Deflation-based FastICA reloaded

K. Nordhausen[1], Pauliina Ilmonen[1], Abhijit Mandal[1], H. Oja[1] and Esa Ollila[2,3]

[1]School of Health Sciences, University of Tampere, Finland

[2]Signal Processing and Acoustics, Aalto University, Finland [3]Dept. of Electrical Engineering, Princeton University, USA

Introduction

Deflation-based FastICA, where independent components (IC's) are extracted one-by-one, is among the most popular methods for estimating an unmixing matrix in the independent component analysis (ICA) model $\mathbf{x} = \mathbf{A}\mathbf{s}$. The method is usually given for a p-variate random sample $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and a nonlinearity function $g = G'$ as the algorithm:

$\mathbf{x}_i \leftarrow \hat{\mathbf{S}}^{-1/2}(\mathbf{x}_i - \bar{\mathbf{x}})$ {Whiten the data}
$\mathbf{u}_{k,0} \leftarrow \mathbf{u}_{k,init}$ {Choose an initial value}
$\Delta = \infty$
while $\epsilon < \Delta$ **do**
 $\mathbf{u}_{k,1} \leftarrow \mathrm{ave}(\mathbf{x}_i g(\mathbf{u}_{k,0}^T\mathbf{x}_i)) - \mathrm{ave}(g'(\mathbf{u}_{k,0}^T\mathbf{x}_i))\mathbf{u}_{k,0}$
 $\mathbf{u}_{k,1} \leftarrow \mathbf{u}_{k,1} - \sum_{j=1}^{k-1}(\mathbf{u}_{k,1}^T\hat{\mathbf{u}}_j)\hat{\mathbf{u}}_j$
 $\mathbf{u}_{k,1} \leftarrow \mathbf{u}_{k,1}/\|\mathbf{u}_{k,1}\|$
 $\Delta = \|\mathbf{u}_{k,1} - \mathbf{u}_{k,0}\|$
 $\mathbf{u}_{k,0} \leftarrow \mathbf{u}_{k,1}$
end while
RETURN $\hat{\mathbf{u}}_k = \mathbf{u}_{k,1}$

where $\hat{\mathbf{S}} = \hat{\mathrm{COV}}(\mathbf{X})$. The FastICA estimator of the unmixing matrix $\mathbf{W}$ is thus $\hat{\mathbf{W}} = \hat{\mathbf{U}}\hat{\mathbf{S}}^{-1/2}$ with $\hat{\mathbf{U}}$ coming from the algorithm.
Only recently, the statistical properties of deflation-based FastICA have been derived in a series of papers.

ICA equivariance of deflation-based FastICA

The order in which the sources are found by the deflation-based FastICA estimator depends heavily on the initial value $\mathbf{U}_{init} = (\mathbf{u}_{1,init}, \ldots, \mathbf{u}_{p,init})^T$. Write next $\mathbf{W}(\mathbf{U}, \mathbf{X})$ for the estimate based on the data $\mathbf{X}$ and the initial value $\mathbf{U}_{init} = \mathbf{U}$. If $\mathbf{U}$ is random, then the estimate $\mathbf{W}(\mathbf{U}, \mathbf{X})$ may get $p!$ different values depending on random $\mathbf{U}$, and the different solutions may not be ICA equivalent, that is, $\mathbf{W}(\mathbf{U}, \mathbf{B}\mathbf{X})$ may not equal $\mathbf{W}(\mathbf{U}, \mathbf{X})\mathbf{B}^{-1}$.

Estimation equations of deflation-based FastICA

To facilitate statistical analysis, it is appropriate to formulate the method as an estimator verifying a set of estimating equations.
Let $\mathbf{T}(F_\mathbf{x}) = \mathrm{E}(\mathbf{x})$ denote the mean vector (functional). The deflation-based FastICA functional $\mathbf{w}_k(F_\mathbf{x})$, $k = 1, \ldots, p-1$, may be seen as an optimizer of

$$|\,\mathrm{E}[G(\mathbf{w}_k^T(\mathbf{x} - \mathbf{T}(F_\mathbf{x})))]\,|$$

under the constraints (i) $\mathbf{w}_k^T\mathbf{S}(F_\mathbf{x})\mathbf{w}_k = 1$ and (ii) $\mathbf{w}_j^T\mathbf{S}(F_\mathbf{x})\mathbf{w}_k = 0$ for $j = 1, \ldots, k-1$. (For $\mathbf{w}_1$, only the first constraint is needed.) Note that, for the definition of the functional $\mathbf{w}_k$, we need functionals $\mathbf{T}$, $\mathbf{S}$, and $\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}$.
Note that, if $\mathbf{s} = \mathbf{W}\mathbf{x}$ has independent components, then $\mathbf{W}$ solves the corresponding estimating equations. It is also important to note that, for all permutation matrices $\mathbf{P}$, also $\mathbf{P}\mathbf{W}$ then solves the estimating equations, and therefore the estimating equations do not fix the order of the unmixing vectors $\mathbf{w}_1, \ldots, \mathbf{w}_p$.

Limiting distribution of deflation-based FastICA

Without loss of generality we assume that $\mathrm{E}(\mathbf{x}_i) = \mathbf{0}$, $\mathrm{COV}(\mathbf{x}_i) = \mathbf{I}_p$, and the true mixing matrix is $\mathbf{A} = \mathbf{I}_p = (\mathbf{e}_1, \ldots, \mathbf{e}_p)^T$.
If the first four moments of $\mathbf{s}$ exist, then by the central limit theorem, the joint distribution of $\sqrt{n}\bar{\mathbf{x}}$ and $\sqrt{n}\mathrm{vec}(\hat{\mathbf{S}} - \mathbf{I}_p)$ is asymptotically normal. Furthermore, the existence of the expected values $\mu_{g,k} = \mathrm{E}[g(\mathbf{e}_k^T\mathbf{x}_i)]$, $\sigma_{g,k}^2 = Var[g(\mathbf{e}_k^T\mathbf{x}_i)]$, $\lambda_{g,k} = \mathrm{E}[g(\mathbf{e}_k^T\mathbf{x}_i)\mathbf{e}_k^T\mathbf{x}_i]$, $\delta_{g,k} = \mathrm{E}[g'(\mathbf{e}_k^T\mathbf{x}_i)]$ and $\tau_{g,k} = \mathrm{E}[g'(\mathbf{e}_k^T\mathbf{x}_i)\mathbf{e}_k^T\mathbf{x}_i]$ are required.

We also need to assume that $\delta_{g,k} \neq \lambda_{g,k}$, $k = 1, \ldots, p-1$, and we write

$$\alpha_{g,k} = \frac{\sigma_{g,k}^2 - \lambda_{g,k}^2}{(\lambda_{g,k} - \delta_{g,k})^2}, \quad k = 1, \ldots, p. \tag{1}$$

Then, under general assumptions holds:

Result 1 *For $\mathbf{A} = \mathbf{I}_p$, the asymptotic covariance matrix (ASV) of the k-th source $\hat{\mathbf{w}}_k$ is*

$$ASV(\hat{\mathbf{w}}_k) = \sum_{j=1}^{k-1}(\alpha_{g,j} + 1)\mathbf{e}_j\mathbf{e}_j^T + \kappa_k\mathbf{e}_k\mathbf{e}_k^T + \alpha_{g,k}\sum_{l=k+1}^{p}\mathbf{e}_l\mathbf{e}_l^T.$$

where $\kappa_k = (\mathrm{E}(x_{ik}^4) - 1)/4$ and $\alpha_{g,j}$ is defined in (1). Note that the asymptotic variances of the diagonal elements of $\hat{\mathbf{W}}$ do not depend on the choice of the function $g(\cdot)$, but only on the kurtosis of the corresponding source.

Result 2 *If $\frac{1}{\sqrt{n}}\sum_{i=1}^n(g(\mathbf{e}_k^T\mathbf{x}_i) - \mu_{g,k})\mathbf{x}_i$, $k = 1, ..., p$, and $\sqrt{n}\mathrm{vec}(\hat{\mathbf{S}} - \mathbf{I}_p)$ have a joint limiting multivariate distribution, the limiting distribution of $\sqrt{n}\mathrm{vec}(\hat{\mathbf{W}} - \mathbf{I}_p)$ is also multivariate normal. Interestingly, the limiting distributions of the estimated directions $\hat{\mathbf{w}}_1, ..., \hat{\mathbf{w}}_p$ depend on the order in which they are found. The initial value $\mathbf{U}_{init}$ in the FastICA algorithm mainly determines the order of the extracted sources in practice and hence plays a crucial role in the performance of the estimator.*

MD criterion

In this paper we use the so called minimum distance (MD) measure which is defined as

$$MD(\hat{\mathbf{W}}, \mathbf{A}) = \frac{1}{\sqrt{p-1}}\inf_{\mathbf{PD}}\|\mathbf{P}\mathbf{D}\hat{\mathbf{W}}\mathbf{A} - \mathbf{I}_p\|,$$

where $\mathbf{P}$ is a permutation matrix and $\mathbf{D}$ a diagonal matrix with nonzero diagonal entries. This index is independent of the model specification and surprisingly easy to compute.
If an ICA-equivariant estimator $\hat{\mathbf{W}}$ satisfies $\sqrt{n}\mathrm{vec}(\hat{\mathbf{W}} - \mathbf{I}_p) \to_d N_{p^2}(0, \mathbf{\Sigma})$, then the limiting distribution of $nMD^2(\hat{\mathbf{W}}, \mathbf{A})$ is that of a weighted sum of independent chi-square variables.
Also, the expected value $n(p-1)\mathrm{E}[MD^2(\hat{\mathbf{W}}, \mathbf{A})]$ converges to the sum of the limiting variances of the off-diagonal elements of $\hat{\mathbf{W}}$ as $n \to \infty$.

Reloaded deflation-based FastICA

In order to achieve optimal performance in terms of the MD measure, we thus should minimize the sum of the variances of the off-diagonal elements of the FastICA estimator. Using Result 1 it is easy to see that, for $\mathbf{A} = \mathbf{I}_p$,

$$\sum_{i \neq j} ASV(\hat{\mathbf{w}}_{ij}) = 2\sum_{i=1}^{p}(p-i)\alpha_{g,i} + \frac{p(p-1)}{2},$$

which is minimized if the $\alpha_{g,i}$'s are in the increasing order of magnitude.
To optimize the performance of the deflation-based FastICA, we therefore suggest the following simple procedure.

1. Find any equivariant and consistent estimate $\hat{\mathbf{W}}_0$ (e.g. FOBI) such that $\hat{\mathbf{S}}(\hat{\mathbf{W}}_0\mathbf{X}) = \mathbf{I}_p$.
2. Find the estimated sources $\hat{\mathbf{Z}} = \hat{\mathbf{W}}_0(\mathbf{X} - \bar{\mathbf{x}}\mathbf{1}_n^T)$.
3. Find estimates $\hat{\alpha}_{g,k}$, $k = 1, \ldots, p$, based on $\hat{\mathbf{Z}}$ by replacing the expected values by averages in (1).
4. Find the permutation matrix $\hat{\mathbf{P}}$ such that, for the permutated sources, the $\hat{\alpha}_{g,k}$ are in an increasing order.
5. Reload the deflation-based FastICA algorithm with a new initial value: The estimate is $\mathbf{W}(\mathbf{U}(\mathbf{X}), \mathbf{X})$ where $\mathbf{U}(\mathbf{X}) = \hat{\mathbf{P}}\hat{\mathbf{W}}_0\hat{\mathbf{S}}^{1/2}$.

It is easy to see that $\mathbf{W}(\mathbf{U}(\mathbf{X}), \mathbf{X})$ is fully affine equivariant. We conjecture that this new estimator has the same limiting distribution as the simple FastICA estimator which extracts the sources in the (same) optimal order.

Simulation

The data used in our simulations come from a three-variate distribution; the independent source distributions are (i) the exponential distribution (E), (ii) the chi-square distribution with 8 df (C), and (iii) the Laplace distribution (L). All three distributions are centered and scaled to have expected value 0 and variance 1. The mixing matrix used in our simulations is $\mathbf{A} = \mathbf{I}_3$. The sequence ECL, for example, means the extraction order exponential-chi-square-Laplace. We considered two nonlinearity functions $g = pow3$ and $g = tanh$. The values of corresponding $\alpha_{g,k}$, given in Table 1, were obtained from (1), where the expectations were calculated using numerical integration. The values of $n(p-1)\mathrm{E}[MD^2(\hat{\mathbf{W}}, \mathbf{A})]$ for different extraction orders are given in Table 2.

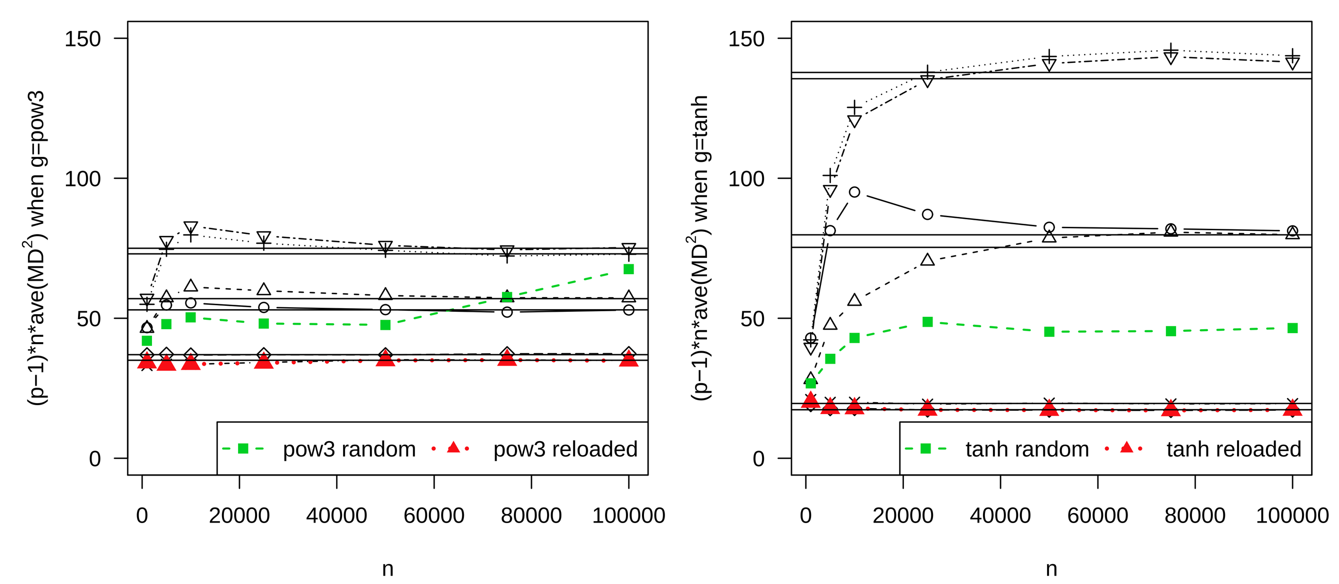
Table 1:

$g(\cdot)$	$\alpha_{g,E}$	$\alpha_{g,C}$	$\alpha_{g,L}$
pow3	5	15	6
tanh	3.14	32.13	2.01

Table 2:

$g(\cdot)$	LCE	LEC	CEL	ECL	CLE	ELC
pow3	57	37	73	53	75	35
tanh	75.32	17.33	137.79	79.80	135.55	19.57

To see whether the expected behavior is observed in finite sample sizes we repeated the estimation of the unmixing matrix 5000 times for different sample sizes using all six possible extraction orders for both nonlinearities. The extraction order can be controlled using six different 3×3 permutation matrices $\mathbf{P}$ as initial values $\mathbf{U}_{init}$. For the reloaded deflation-based FastICA we chose FOBI as the initial estimate. In this simulation study we included the FastICA estimators using random initial values as well.
It is clear that the average MD of the reloaded FastICA corresponds to the minimum value among the six possible cases. Therefore, the reloaded FastICA behaves as expected and is basically equivalent with the best extraction order for that given nonlinearity.



Summary

One important curious property of FastICA is that the extraction order has a huge impact on the separation performance. We used this property and suggested the use of the reloaded deflation-based FastICA to achieve the optimal extraction order. In our approach, we first need to run some ICA procedure that provides a consistent and affine equivariant unmixing matrix estimate. Then the extracted sources are permuted based on the nonlinearity used, and finally the regular deflation-based FastICA is performed using the estimated and permuted sources as whitened data and the identity matrix as an initial value of the rotation matrix. Reloading FastICA this way yields the best extraction order and renders the algorithm more stable at small sample sizes as validated by our simulation studies.

Key References

P. Ilmonen, K. Nordhausen, H. Oja, and E. Ollila (2010). A new performance index for ICA: properties, computation and asymptotic analysis. In Proc. of LVA/ICA 2010, pp. 229–236.

P. Ilmonen, K. Nordhausen, H. Oja, and E. Ollila (2011). Independent component (IC) functionals and a new performance index. Submitted.

E. Ollila (2010). The deflation-based FastICA estimator: statistical analysis revisited. IEEE Trans. Signal Processing, 58, pp. 1527–1541.