Comparative Speed Analysis of FastICA

Vicente Zarzoso and Pierre Comon

Laboratoire I3S, CNRS/UNSA
Les Algorithmes – Euclide-B, BP 121
06903 Sophia Antipolis Cedex, France
{zarzoso,pcomon}@i3s.unice.fr

Abstract. FastICA is arguably one of the most widespread methods for independent component analysis. We focus on its deflation-based implementation, where the independent components are extracted one after another. The present contribution evaluates the method's speed in terms of the overall computational complexity required to reach a given source extraction performance. FastICA is compared with a simple modification referred to as RobustICA, which merely consists of performing exact line search optimization of the kurtosis-based contrast function. Numerical results illustrate the speed limitations of FastICA.

1 Introduction

Independent component analysis (ICA) aims at decomposing an observed random vector into statistically independent variables [1]. Among its numerous applications, ICA is the most natural tool for blind source separation (BSS) in instantaneous linear mixtures when the source signals are assumed to be independent. Under certain identifiability conditions, the independent components correspond to the sources up to admissible scale and permutation indeterminacies [1].

Two main approaches to ICA have been proposed to date. In the original definition of ICA carried out in early works such as [1] and [2], the independent components are extracted jointly or simultaneously, an approach sometimes called symmetric. On the other hand, the deflation approach estimates the sources one after another [3], and has also been shown to work successfully to separate convolutive mixtures [4]. Due to error accumulation throughout successive deflation stages, it is generally acknowledged that joint algorithms outperform deflationary algorithms without necessarily incurring an excessive computational cost.

The FastICA algorithm [5], [6], [7], originally put forward in deflation mode, appeared when many other ICA methods had already been proposed, such as COM2 [1], JADE [2], COM1 [8], or the deflation methods by Tugnait [4] or Delfosse-Loubaton [3]. A thorough comparative study was carried out in [9], where FastICA is found to fail for weak or highly spatially correlated sources. More recently, its convergence has been shown to slow down or even fail in the presence of saddle points, particularly for short block sizes [10].

The objective of the present contribution is to carry out a brief critical review and experimental assessment of the deflationary kurtosis-based FastICA algorithm. In particular, we aim at evaluating objectively the algorithms' speed and

M.E. Davies et al. (Eds.): ICA 2007, LNCS 4666, pp. 293–300, 2007.

efficiency. For the sake of fairness, FastICA is not compared to joint extraction algorithms [1], [2], [3] but only to a simple modification called RobustICA, possibly the simplest deflation algorithm that can be thought of under the same general conditions.

2 Signal Model

The observed random vector $\mathbf{x} \in \mathbb{C}^L$ is assumed to be generated from the instantaneous linear mixing model:

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n} \tag{1}$$

where the source vector $\mathbf{s} = [s_1, s_2, \ldots, s_K]^{\mathrm{T}} \in \mathbb{C}^K$ is made of $K \leq L$ unknown mutually independent components. The elements of mixing matrix $\mathbf{H} \in \mathbb{C}^{L\times K}$ are also unknown, and so is the noise vector $\mathbf{n}$, which is only assumed to be independent of the sources. Our focus is on block implementations, which, contrary to common belief, are not necessarily more costly than adaptive (recursive, on-line, sample-by-sample) algorithms, and are able to use more effectively the information contained in the observed signal block. Given a sensor-output signal block composed of T samples, ICA aims at estimating the corresponding T-sample realization of the source vector.

3 FastICA Revisited

3.1 Optimality Criteria

In the deflation approach, an extracting vector $\mathbf{w}$ is sought so that the estimate

$$z \stackrel{\text{def}}{=} \mathbf{w}^{\mathrm{H}}\mathbf{x} \tag{2}$$

maximizes some optimality criterion or contrast function, and is hence expected to be a component independent from the others. A widely used contrast is the normalized kurtosis, which can be expressed as:

$$\mathcal{K}(\mathbf{w}) = \frac{\mathrm{E}\{|z|^4\} - 2\mathrm{E}^2\{|z|^2\} - |\mathrm{E}\{z^2\}|^2}{\mathrm{E}^2\{|z|^2\}}. \tag{3}$$

This criterion is easily seen to be insensitive to scale, i.e., $\mathcal{K}(\lambda\mathbf{w}) = \mathcal{K}(\mathbf{w})$, $\forall\lambda \neq 0$. Since this scale indeterminacy is typically unimportant, we can impose, without loss of generality, the normalization $\|\mathbf{w}\| = 1$ for numerical convenience. The kurtosis maximization (KM) criterion started to receive attention with the pioneering work of Donoho [11] and Shalvi-Weinstein [12] on blind equalization, and was later employed for source separation even in the convolutive-mixture case [4]. Contrast (3) is quite general in that it does not require the observations to be prewhitened and can be applied to real- or complex-valued sources without any modification.

To simplify the source extraction, the kurtosis-based FastICA algorithm [5], [6], [7] first applies a prewhitening operation resulting in transformed observations with an identity covariance matrix, $\mathbf{R}_x \stackrel{\text{def}}{=} \mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\} = \mathbf{I}$. In the real-valued case, contrast (3) then becomes equivalent to the fourth-order moment criterion:

$$\mathcal{M}(\mathbf{w}) = \mathrm{E}\{|z|^4\}, \tag{4}$$

which must be optimized under a constraint, e.g., $\|\mathbf{w}\| = 1$, to avoid arbitrarily large values of z. Under the same constraint, criteria (3) and (4) are also equivalent if the sources are complex-valued but second-order circular, i.e., the non-circular second-moment matrix $\mathbf{C}_x \stackrel{\text{def}}{=} \mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\}$ is null. Consequently, contrast (4) is less general than criterion (3) in that it requires the observations to be prewhitened and the sources to be real-valued, or complex-valued but circular. Indeed, the extension of the FastICA algorithm to complex signals [13], [14] is only valid for second-order circular sources. In the remainder, we shall restrict our attention to sources fulfilling these requirements.

3.2 Contrast Optimization

Under the constraint $\|\mathbf{w}\| = 1$, the stationary points of $\mathcal{M}(\mathbf{w})$ are obtained as a collinearity condition on $\mathrm{E}\{\mathbf{x}zz^{*2}\}$:

$$\mathrm{E}\big\{|\mathbf{w}^{\mathrm{H}}\mathbf{x}|^2\mathbf{x}\mathbf{x}^{\mathrm{H}}\big\}\mathbf{w} = \lambda\mathbf{w} \tag{5}$$

where λ is a Lagrangian multiplier. As opposed to the claims of [5], eqn. (5) is a fixed-point equation only if λ is known, which is not the case here; λ must be determined so as to satisfy the constraint, and thus it depends on $\mathbf{w}_0$, the optimal value of $\mathbf{w}$: $\lambda = \mathcal{M}(|\mathbf{w}_0{}^{\mathrm{H}}\mathbf{x}|^4\}$.

For the sake of simplicity, λ is arbitrarily set to a deterministic fixed value [5], [7], so that FastICA becomes an approximate standard Newton algorithm, as eventually pointed out in [6]. In the real-valued case, the Hessian matrix of $\mathcal{M}(\mathbf{w})$ is approximated as

$$\mathrm{E}\{(\mathbf{w}^{\mathrm{T}}\mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{w})\mathbf{x}\mathbf{x}^{\mathrm{T}}\} \approx \mathrm{E}\{\mathbf{w}^{\mathrm{T}}\mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{w}\}\mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\} = \mathbf{w}^{\mathrm{T}}\mathbf{w} = \mathbf{I} \tag{6}$$

As a result, the kurtosis-based FastICA reduces to a gradient-descent algorithm with a judiciously chosen fixed step size leading to cubic convergence:

$$\mathbf{w}^{+} = \mathbf{w} - \frac{1}{3}\,\mathrm{E}\{\mathbf{x}(\mathbf{w}^{\mathrm{T}}\mathbf{x})^3\} \tag{7}$$

$$\mathbf{w}^{+} \leftarrow \mathbf{w}^{+}/\|\mathbf{w}^{+}\|. \tag{8}$$

This is a particular instance of the family of algorithms proposed in [4].

4 RobustICA

A simple quite natural modification of FastICA consists of performing exact line search of the kurtosis contrast (3):

$$\mu_{\mathrm{opt}} = \arg\max_{\mu} \mathcal{K}(\mathbf{w} + \mu\mathbf{g}). \tag{9}$$

The search direction $\mathbf{g}$ is typically (but not necessarily) the gradient: $\mathbf{g} = \nabla_{\mathbf{w}}\mathcal{K}(\mathbf{w})$. Exact line search is in general computationally intensive and presents other limitations [15], which explains why, despite being a well-known optimization method, it is very rarely used. However, for criteria that can be expressed as rational functions of μ, such as the kurtosis, the constant modulus [16], [17] and the constant power [18], [19] contrasts, the optimal step size $\mu_{\rm opt}$ can easily be determined by finding the roots of a low-degree polynomial.

At each iteration, optimal step-size (OS) optimization performs the following steps:

S1) Compute OS polynomial coefficients.

For the kurtosis contrast, the OS polynomial is given by:

$$p(\mu) = \sum_{k=0}^{4} a_k \mu^k. \tag{10}$$

The coefficients $\{a_k\}_{k=0}^4$ can easily can be obtained at each iteration from the observed signal block and the current values of $\mathbf{w}$ and $\mathbf{g}$ (their expressions are not reproduced here due to the lack of space; see [20] for details). Numerical conditioning in the determination of $\mu_{\rm opt}$ can be improved by normalizing the gradient vector.

S2) Extract OS polynomial roots $\{\mu_k\}_{k=1}^4$.

The roots of the 4th-degree polynomial (quartic) can be found at practically no cost using standard algebraic procedures known since the 16th century such as Ferrari's formula [15].

S3) Select the root leading to the absolute maximum:

$$\mu_{\rm opt} = \arg\max_k \mathcal{K}(\mathbf{w} + \mu_k \mathbf{g}).$$

S4) Update $\mathbf{w}^+ = \mathbf{w} + \mu_{\rm opt}\mathbf{g}$.

S5) Normalize as in (8).

For sufficient sample size, the computational cost per iteration is $(5L + 12)T$ flops whereas that of FastICA's iteration (7) is $2(L + 1)T$ flops. A flop is conventionally defined as a real product followed by an addition.

To extract more than one independent component, the Gram-Schmidt-type deflationary orthogonalization procedure proposed for FastICA [5], [6], [7] can also be used in conjunction with RobustICA. After step S4, the updated extracting vector is constrained to lie in the orthogonal subspace of the extracting vectors previously found.

5 Numerical Experiments

The experimental analysis of this section aims at evaluating objectively the speed and efficiency of FastICA and RobustICA in several simulation conditions. The influence of prewhitening on the methods' performance is also assessed.

Performance-complexity trade-off. Noiseless unitary random mixtures of K independent unit-power BPSK sources are observed at the output of an $L = K$ element array in signal blocks of T samples. The search for each extracting vector is initialized with the corresponding canonical basis vector, and is stopped at a fixed number of iterations. The total cost of the extraction can then be computed as the product of the number of iterations, the cost per iteration per source (Sec. 4) and the number of sources. Prewhitening, if used, also adds to the total cost. The complexity per source per sample is given by the total cost divided by KT. As a measure of extraction quality, we employ the signal mean square error (SMSE), a contrast-independent criterion defined as

$$\text{SMSE} = \frac{1}{K}\sum_{k=1}^{K} \text{E}\big\{|s_k - \hat{s}_k|^2\big\}. \tag{11}$$

The estimated sources are optimally scaled and permuted before evaluating the SMSE. This performance index is averaged over 1000 independent random realizations of the sources and the mixing matrix. Extraction solutions are computed directly from the observed unitary mixtures (methods labelled as 'FastICA' and 'RobustICA') and after a prewhitening stage based on the SVD of the observed data matrix ('pw+FastICA', 'pw+RobustICA'). The cost of the prewhitening stage is of the order of $2K^2T$ flops.

Fig. 1 summarizes the performance-complexity variation obtained for $T = 150$ samples and different values of the mixture size K. Clearly, the best fastest performance is provided by RobustICA without prewhitening: a given performance

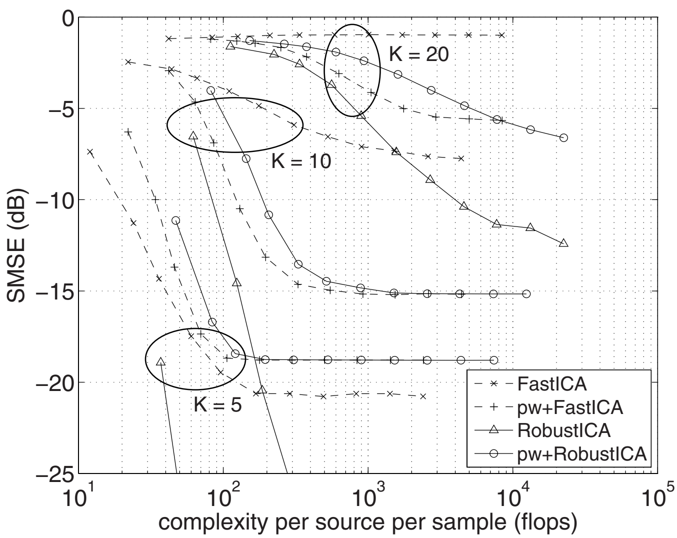

Fig. 1. Average extraction quality against computational cost for different mixture sizes K, with signal blocks of $T = 150$ samples

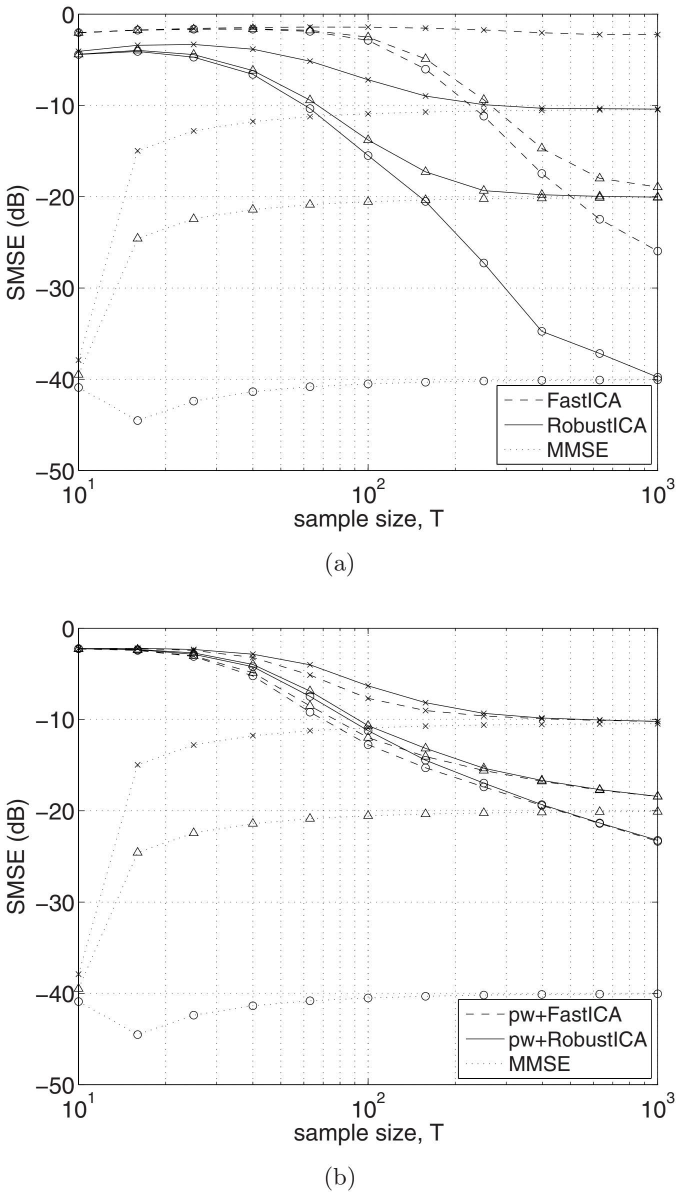

Fig. 2. Average extraction quality against signal block size for unitary mixtures of $K = 10$ sources and a total complexity of 400 flops/source/sample: (a) without prewhitening, (b) with prewhitening. '$\times$': SNR = 10 dB; '$\triangle$': SNR = 20 dB; '$\circ$': SNR = 40 dB

level is achieved with lower cost or, alternatively, an improved extraction quality is reached with a given complexity. The use of prewhitening worsens RobustICA's performance-complexity trade-off and, due to the finite sample size, imposes the same SMSE bound for two methods. Using prewhitening, FastICA improves considerably and becomes slightly faster than RobustICA.

Efficiency. We now evaluate the methods' performance for a varying block sample size T. Extractions are obtained by limiting the number of iterations per source, as explained above. To make the comparison meaningful, the overall complexity is fixed at 400 flops/source/sample for all tested methods. Accordingly, since RobustICA is more costly per iteration than FastICA, it performs fewer iterations per source. Isotropic additive white real Gaussian noise is present at the sensor output, with signal-to-noise ratio:

$$\mathrm{SNR} = \frac{\mathrm{trace}(\mathbf{H}\mathbf{H}^{\mathrm{T}})}{\sigma_n^2 L}. \tag{12}$$

Results for the minimum mean square error (MMSE) receiver are also obtained by jointly estimating the separating vectors assuming that all transmitted symbols are used for training. The MMSE can be considered as a performance bound for linear extraction.

Fig. 2(a) shows the results without prewhitening for random unitary mixtures of $K = 10$ sources and three different SNR values (10 dB, 20 dB and 40 dB). RobustICA attains the MMSE bound for block sizes of about 1000 samples for the tested SNR levels; the required block size can be shown to decrease for smaller K. FastICA seems to require longer block sizes, particularly for noisier conditions at the given overall complexity. As shown in Fig. 2(b), the use of prewhitening in the same experiment worsens the performance-complexity ratio of RobustICA while improving that of FastICA, making both methods' efficiency comparable.

6 Conclusions

The computational complexity required to reach a given source extraction quality is put forward as a natural objective measure of convergence speed for BSS/ICA algorithms. The kurtosis-based FastICA method can be considered as a gradient-based algorithm with constant step size. Its speed is shown to depend heavily on prewhitening and sometimes on initialization. Without the performance limitations imposed by the second-order preprocessing, a simple algebraic line optimization of the more general kurtosis contrast proves computationally faster and more efficient than FastICA even in scenarios favouring this latter method. Although not demonstrated in this paper, RobustICA is also more robust to initialization [20], and the optimal step-size technique it relies on proves less sensitive to saddle points or local extrema [17], [19].

References

1. Comon, P.: Independent component analysis, a new concept? Signal Processing 36(3), 287–314 (1994)
2. Cardoso, J.F., Souloumiac, A.: Blind beamforming for non-Gaussian signals. IEE Proceedings-F 140(6), 362–370 (1993)

3. Delfosse, N., Loubaton, P.: Adaptive blind separation of independent sources: a deflation approach. Signal Processing 45(1), 59–83 (1995)
4. Tugnait, J.K.: Identification and deconvolution of multichannel non-Gaussian processes using higher order statistics and inverse filter criteria. IEEE Transactions on Signal Processing 45, 658–672 (1997)
5. Hyvärinen, A., Oja, E.: A fast fixed-point algorithm for independent component analysis. Neural Computation 9(7), 1483–1492 (1997)
6. Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks 10(3), 626–634 (1999)
7. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons, New York (2001)
8. Comon, P., Moreau, E.: Improved contrast dedicated to blind separation in communications. In: Proc. ICASSP-97, 22nd IEEE International Conference on Acoustics, Speech and Signal Processing, Munich, Germany, April 20-24, 1997, pp. 3453–3456. IEEE Computer Society Press, Los Alamitos (1997)
9. Chevalier, P., Albera, L., Comon, P., Ferreol, A.: Comparative performance analysis of eight blind source separation methods on radiocommunications signals. In: Proc. International Joint Conference on Neural Networks, Budapest, Hungary (July 25-29, 2004)
10. Tichavsky, P., Koldovsky, Z., Oja, E.: Performance analysis of the FastICA algorithm and Cramér-Rao bounds for linear independent component analysis. IEEE Transactions on Signal Processing 54(4), 1189–1203 (2006)
11. Donoho, D.: On minimum entropy deconvolution. In: Proc. 2nd Applied Time Series Analysis Symposium, Tulsa, OK, pp. 565–608 (1980)
12. Shalvi, O., Weinstein, E.: New criteria for blind deconvolution of nonminimum phase systems (channels). IEEE Transactions on Information Theory 36(2), 312–321 (1990)
13. Bingham, E., Hyvärinen, A.: A fast fixed-point algorithm for independent component analysis of complex valued signals. International Journal of Neural Systems 10(1), 1–8 (2000)
14. Ristaniemi, T., Joutsensalo, J.: Advanced ICA-based receivers for block fading DS-CDMA channels. Signal Processing 82(3), 417–431 (2002)
15. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C. The Art of Scientific Computing, 2nd edn. Cambridge University Press, Cambridge (1992)
16. Godard, D.N.: Self-recovering equalization and carrier tracking in two-dimensional data communication systems. IEEE Transactions on Communications 28(11), 1867–1875 (1980)
17. Zarzoso, V., Comon, P.: Optimal step-size constant modulus algorithm. IEEE Transactions on Communications (to appear, 2007)
18. Grellier, O., Comon, P.: Blind separation of discrete sources. IEEE Signal Processing Letters 5(8), 212–214 (1998)
19. Zarzoso, V., Comon, P.: Blind and semi-blind equalization based on the constant power criterion. IEEE Transactions on Signal Processing 53(11), 4363–4375 (2005)
20. Zarzoso, V., Comon, P., Kallel, M.: How fast is FastICA? In: Proc. EUSIPCO-2006, XIV European Signal Processing Conference, Florence, Italy (September 4-8, 2006)