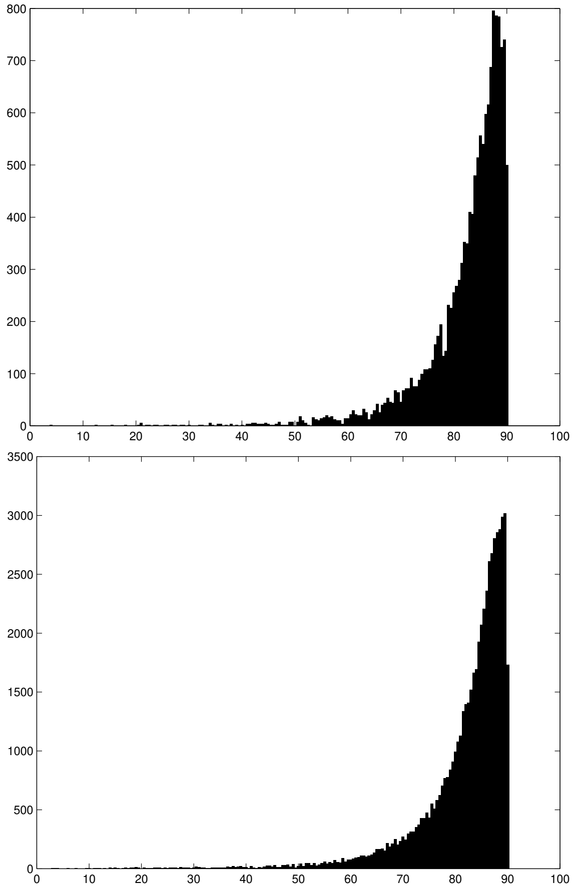

Figure 3: The histograms of the angles between all possible couples of the basis vectors. The upper histogram is for the basis in Figure 1, and the lower one is for the basis in Figure 2. Practically all the angles are above 70 degrees, which shows that the quasi-decorrelation methods were effective.

6. REFERENCES

[1] S. Amari, A. Cichocki, and H.H. Yang. A new learning algorithm for blind source separation. In *Advances in Neural Information Processing Systems 8*, pages 757–763. MIT Press, Cambridge, MA, 1996.

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

[3] J.-F. Cardoso and B. Hvam Laheld. Equivariant adaptive source separation. *IEEE Trans. on Signal Processing*, 44(12):3017–3030, 1996.

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

[5] N. Delfosse and P. Loubaton. Adaptive blind separation of independent sources: a deflation approach. *Signal Processing*, 45:59–83, 1995.

[6] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 1999. To appear.

[7] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[8] A. Hyvärinen and E. Oja. Independent component analysis by general nonlinear Hebbian-like learning rules. *Signal Processing*, 64(3):301–313, 1998.

[9] A. Hyvärinen, E. Oja, P. Hoyer, and J. Hurri. Image feature extraction by sparse coding and independent component analysis. In *Proc. Int. Conf. on Pattern Recognition (ICPR'98)*, pages 1268–1273, Brisbane, Australia, 1998.

[10] C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[11] T. Kohonen. *Self-Organizing Maps.* Springer-Verlag, Berlin, Heidelberg, New York, 1995.

[12] M. Lewicki and B. Olshausen. Inferring sparse, overcomplete image codes using an efficient coding framework. In *Advances in Neural Information Processing Systems 10 (Proc. NIPS*97)*, pages 815–821. MIT Press, 1998.

[13] M. Lewicki and T. J. Sejnowski. Learing overcomplete representations. In *Advances in Neural Information Processing Systems 10 (Proc. NIPS*97)*, pages 556–562. MIT Press, 1998.

[14] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325, 1997.

[15] P. Pajunen. Blind separation of binary sources with less sensors than sources. In *Proc. ICNN'97*, Houston, Texas, 1997.