HOW FAST IS FASTICA?

Vicente Zarzoso, Pierre Comon

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

Mariem Kallel

Departement TIC, Laboratoire U2S, ENIT
Campus Universitaire "Le Belvédère"
1002 Tunis, Tunisia

ABSTRACT

The present contribution deals with the statistical tool of Independent Component Analysis (ICA). The focus is on the deflation approach, whereby the independent components are extracted one after another. The kurtosis-based FastICA is arguably one of the most widespread methods of this kind. However, its features, particularly its speed, have not been thoroughly evaluated or compared, so that its popularity seems somewhat unfounded. To substantiate this claim, a simple quite natural modification is put forward and assessed in this paper. It merely consists of performing exact line search optimization of the contrast function. Speed is objectively measured in terms of the computational complexity required to reach a given source extraction performance. Illustrative numerical results demonstrate the faster convergence and higher robustness to initialization of the proposed approach, which is thus referred to as RobustICA.

1. INTRODUCTION

Independent Component Analysis (ICA) transforms an observed random vector into mutually statistically independent components [1]. Its numerous applications have spurred an increasing research interest in this technique; for instance, ICA is the basic statistical tool to perform Blind Source Separation (BSS) [1, 2, 3]. In its original definition (see [1, 4], among other early works), ICA extracts all the sources jointly or simultaneously; this is the so-called "symmetric" approach. ICA can also be performed by estimating the sources sequentially or one by one. This alternative procedure, referred to as *deflation*, was originally proposed in [5], and used successfully in the separation of convolutive mixtures [6]. Deflation has later been widely promoted in the machine learning community [3]. Joint algorithms are usually thought to outperform deflationary algorithms due to errors accumulated in successive subtractions (regressions) of the estimated source contribution to the observation. This shortcoming is generally claimed to be compensated by a significant gain in computations, although this claim still requires closer examination.

The FastICA [7, 8], originally put forward in deflation mode, features among the most popular ICA algorithms. Although it appeared when many other ICA methods had already been proposed, the deflationary FastICA has never been compared by the authors of [3] with earlier joint algorithms such as COM2 [1], JADE [4], COM1 [9], or the deflation methods by Tugnait [6] or Delfosse-Loubaton [5]. In fact, to our knowledge, FastICA (both in its deflation and symmetric implementations) has only been compared with neural-based adaptive algorithms and principal component analysis (PCA), that most ICA algorithms are known to outperform. Its popularity has been justified on the grounds of the satisfactory performance offered by the method in several applications, as well as its simplicity. However, these features, and in particular its speed, have never been substantiated by a thorough comparison with other techniques. A first serious attempt has been made in [10], where FastICA is found to fail for weak or highly spatially correlated sources. In spite of its comprehensiveness, the comparative analysis of [10] is perhaps unfortunate in contrasting the deflationary FastICA with joint methods such as COM2, JADE and COM1. On the other hand, recent studies have put in evidence some deficiencies of FastICA, such as the detrimental effects of saddle points on its performance [11].

Given the assiduous attention the method has received over the last decade, these gaps are somewhat surprising. Indeed, it does not seem difficult to envisage a very simple, quite natural deflation algorithm that would outperform FastICA. The goal of this work is to put forward such a method, which we refer to as *RobustICA*, and compare it with FastICA. The new method simply consists of carrying out exact line search of the contrast function, the normalized kurtosis [12]. Exact line search is achieved at low cost, since the optimal step size (OS) leading to the global maximum along the search direction can algebraically be found at each iteration among the roots of a low-degree polynomial. The OS methodology, which has already been proposed in the time equalization context [13, 14, 15, 16], can be used in conjunction with a variety of alternative criteria such as the constant modulus [17] and the constant power [14, 18]. As part of our experimental study, we evaluate the computational complexity required to reach a given source extraction performance. The algorithms' speed and efficiency can thus be compared objectively.

It is now generally acknowledged that adaptive (also known as on-line, recursive or sample-by-sample) algorithms are not always computationally cheaper than block (off-line, windowed) algorithms, and that they are rarely better in terms of precision. On this account, block implementations are the focus of this paper.

2. MODEL AND NOTATION

Let an L-dimensional random vector $\mathbf{x}$ denote the observation, which is assumed to stem from the linear sta-

In *Proc. EUSIPCO-2006*, Florence, Italy, Sept. 4-8, 2006.

tistical model:

$$\mathbf{x} = \mathbf{Hs} + \mathbf{v}. \quad (1)$$

The source vector $\mathbf{s} = [s_1, s_2, \ldots, s_K]^{\mathrm{T}}$ is made of K statistically mutually independent components. The noise term $\mathbf{v}$ will be ignored throughout, except in the numerical experiments. In fact, its distribution is assumed to be unknown, so that it can at most be considered as a nuisance; otherwise, a maximum likelihood approach could be employed, which is beyond the scope of the present comparison. The goal of ICA can be expressed as follows: given a sensor-output signal block composed of T samples, estimate the corresponding T-sample realization of the source vector.

Vectors and matrices will be typeset in boldface lowercase and boldface uppercase symbols, respectively; superscripts $(^{\mathrm{T}})$, $(^{\mathrm{H}})$, and $(^*)$ denote respectively transposition, conjugate transposition, and complex conjugation. Unless otherwise specified, the components of random vectors $\mathbf{x}$, $\mathbf{s}$ and $\mathbf{v}$ take their values in the complex field.

3. OPTIMALITY CRITERIA

The deflation approach to ICA consists of searching for an extracting vector $\mathbf{w}$ so that its scalar output

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

maximizes some optimality criterion or contrast function. A widely used contrast is the normalized kurtosis of the separator output:

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

This criterion is easily seen to be insensitive to scale, i.e., $\mathcal{K}(\lambda\mathbf{w}) = \mathcal{K}(\mathbf{w})$, $\forall\lambda \neq 0$. This scale indeterminacy is inherent in BSS, and we can thus impose $\|\mathbf{w}\| = 1$ for numerical convenience. Other criteria are the widespread constant modulus (CM) [17]:

$$\mathcal{C}(\mathbf{w}) = \mathrm{E}\{\left(|z|^2 - 1\right)^2\} \quad (4)$$

and the constant power (CP) [14, 18]:

$$\mathcal{P}_r(\mathbf{w}) = \mathrm{E}\{|z^r - 1|^2\}. \quad (5)$$

Another type of objective functions need the data to be *prewhitened*, so that the sensor outputs are assumed to have an identity covariance matrix, $\mathbf{R}_x \stackrel{\text{def}}{=} \mathrm{E}\{\mathbf{xx}^{\mathrm{H}}\} = \mathbf{I}$. One criterion that we shall be particularly interested in is the separator-output fourth-order moment:

$$\mathcal{M}(\mathbf{w}) = \mathrm{E}\{|z|^4\}. \quad (6)$$

This criterion must be optimized under a constraint to avoid arbitrarily large values of z. Assuming $\|\mathbf{w}\| = 1$, it is simple to realize that (6) is equivalent to (3) after prewhitening in two cases: if all sources and mixtures are real-valued, and if the sources are complex-valued but second-order circular, i.e., the non-circular second-moment matrix $\mathbf{C}_x \stackrel{\text{def}}{=} \mathrm{E}\{\mathbf{xx}^{\mathrm{T}}\}$ is null. For instance, in the case where the mixture and noise are complex but the sources are real, criteria (6) and (3) are not equivalent.

4. KURTOSIS-BASED FASTICA

The stationary values of the kurtosis contrast $\mathcal{K}(\mathbf{w})$ are given by the cancellation of its gradient, which is proportional to:

$$\mathrm{E}\{\mathbf{x}zz^{*2}\} - (\mathbf{w}^{\mathrm{T}}\mathbf{C}_x^*\mathbf{w})\mathbf{C}_x\mathbf{w}^* - (\mathbf{w}^{\mathrm{H}}\mathbf{R}_x\mathbf{w})^{-1}\big[\mathrm{E}\{|z|^4\} - |\mathbf{w}^{\mathrm{H}}\mathbf{C}_x\mathbf{w}^*|^2\big]\mathbf{R}_x\mathbf{w}. \quad (7)$$

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

$$\mathrm{E}\{(\mathbf{w}^{\mathrm{H}}\mathbf{xx}^{\mathrm{H}}\mathbf{w})\mathbf{xx}^{\mathrm{H}}\}\mathbf{w} = \lambda\mathbf{w} \quad (8)$$

where λ is some Lagrangian multiplier. It is easy to verify that the same result is obtained by performing the unconstrained optimization of $\mathcal{M}(\mathbf{w})/\|\mathbf{w}\|^4$.

Equation (8) is a fixed-point equation as claimed in [7] only when λ is known, which is not the case here; λ must be determined so as to satisfy the constraint, and thus unfortunately it depends again on $\mathbf{x}$ and $\mathbf{w}$. In [3, 7], λ is arbitrarily set to a deterministic fixed value, which allows to spare computations. For this reason, as eventually pointed out in [8], FastICA is actually an approximate standard Newton algorithm rather than a fixed-point algorithm. As a result of the Hessian matrix approximation carried out under the prewhitening assumption, the kurtosis-based FastICA reduces to a conventional gradient-descent algorithm with a fixed step size, and is hence a particular case of [6]. In the real-valued scenario, FastICA's update rule reads:

$$\mathbf{w}^+ = \mathbf{w} - \frac{1}{3}\,\mathrm{E}\{\mathbf{x}(\mathbf{w}^{\mathrm{T}}\mathbf{x})^3\} \quad (9)$$

$$\mathbf{w}^+ \leftarrow \mathbf{w}^+/\|\mathbf{w}^+\|. \quad (10)$$

The Hessian matrix approximation is somewhat fortunate in that, under the source statistical independence assumption, it theoretically endows the resulting method with global cubic convergence. It is likely that the algorithms described in the next section inherit analogous convergence properties. Nevertheless, the FastICA algorithm sometimes gets stuck at saddle points, particularly for short sample sizes [11].

5. OPTIMAL STEP SIZE: ROBUSTICA

As we have just recalled, FastICA attempts to maximize the normalized kurtosis of the extractor output by means of a approximate Newton algorithm. The Hessian simplification reduces the Newton update to a gradient-based update with fixed step size. For the kurtosis as well as analogous contrast functions commonly encountered in blind signal processing, a more efficient optimization method exists that can improve performance while accelerating convergence. This method, theoretically straightforward yet effective in practice, is exact line maximization.

Line maximization of a generic cost function $\mathcal{J}(\mathbf{w})$ consists of finding its global maximum along a given search direction:

$$\mu_{\rm opt} = \arg\max_{\mu} \mathcal{J}(\mathbf{w} + \mu\mathbf{g}). \quad (11)$$

The direction is typically (but not necessarily) the gradient: $\mathbf{g} = \nabla_{\mathbf{w}}\mathcal{J}(\mathbf{w})$. Exact line search is in general computationally intensive and presents other limitations [19], which explains why, despite being a well-known optimization method, it has largely been disregarded. However, for criteria such as the kurtosis, the CM and the CP contrasts, $\mathcal{J}(\mathbf{w} + \mu\mathbf{g})$ is a low-degree rational function in μ. As a result, the optimal step size $\mu_{\rm opt}$ can be found algebraically (in closed form) among the roots of a simple polynomial of degree D:

$$p(\mu) = \sum_{k=0}^{D} a_k \mu^k. \quad (12)$$

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

S1) Compute OS polynomial coefficients

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

S3) Obtain $\mu_{\rm opt} = \arg\max_{k} \mathcal{J}(\mathbf{w} + \mu_k\mathbf{g})$

S4) Update $\mathbf{w}^+ = \mathbf{w} + \mu_{\rm opt}\mathbf{g}$. (13)

The application of the OS methodology on the kurtosis, the CM and the CP criteria results in the OS kurtosis maximization algorithm (OS-KMA), the OS CM algorithm (OS-CMA), and the OS CP algorithm (OS-CPA), respectively. Note that the above procedure also applies when the contrast function is to be minimized: the minimization of the CM and CP criteria can be achieved through the maximization of $-\mathcal{C}(\mathbf{w})$ and $-\mathcal{P}_r(\mathbf{w})$, respectively. Some important aspects of OS optimization are briefly developed next.

Coefficient computation (step S1). The polynomials associated with the OS-KMA has degree $D = 4$. The derivation of its coefficients is tedious but otherwise straightforward. As summarized in the Appendix, they can be obtained at each iteration from the observed signal block and the current values of $\mathbf{w}$ and $\mathbf{g}$. An alternative version is based on the sensor-output statistics computed once before starting the iterations. This statistics-based version becomes more costly than the data-based version for large values of L. The expressions for the OS-CMA polynomial, which has degree $D = 3$, can be found in [15, 16].

Root extraction and selection (steps S2–S3). The roots of polynomial at orders 3 and 4 can be found with standard algebraic procedures such as Cardano's and Ferrari's formulas, respectively [19]. Preliminary experiments point out that, although complex-valued roots may appear as favourite in the sense of the maximization of $\mathcal{J}(\mathbf{w} + \mu_k\mathbf{g})$, the best real-valued candidate root should typically be preferred.

Normalization. To improve numerical conditioning in the determination of $\mu_{\rm opt}$, the normalized version of the gradient vector should be used in the above steps.

Table 1: Computational complexity per iteration of the deflationary ICA algorithms compared in this paper, for signal blocks of T samples observed at the output of L sensors, assuming real-valued sources and mixtures. The figures in the second row are for the simulation scenario of Sec. 7 and Figs. 1–2.

		RobustICA	
	FastICA	OS-KMA	OS-CMA
(L, T)	$2(L+1)T$	$(5L+12)T$	$(3L+10)T$
$(4, 150)$	1500	4800	3300

As observed in Sec. 3, the kurtosis criterion is scale invariant, so that the new extracting vector $\mathbf{w}^+$ should be normalized as in (10) after each OS-KMA iteration.

Computational complexity. The computational cost per iteration of FastICA and the two OS methods (data-based versions) presented above is shown in Table 1. Only the most significant terms have been retained. These dominant terms are of order $\mathrm{O}(T)$, and provide accurate approximations of the exact cost for sufficient sample size T. Complexity is measured in floating point operations (flops). A flop is considered as a real product.

The OS technique in the blind and semi-blind equalization context is fully developed in [14, 15, 16]; details are omitted here due to space limitations. By design, and as confirmed by simulations, OS optimization provides some robustness to local extrema and reduced overall complexity relative to conventional fixed step-size optimization. In the ICA context, the OS methodology naturally gives rise to what could be referred to as *RobustICA* algorithms. Indeed, improved faster convergence and increased robustness to the initial value of the extracting vector will be illustrated in the experiments of Sec. 7.

6. DEFLATION

After convergence, output signal z contains an estimate $\hat{s}_k$ of source component s_k. In most deflation algorithms (except, e.g., [5]), the extracted-source contribution to the sensor output is estimated by linear regression as $\hat{\mathbf{x}}_k = \hat{\mathbf{h}}_k \hat{s}_k$, with

$$\hat{\mathbf{h}}_k = \mathrm{E}\{\mathbf{x}\hat{s}_k^*\}/\mathrm{E}\{|\hat{s}_k|^2\}. \quad (14)$$

This contribution is then subtracted from the observations, producing a new observed vector

$$\mathbf{x} \leftarrow \mathbf{x} - \hat{\mathbf{x}}_k. \quad (15)$$

From the 'deflated' observations, the next source is estimated by running again the same extraction procedure. The deflation procedure is repeated until no sources are left. In practice, the expectations in (14) are substituted by sample averages over the signal block, which accept efficient matrix-vector product formulations.

7. NUMERICAL EXPERIMENTS

Since FastICA heavily relies on the whitening assumption, only real orthogonal mixtures are considered in the following numerical study, as if prewhitening had been previously carried out. By contrast, a feature of deflation algorithms in general, and RobustICA in particular, is that they can directly operate on the observed sensor output without prewhitening. Hence, the orthogonal mixture scenario benefits the FastICA implementation.

A mixture of $K = 4$ independent unit-power BPSK sources is observed at the output of a $L = 4$ element array in signal blocks of 150 samples. Isotropic additive white real Gaussian noise is present at the sensor output, with signal-to-noise ratio:

$$\mathrm{SNR} = \frac{\mathrm{trace}(\mathbf{H}\mathbf{H}^{\mathrm{T}})}{\sigma_v^2 L} = \frac{1}{\sigma_v^2}. \quad (16)$$

Equivalent thresholds on the separating vector variation and a higher limit of $100L = 400$ iterations are employed as convergence tests. Once all sources have been estimated, they are optimally scaled and permuted to allow a meaningful comparison with the original sources. The signal mean square error (SMSE), defined as

$$\mathrm{SMSE}_k = \mathrm{E}\{|s_k - \hat{s}_k|^2\} \quad (17)$$

is used as separation quality index. The minimum mean square error (MMSE) receiver, which jointly estimates the separating vectors assuming that all transmitted symbols are used for training, provides a performance bound. Computational complexity is measured in terms of the number of floating point operations (flops) required to reach a solution. Performance parameters are averaged over 1000 independent random realizations of the sources, the noise and the mixing matrix.

A single-tap initialization, $\mathbf{w}_0 = [0, 1, 0, 0]^{\mathrm{T}}$, is used for all sources to be extracted. Fig. 1(a) shows the SMSE performance variation as a function of SNR. The first source extracted by OS-KMA and OS-CMA attains the MMSE bound, whereas the first source by FastICA can only achieve the performance of the second source by the other two methods. As expected, performance degrades for subsequent extractions. On average, the RobustICA algorithms clearly outperform FastICA, which shows a worse finite sample-size flooring effect due to the increased misadjustment introduced by its constant step size.

The algorithms' computational complexity is displayed in Fig. 1(b). Flop counts are obtained as the number of iterations times the number of flops per iteration (Table 1). OS-CMA's cost decreases as the SNR increases and as more sources are extracted. The OS-KMA shows a similar trend except for the last source, but its average complexity lies just below OS-CMA's. FastICA is only efficient when extracting the first source in sufficient SNR, and often goes over the iteration-count limit for the remaining sources. On average, FastICA turns out to be well over an order of magnitude more expensive than RobustICA in these experiments, even though its cost per iteration (Table 1) is less than a half and a third of OS-CMA's and OS-KMA's, respectively.

To assess their efficiency, the three methods' average extraction quality as a function of complexity is summarized by the '+'-marked plots in Fig. 2. RobustICA's higher efficiency is remarkable, despite its heavier cost per iteration (Table 1). Note that the MMSE is not an iterative method, and so its cost is irrelevant here; its SMSE value is shown in Fig. 2 for reference only. Also displayed in that figure is the average performance for other initial values of the extracting vector: canonical basis and random. In the former, the separating vector aiming to extract the kth source is initialized with the kth canonical basis vector, $\mathbf{e}_k = [\underbrace{0, \ldots, 0}_{(k-1)}, 1, \underbrace{0, \ldots, 0}_{(L-k)}]^{\mathrm{T}}$, $k = 1, \ldots, K$. In the latter, the initial values of the extracting vector taps are independently drawn from a normalized Gaussian distribution. As observed in these plots, RobustICA's consistent behaviour contrasts with FastICA's sensitivity to initialization.

8. CONCLUSIONS

The main purpose of this contribution was to show that FastICA is probably not the best ICA algorithm, and that its popularity is not based on a solid scientific comparison with earlier algorithms. Its fair simplicity is appealing, but its satisfactory fast performance has long been taken for granted by many researchers in the field. The superior efficiency and increased robustness to initialization of the simple RobustICA technique demonstrate that FastICA can indeed be easily improved. The OS methodology giving rise to RobustICA is not exclusive to the kurtosis criterion, but is applicable to any contrast function that can be expressed as a rational function in the step size. Further work will consider the use of the OS strategy for simultaneous ICA, and its comparison with other techniques.

9. APPENDIX: OS-KMA POLYNOMIAL

The OS polynomial of contrast $\mathcal{K}$ at $\mathbf{w}$ along direction $\mathbf{g}$ has coefficients:

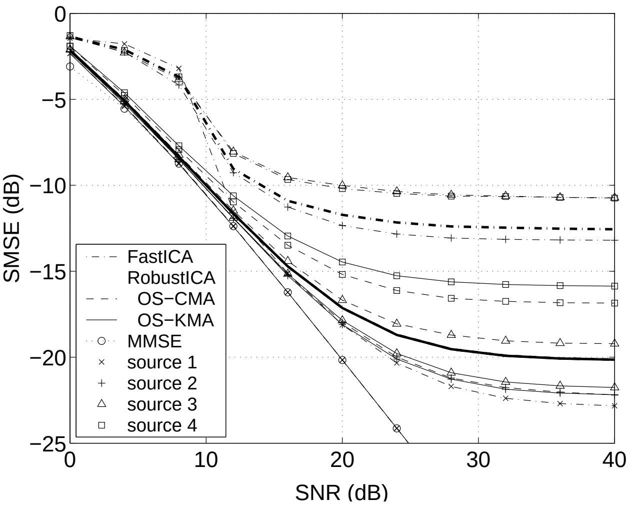
$$a_0 = -2h_0 i_1 + h_1 i_0, \qquad a_1 = -4h_0 i_2 - h_1 i_1 + 2h_2 i_0$$
$$a_2 = -3h_1 i_2 + 3h_3 i_0, \qquad a_3 = -2h_2 i_2 + h_3 i_1 + 4h_4 i_0$$
$$a_4 = -h_3 i_2 + 2h_4 i_1$$

with

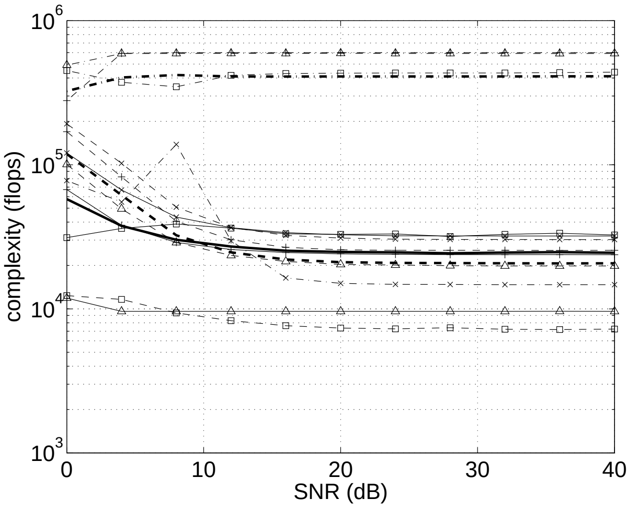
$$\begin{aligned}
h_0 &= \mathrm{E}\{|a|^2\} - 2\mathrm{E}^2\{|a|\} - |\mathrm{E}\{a\}|^2 \\
h_1 &= 4\mathrm{E}\{|a|d\} - 8\mathrm{E}\{|a|\}\mathrm{E}\{d\} - 4\mathrm{Re}(\mathrm{E}\{a\}\mathrm{E}\{c^*\}) \\
h_2 &= 4\mathrm{E}\{d^2\} + 2\mathrm{E}\{|ab|\} - 8\mathrm{E}^2\{d\} - 4\mathrm{E}\{|a|\}\mathrm{E}\{|b|\} \\
&\quad -4|\mathrm{E}\{c\}|^2 - 2\mathrm{Re}(\mathrm{E}\{a\}\mathrm{E}\{b^*\}) \\
h_3 &= 4\mathrm{E}\{|b|d\} - 8\mathrm{E}\{|b|\}\mathrm{E}\{d\} - 4\mathrm{Re}(\mathrm{E}\{b\}\mathrm{E}\{c^*\}) \\
h_4 &= \mathrm{E}\{|b|^2\} - 2\mathrm{E}^2\{|b|\} - |\mathrm{E}\{b\}|^2 \\
i_0 &= \mathrm{E}\{|a|\}, \quad i_1 = 2\mathrm{E}\{d\}, \quad i_2 = \mathrm{E}\{|b|\} \\
a &= y^2, \quad b = g^2, \quad c = yg, \quad d = \mathrm{Re}(yg^*) \\
y &= \mathbf{w}^{\mathrm{H}}\mathbf{x}, \quad g = \mathbf{g}^{\mathrm{H}}\mathbf{x}.
\end{aligned}$$

REFERENCES

[1] P. COMON, "Independent component analysis, a new concept?," *Signal Processing, Elsevier*, vol. 36, no. 3, pp. 287–314, Apr. 1994, Special Issue on Higher-Order Statistics.



(a)



(b)

Figure 1: Performance of deflationary ICA algorithms for single-tap initialization: (a) signal extraction quality, (b) computational complexity. Unmarked thick lines represent performance indices averaged over the 4 sources.

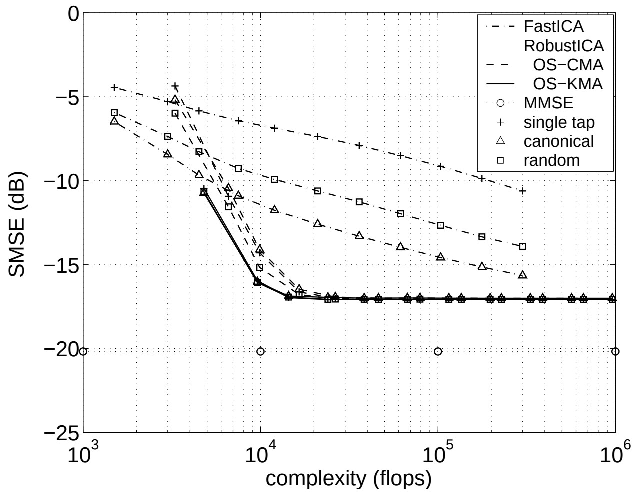


Figure 2: Average extraction quality against computational cost at 20-dB SNR for different extracting vector initializations.

[2] J. F. CARDOSO, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, Jan. 1999.

[3] A. HYVÄRINEN, J. KARHUNEN, and E. OJA, *Independent Component Analysis*, John Wiley & Sons, 2001.

[4] J. F. CARDOSO and A. SOULOUMIAC, "Blind beamforming for non-Gaussian signals," *IEE Proceedings - Part F*, vol. 140, no. 6, pp. 362–370, Dec. 1993, Special issue on Applications of High-Order Statistics.

[5] N. DELFOSSE and P. LOUBATON, "Adaptive blind separation of independent sources: a deflation approach," *Signal Processing*, vol. 45, pp. 59–83, 1995.

[6] J. K. TUGNAIT, "Identification and deconvolution of multichannel non-Gaussian processes using higher order statistics and inverse filter criteria," *IEEE Trans. Sig. Proc.*, vol. 45, pp. 658–672, Mar. 1997.

[7] A. HYVÄRINEN, "A family of fixed-point algorithms for independent component Analysis," in: *Proc. ICASSP*, Munich, Germany, April 20-24 1997, pp. 3917–3920.

[8] A. HYVÄRINEN, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[9] P. COMON and E. MOREAU, "Improved contrast dedicated to blind separation in communications," in: *Proc. ICASSP*, Munich, April 20-24 1997, pp. 3453–3456.

[10] P. CHEVALIER, L. ALBERA, P. COMON, and A. FERREOL, "Comparative performance analysis of eight blind source separation methods on radiocommunications signals," in: *Proc. Intl. Joint Conf. on Neural Networks*, Budapest, Hungary, July 25–29, 2004.

[11] P. TICHAVSKÝ, Z. KOLDOVSKÝ, and E. OJA, "Performance analysis of the FastICA algorithm and Cramér-Rao bounds for linear independent component analysis," *IEEE Trans. Sig. Proc.*, vol. 54, no. 4, pp. 1189–1203, Apr. 2006.

[12] O. SHALVI and E. WEINSTEIN, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Transactions on Information Theory*, vol. 36, no. 2, pp. 312–321, Mar. 1990.

[13] P. COMON, "Independent component analysis, contrasts and convolutive mixtures," in: *Proc. 2nd IMA Conference on Mathematics in Communications*, Lancaster, UK, Dec. 16-18, 2002, pp. 10–17, invited.

[14] V. ZARZOSO and P. COMON, "Blind and semi-blind equalization based on the constant power criterion," *IEEE Trans. Sig. Proc.*, vol. 53, no. 11, Nov. 2005, pp. 4363–4375.

[15] V. ZARZOSO and P. COMON, "Blind channel equalization with algebraic optimal step size," in: *Proc. EUSIPCO, XIII European Signal Processing Conference*, Antalya, Turkey, Sept. 4–8, 2005.

[16] V. ZARZOSO and P. COMON, "Optimal step-size constant modulus algorithm," *IEEE Transactions on Communications*, submitted. (`www.i3s.unice.fr/%7Emh/RR/2004/RR-04.23-V.ZARZOSO.pdf`)

[17] D. N. GODARD, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Transactions on Communications*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.

[18] O. GRELLIER and P. COMON, "Blind separation of discrete sources," *IEEE Signal Processing Letters*, vol. 5, no. 8, Aug. 1998, pp. 212–214.

[19] W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING, *et al.*, *Numerical Recipes in C. The Art of Scientific Computing*, 2nd Ed., Cambridge University Press, 1992.