

FAST EQUIVARIANT JADE

Jari Miettinen* Klaus Nordhausen† Hannu Oja‡ Sara Taskinen*

* Department of Mathematics and Statistics, 40014 University of Jyväskylä, Finland

† School of Information Sciences, 33014 University of Tampere, Finland

‡ School of Health Sciences, 33014 University of Tampere, Finland

ABSTRACT

Independent component analysis (ICA) is a widely used signal processing tool having applications in various fields of science. In this paper we focus on affine equivariant ICA methods. Two such well-established estimation methods, FOBI and JADE, diagonalize certain fourth order cumulant matrices to extract the independent components. FOBI uses one cumulant matrix only, and is therefore computationally very fast. However, it is not able to separate identically distributed components which is a major drawback. JADE overcomes this restriction. Unfortunately, JADE uses a huge number of cumulant matrices and is computationally very heavy in high-dimensional cases. In this paper, we hybridize these two methods. The affine equivariant FOBI estimate is used as an initial value for JADE, and only a small subset of most informative cumulant matrices is then diagonalized. In simulation studies we show that the new affine equivariant estimate is almost as good as JADE, and it is computationally much faster.

Index Terms— FOBI, Independent component analysis, Minimum distance index, SHIBBS

1. INTRODUCTION

The basic independent component (IC) model assumes that the observed p -variate random vector $\mathbf{x} = (x_1, \dots, x_p)'$ is generated by

$$\mathbf{x} = A\mathbf{s}, \quad (1)$$

where A is a full rank $p \times p$ mixing matrix and $\mathbf{s} = (s_1, \dots, s_p)'$ is the unobserved random vector of p mutually independent sources. We also assume that

$$(A1) \quad E(s_i) = 0 \text{ and } E(s_i^2) = 1, \text{ for } i = 1, \dots, p.$$

(A2) At most one of the source components is gaussian.

In independent component analysis (ICA) the aim is to use observations $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ from the distribution of \mathbf{x} to find an estimate \hat{W} of an unmixing matrix W such that $W\mathbf{x}$ has independent components. Recall that A and \mathbf{s} in model (1) are confounded in the sense that the order and

signs of the components of \mathbf{s} (and the columns of A) are not uniquely defined. Write

$$\mathcal{C} = \{C : \text{each row and column of } C \text{ has exactly one non-zero element.}\}. \quad (2)$$

Then, if W is an unmixing matrix in model (1), so is CW for any $C \in \mathcal{C}$. We say that W_1 and W_2 are equivalent if $W_1 = CW_2$ for some $C \in \mathcal{C}$, and write $W_1 \sim W_2$.

Let $W(F_{\mathbf{x}})$ be the value of an unmixing matrix (functional) at the distribution $F_{\mathbf{x}}$ of \mathbf{x} . In independent component analysis it is usually required that the separation result $\mathbf{z} = W(F_{\mathbf{x}})\mathbf{x}$ does not depend on the mixing matrix A . This is formalized in the following definition. See also [1].

Definition 1 Let $F_{\mathbf{x}}$ denote the cdf of \mathbf{x} . The functional $W(F_{\mathbf{x}})$ is an IC functional if (i) $W(F_{\mathbf{s}}) \sim I_p$ for any \mathbf{s} with independent components and at most one gaussian component, and (ii) $W(F_{\mathbf{x}})$ is affine equivariant in the sense that $W(F_{B\mathbf{x}}) \sim W(F_{\mathbf{x}})B^{-1}$ for all full rank $p \times p$ matrices B .

The corresponding estimator $\hat{W} = W(X)$ is obtained when the IC functional is applied to the empirical distribution function of $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)$. Naturally, the estimator is then also affine equivariant in the sense that $\hat{W}(BX) \sim \hat{W}(X)B^{-1}$.

One of the first methods to solve the ICA problem was the so called FOBI (fourth order blind identification) estimate, see [2]. The affine equivariant FOBI functional diagonalizes both the covariance matrix and a certain matrix of fourth moments. It was noticed later ([3]) that a better performance is obtained if a large number of fourth moments matrices are simultaneously used in the estimation. In JADE, p^2 fourth moment matrices are selected so that the procedure is affine equivariant in the sense of Definition 1. The drawback of the JADE (joint approximate diagonalization of eigenmatrices) method is that, due to a huge number of matrices to be diagonalized, it cannot be applied to high-dimensional data sets.

In this paper, we propose a simple affine equivariant IC functional that combines FOBI and JADE. The structure of the paper is the following. In Section 2 we first recall the classical fourth moment based methods and an earlier approach to

speed up JADE in high-dimensional cases. In Section 3 a new estimator hybridizing FOBI and JADE is proposed. The efficiencies and computation times of different estimators are compared in simulation studies in Section 4.

Throughout the paper, we use the following notation. For a $p \times p$ matrix A , $\text{diag}(A)$ is a diagonal matrix with the same diagonal elements as A and $\text{off}(A) = A - \text{diag}(A)$. $\|A\| = \sqrt{\sum_{i=1}^p \sum_{j=1}^p A_{ij}^2}$ is the matrix (Frobenius) norm of A . Also, e_i is a p -vector with i th element one and other elements zero, $i = 1, \dots, p$, and $E_{ij} = e_i e_j^T$, $i, j = 1, \dots, p$.

2. SOME IC FUNCTIONALS BASED ON FOURTH ORDER CUMULANTS

For notational simplicity, we assume that, in the model (1), $E(\mathbf{x}) = E(\mathbf{s}) = 0$. Write $\text{Cov}(\mathbf{x}) = E(\mathbf{x}\mathbf{x}^T)$ for the covariance matrix, and $\mathbf{z} = \text{Cov}(\mathbf{x})^{-1/2}\mathbf{x}$ for the whitened random vector. Consider first the $p \times p$ matrix of fourth moments

$$\text{Cov}_4(\mathbf{x}) = E[(\mathbf{x}^T \mathbf{x})\mathbf{x}\mathbf{x}^T].$$

We then have the following definition, see [2].

Definition 2 *The FOBI functional is defined as follows.*

1. Whiten the data: $\mathbf{z} = \text{Cov}(\mathbf{x})^{-1/2}\mathbf{x}$.
2. Find an orthogonal matrix U to minimize

$$D(U) = \|\text{off}(U\text{Cov}_4(\mathbf{z})U^T)\|^2.$$

3. The FOBI functional is then $W = U\text{Cov}(\mathbf{x})^{-1/2}$.

When does the FOBI functional find the latent independent components s_1, \dots, s_p ? First note that

$$\text{Cov}_4(\mathbf{s}) = \sum_{i=1}^p (\beta_{2,i} + p + 2)E_{ii}$$

where $\beta_{2,i} = E(s_i^4) - 3$. Moreover Cov_4 is equivariant under orthogonal transformations so that

$$\text{Cov}_4(U\mathbf{s}) = U\text{Cov}_4(\mathbf{s})U^T$$

for all orthogonal matrices U . As $\mathbf{z} = V\mathbf{s}$ for some orthogonal V , the FOBI functional is an IC functional if all kurtosis values $\beta_{2,i}$ are distinct. However, the equality of some kurtosis values has no effect on the separation of the components with distinct kurtosis values. As the second step simply finds the matrix of eigenvectors of $\text{Cov}_4(\mathbf{z})$, FOBI is computationally highly efficient. However, the inability to separate components with identical kurtosis values, e.g. iid components, is a major drawback.

JADE [3] overcomes the restrictions of FOBI. In JADE, one considers p^2 fourth moment matrices

$$C_{ij}(\mathbf{x}) = E[(\mathbf{z}^T E_{ij} \mathbf{z})\mathbf{z}\mathbf{z}^T] - E_{ij} - E_{ij}^T - \text{tr}(E_{ij})I_p,$$

$i, j = 1, \dots, p$. Note that FOBI uses only

$$\text{Cov}_4(\mathbf{x}) = \sum_{i=1}^p C_{ii}(\mathbf{x}) + (p+2)I_p.$$

We then have the following definition, see [3].

Definition 3 *The JADE functional is defined as follows.*

1. Whiten the data: $\mathbf{z} = \text{Cov}(\mathbf{x})^{-1/2}\mathbf{x}$.
2. Find an orthogonal matrix U to minimize

$$D(U) = \sum_{i=1}^p \sum_{j=1}^p \|\text{off}(UC_{ij}(\mathbf{z})U^T)\|^2.$$

3. The JADE functional is then $W = U\text{Cov}(\mathbf{x})^{-1/2}$.

It is easy to see that

$$\begin{aligned} C_{ij}(\mathbf{s}) &= 0, \text{ for } i, j = 1, \dots, p \text{ and } i \neq j \text{ and} \\ C_{ii}(\mathbf{s}) &= \beta_{2,i}E_{ii}, \text{ for } i = 1, \dots, p. \end{aligned}$$

The matrices C_{ij} are not equivariant under orthogonal transformations but, surprisingly, the optimization at step 2 is rotation equivariant: If U is the minimizer for \mathbf{z} then UV^T is a minimizer for $V\mathbf{z}$ for all orthogonal V . Again, as the whitened $\mathbf{z} = V\mathbf{s}$ for some orthogonal V , the JADE functional is an IC functional if at most one of the kurtosis values $\beta_{2,i}$ are zero. (If s_i is gaussian then $\beta_{2,i} = 0$.)

For practical data sets, the p^2 matrices can of course be diagonalized only approximately. There are several algorithms available for a simultaneous diagonalization of several symmetric non-negative definite matrices, and the properties of the estimate depend on the chosen algorithm. In our simulation studies, we use a symmetric algorithm based on the Jacobi rotation technique suggested in [4].

The JADE estimator is affine equivariant provided that all the p^2 cumulant matrices are included. The statistical properties of the JADE estimator will be studied in detail in an extended version of this paper. Obviously $C_{ij} = C_{ji}$, and thus only $p(p+1)/2$ matrices need to be diagonalized. Nevertheless, JADE becomes computationally heavy as p grows (see also Section 4).

To reduce computation time of JADE, [5] suggested the SHIBBS (Shifted Block Blind Separation) algorithm:

Definition 4 *The SHIBBS functional is defined as follows.*

1. Whiten the data: $\mathbf{z} = \text{Cov}(\mathbf{x})^{-1/2}\mathbf{x}$, and choose an initial value U .
2. Repeat the following two steps until convergence:
 - Find an orthogonal matrix V to minimize

$$D(V) = \sum_{i=1}^p \|\text{off}(VC_{ii}(U\mathbf{z})V^T)\|^2.$$

- $U = VU$

3. The SHIBBS functional is then $W = U Cov(\mathbf{x})^{-1/2}$.

Compared to JADE, SHIBBS reduces the number of operations if the number of iterations at stage 2 is small relative to p . It is not clear, however, whether the SHIBBS estimate is affine equivariant. For small sample sizes, the convergence to the global optimum cannot be guaranteed. For large (depending on the components and on the validity of the model) number of observations this problem seems to vanish.

3. A NEW EQUIVARIANT ESTIMATOR BASED ON FOURTH ORDER CUMULANTS

In [5] Cardoso noticed that SHIBBS and JADE algorithms separate the sources almost identically, and the simulation results in section 4 support that. This suggests that the use of matrices C_{ii} , $i = 1, \dots, p$, may be sufficient in practice. Also, if one is close to the solution then, for a small k , the matrices C_{ij} , $|i - j| < k$, may carry most of the information for further steps towards the correct solution. To find a good initial solution, one can then use the affine equivariant FOBI algorithm. This is done in the following definition of a new k -JADE functional.

Definition 5 The k -JADE functional is defined as follows.

1. Whiten the data: $\mathbf{z} = Cov(\mathbf{x})^{-1/2}\mathbf{x}$.

2. Find an orthogonal matrix U to minimize

$$D(U) = \|\text{off}(UCov_4(\mathbf{z})U^T)\|^2.$$

Write $\mathbf{z}^* = U\mathbf{z}$ (FOBI solution).

3. Find an orthogonal matrix V to minimize

$$D(V) = \sum_{|i-j|<k} \|\text{off}(VC_{ij}(\mathbf{z}^*)V^T)\|^2.$$

4. The k -JADE functional is then $W = VU Cov(\mathbf{x})^{-1/2}$.

The first remark is that p -JADE is equivalent to regular JADE. Recall that FOBI separates only components with distinct kurtosis values. Therefore, if i th and $(i + 1)$ th largest kurtosis values of the components are equal, then, after the FOBI step, one can expect that $C_{ii}(\mathbf{z}^*)$ and $C_{i+1,i+1}(\mathbf{z}^*)$ (as well as $C_{i,i+1}(\mathbf{z}^*)$) are useful for separating these two components. It is shown in the simulation study that the performance of k -JADE in separating identically distributed components is superior as compared to the performance of FOBI. When choosing k , there is obviously a trade-off between the efficiency and the computation time. The equivariance of k -JADE follows from the equivariance of FOBI.

4. SIMULATION STUDY

All simulations in this paper are done using R 2.15.1 [6] and the package JADE [7].

The performances of FOBI, JADE, SHIBBS and k -JADE were compared in three different models. In model (i) there were four different source components with uniform, normal, t_8 - and t_5 -distributions. The ten source components in model (ii) were all t_5 -distributed. In model (iii) there were two uniformly and two t_5 -distributed components. The source components were centered and scaled to have zero means and unit variances. To compare the estimates, 10000 random samples of size $n = 10000$ were generated from all three models.

To measure the performance of the estimators we used the minimum distance (MD) index [8] defined by

$$\hat{D} = D(\hat{W}A) = \frac{1}{\sqrt{p-1}} \inf_{C \in \mathcal{C}} \|C\hat{W}A - I_p\| \quad (3)$$

with the matrix (Frobenius) norm $\|\cdot\|$ and \mathcal{C} as defined in (2). The minimum distance index is affine invariant and it is scaled to have a maximum value 1. Notice also that $\hat{D} = 0$ iff $\hat{W}A \sim I_p$ indicating a perfect separation of the components.

The simulation results are given in Figure 1. The figure shows that if the source components are different, FOBI yields reasonable unmixing matrix estimates and consequently estimates based on 1-JADE perform as well as JADE. In model (iii) with two identically distributed pairs of components, 2-JADE seems to be sufficient. 1-JADE produces good estimates most of the time, but it “fails” sometimes. Finally, 1-JADE does not break down even though FOBI estimates are poor, and 2-JADE is already quite close to JADE. In all three models the differences between JADE, SHIBBS and 2-JADE seem rather minimal so that computational efficiency might provide a reason for which method to choose.

To compare computation times of JADE, SHIBBS, 1-JADE, 2-JADE and FOBI we used model (ii) with identical t_5 distributions but with the different dimensions $p = 5, 6, \dots, 59, 60, 70, 80, 90, 100$. Figure 4 shows, on a log scale, the average computation time based on 50 runs on a Intel(R) Xeon(R) CPU X5650 with 2.67GHz and 24GB of memory running a 64-bit RedHat Linux. For JADE, SHIBBS, 1-JADE and 2-JADE the convergence criterion used in this simulation study was that $\|U_{k+1} - U_k\| \leq \epsilon = 0.0001$.

The figure shows clearly that JADE is not really practical for data with large number of dimensions. In comparison SHIBBS needs much less computation time than JADE but both are clearly slower than 1-JADE and 2-JADE. The difference of 1-JADE and 2-JADE on the contrary is rather minor and for example the average computation for 1-JADE is 85.31s, for 2-JADE 142.53s, for SHIBBS 520.30s and for JADE 2453.20s for $p = 100$. Since FOBI is basically just an eigenvalue eigenvector decomposition it is of course much faster than any of the other methods but lacks on the other side in performance quality.

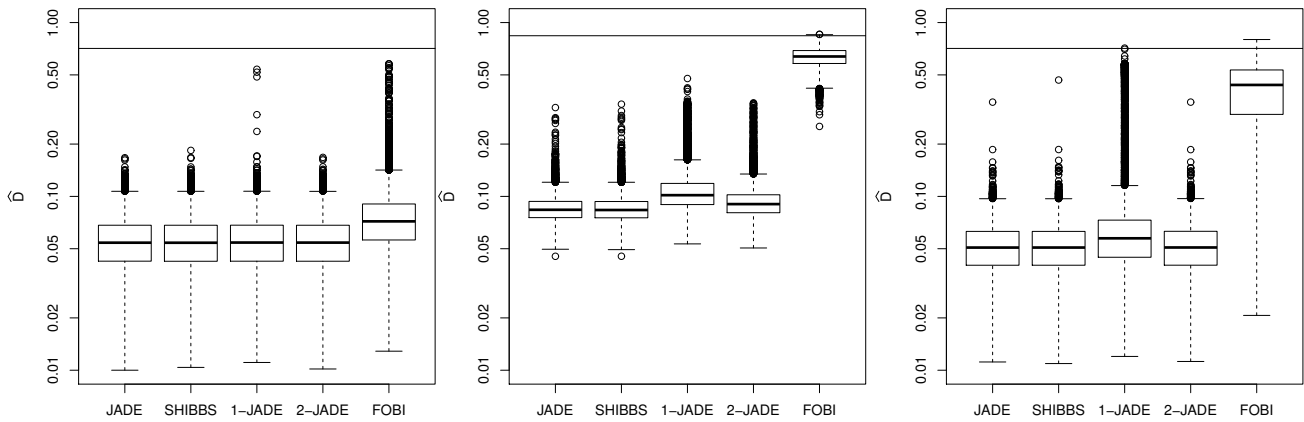


Fig. 1. Comparison of JADE, SHIBBS, 1-JADE, 2-JADE and FOBI estimators in models (i)-(iii) [from left to right] using \hat{D} performance index. The horizontal line indicates the average index value of a random guess. The values are plotted in log scale to make the differences visible.

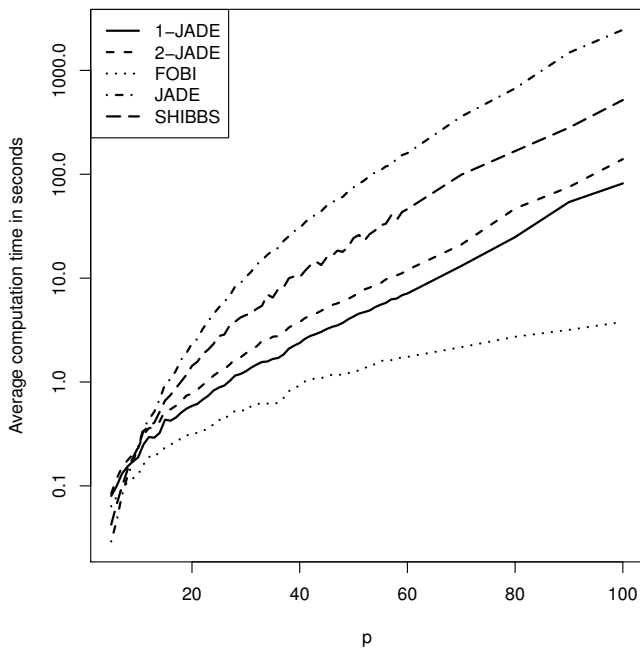


Fig. 2. Average computation time in seconds of JADE, SHIBBS, 1-JADE, 2-JADE and FOBI on log scale for different dimensions p .

5. DISCUSSION

FOBI and JADE are two well-known ICA methods. In this paper we show some connections between the two methods which explain the well-known fact that JADE has a better performance. Although JADE is a popular method, it suffers from the problem of being computationally expensive for high-dimensions. The computational cost is due to the need to compute $p(p+1)/2$ cumulant matrices. While it was early recognized that not all the cumulant matrices are actu-

ally needed, JADE loses affine equivariance if not all of them are used. In SHIBBS algorithm only p cumulant matrices are used. However they need to be computed repeatedly, which seems to provide only little time gain in computation time. In this paper we introduced a new k -JADE method which starts with FOBI and then jointly diagonalizes only a limited number of cumulant matrices controlled by the value of k . The new method is affine equivariant and at least in our simulation studies for $k = 2$ almost as efficient as JADE and SHIBBS, however much faster. The new method thus provides a very efficient tool for practical data analysis. Notice that the equivariance problems of SHIBBS can be avoided by using the same FOBI based whitening as in k -JADE. Another aspect we will pursue is to study how the choice of the joint diagonalization method affects on performance and computation times of the different algorithms. In an extended version of this paper we also plan to derive the asymptotic distribution of JADE and k -JADE.

6. RELATION TO PRIOR WORK

FOBI and JADE are well-established ICA methods with a rich literature in recent years. FOBI was generalized and robustified in [9, 10] by using any two scatter functionals having so called independence property. The asymptotic distribution of FOBI was derived in [11]. Two considerations to speed up JADE were proposed for example by Cardoso and Souloumiac in [3] or as SHIBBS in [5]. However, Cardoso himself recommends against his approach in [3]. SHIBBS was reintroduced as SJAD in [12] and in [13] the authors argue that the computation time of SHIBBS/SJAD can be reduced by relating the convergence criterion to the signal to noise ratio. The limiting distribution of JADE (but without the whitening step) was considered in [14].

7. REFERENCES

- [1] P. Ilmonen, K. Nordhausen, H. Oja, and E. Ollila, "On asymptotics of ICA estimators and their performance indices," <http://arxiv.org/abs/1212.3953>.
- [2] J.-F. Cardoso, "Source separation using higher order moments," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Glasgow, UK, 1989, pp. 2109–2112.
- [3] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals," in *IEE Proceedings-F*, vol. 140, pp. 362–370. IEEE, 1993.
- [4] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 17, pp. 161–164, 1996.
- [5] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, pp. 157–192, 1999.
- [6] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012, ISBN 3-900051-07-0.
- [7] K. Nordhausen, J.F. Cardoso, J. Miettinen, H. Oja, E. Ollila, and S. Taskinen, *JADE: JADE and other BSS methods as well as some BSS performance criteria*, 2012, R package version 1.1-0.
- [8] P. Ilmonen, K. Nordhausen, H. Oja, and E. Ollila, "A new performance index for ica: Properties, computation and asymptotic analysis," in *Latent Variable Analysis and Signal Separation - 9th International Conference, LVA/ICA 2010, St. Malo, France, September 27-30, 2010. Proceedings*, V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent, Eds., vol. 6365 of *Lecture Notes in Computer Science*, pp. 229–236. Springer, 2010.
- [9] H. Oja, S. Sirkia, and J. Eriksson, "Scatter matrices and independent component analysis," *Austrian Journal of Statistics*, vol. 35, pp. 175–189, 2006.
- [10] K. Nordhausen, H. Oja, and E. Ollila, "Robust independent component analysis based on two scatter matrices," *Austrian Journal of Statistics*, vol. 37, pp. 91–100, 2008.
- [11] P. Ilmonen, J. Nevalainen, and H. Oja, "Characteristics of multivariate distributions and the invariant coordinate system," *Statistics and Probability Letters*, vol. 80, pp. 1844–1853, 2010.
- [12] X. Liu and R.B. Randall, "A new efficient independent component algorithm: joint approximate diagonalization of simplified cumulant matrices," in *Proceedings of the 16th National Congress of Australian Institute of Physics*, pp. 29–34. Canberra, Australia, 2005.
- [13] X. Liu, J.-F. Cardoso, and R. B. Randall, "Very fast blind source separation by signal to noise ratio based stopping threshold for the SHIBBS/SJAD algorithm," *Mechanical Systems and Signal Processing*, vol. 24, pp. 2096–2103, 2010.
- [14] S. Bonhomme and J.-M. Robin, "Consistent noisy independent component analysis," *Journal of Econometrics*, vol. 149, pp. 12–25, 2009.