LABORATOIRE

INFORMATIQUE, SIGNAUX ET SYSTÈMES
DE SOPHIA ANTIPOLIS
UMR 6070

# BLOCK DEFLATION ICA ALGORITHMS

*Vicente ZARZOSO, Pierre COMON*

*Projet* ASTRE

Rapport de recherche
ISRN I3S/RR–2005-36–FR

Novembre 2005

LABORATOIRE I3S: Les Algorithmes / Euclide B – 2000 route des Lucioles – B.P. 121 –
06903 Sophia-Antipolis Cedex, France – Tél. (33) 492 942 701 – Télécopie : (33) 492 942 898
http://www.i3s.unice.fr/I3S/FR/

RÉSUMÉ :

Le présent rapport porte sur l'outil statistique appelé Analyse en Composantes Indépendantes (ICA). On s'intéresse plus particulièrement à l'approche par déflation, dans laquelle les composantes indépendantes sont extraites l'une après l'autre. Après avoir rappelé l'algorithme connu sous le nom de FastICA, une autre méthodologie baptisée RobustICA est proposée. La nouvelle approche est basée sur une règle de mise à jour de type ascente de gradient, qui calcule algébriquement à chaque itération le pas optimal menant au maximum absolu d'une fonction de contraste dans la direction de recherche. La technique à pas optimal peut être utilisée avec une grande variété de fonctions de contraste (telles que le kurtosis, le module constant et la puissance constante), fournit certaine robustesse aux extrema locaux, et accélère notoirement la convergence. Des résultats expérimentaux démontrent la supériorité des performances de RobustICA par rapport à FastICA, à la fois en termes de précision et de charge de calcul. Bien qu'ayant été proposé il y a presque une dizaine d'années, c'est la première fois que l'algorithme FastICA est évalué de manière sérieuse.

MOTS CLÉS :

algorithme du module constant, analyse en composantes indépendantes, FastICA, fonctions de contraste, maximisation du kurtosis, non gaussien, pas optimal, RobustICA, séparation aveugle de sources, statistiques d'ordre supérieur

---

ABSTRACT:

The present report 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. After reviewing the so-called FastICA algorithm, another methodology named RobustICA is put forward. The new approach is based on a special gradient-ascent update rule where the optimal step size finding the global maximum of the contrast function along the search direction is algebraically computed at each iteration. The optimal step-size technique can be used with a variety of contrast functions (such as kurtosis, constant modulus and constant power), provides some robustness to local extrema, and notably accelerates convergence. Experimental results demonstrate the superior performance of RobustICA compared to FastICA both in terms of accuracy and computational load. Although it was proposed nearly a decade ago, it is the first time that the FastICA algorithm is seriously evaluated.

KEY WORDS :

blind source separation, independent component analysis, constant modulus algorithm, constrast functions, FastICA, higher-order statistics, kurtosis maximization, non-Gaussian, optimal step size, RobustICA

# Block Deflation ICA Algorithms†

Vicente Zarzoso and Pierre Comon

Laboratoire I3S, UNSA/CNRS, Les Algorithmes – Euclide-B, BP 121

06903 Sophia Antipolis Cedex, France

{zarzoso, comon}@i3s.unice.fr

**Abstract**

The present report 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. After reviewing the so-called FastICA algorithm, another methodology named RobustICA is put forward. The new approach is based on a special gradient-ascent update rule where the optimal step size finding the global maximum of the contrast function along the search direction is algebraically computed at each iteration. The optimal step-size technique can be used with a variety of contrast functions (such as kurtosis, constant modulus and constant power), provides some robustness to local extrema, and notably accelerates convergence. Experimental results demonstrate the superior performance of RobustICA compared to FastICA both in terms of accuracy and computational load. Although it was proposed nearly a decade ago, it is the first time that the FastICA algorithm is seriously evaluated.

**Résumé**

Le présent rapport porte sur l'outil statistique appelé Analyse en Composantes Indépendantes (ICA). On s'intéresse plus particulièrement à l'approche par déflation, dans laquelle les composantes indépendantes sont extraites l'une après l'autre. Après avoir rappelé l'algorithme connu sous le nom de FastICA, une autre méthodologie baptisée RobustICA est proposée. La nouvelle approche est basée sur une règle de mise à jour de type ascente de gradient, qui calcule algébriquement à chaque itération le pas optimal menant au maximum absolu d'une fonction de contraste dans la direction de recherche. La technique à pas optimal peut être utilisée avec une grande variété de fonctions de contraste (telles que le kurtosis, le module constant et la puissance constante), fournit certaine robustesse aux extrema locaux, et accélère notoirement la convergence. Des résultats expérimentaux démontrent la supériorité des performances de RobustICA par rapport à FastICA, à la fois en termes de précision et de charge de calcul. Bien qu'ayant été proposé il y a presque une dizaine d'années, c'est la première fois que l'algorithme FastICA est évalué de manière sérieuse.

† A shortened version of this report has been submitted to the *31st IEEE International Conference on Accoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France, May 14–19, 2006.

## I. INTRODUCTION

Independent Component Analysis (ICA) aims at transforming an observed random vector so that its components become mutually independent. A strong research interest in this statistical technique has been raised by its numerous applications; for instance, ICA is the basic tool for Blind Source Separation (BSS) [1], [2], [3]. The "symmetric" approach to ICA, originally proposed in [1], [4] among other early works, extracts all sources jointly or simultaneously. ICA can also be performed by estimating the sources 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, but this claim has not been thoroughly examined. Despite the arousing interest in the deflation approach and its apparent computational appeal, no comparison has yet been published in the open literature confirming the benefits of deflationary relative to joint ICA.

A good example of this gap is the well-known FastICA algorithm [7], [8], arguably the most popular ICA method and originally put forward in deflation mode. Surprisingly, the deflationary FastICA algorithm has never been compared by the authors of [3] with earlier joint algorithms such as 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.

The goal of the present work is to shed some light on whether deflation algorithms are indeed attractive, and to understand their specific advantages and drawbacks. As a first step towards this objective, three deflationary algorithms are compared in this report. The first method is the popular kurtosis-based FastICA algorithm [8]. The two other methods rely on the widespread constant modulus (CM) [10] and kurtosis maximization (KM) [11] criteria. The CM and KM contrast functions are optimized through an optimal step-size (OS) gradient-ascent algorithm, giving rise to the so-called OS-CMA and OS-KMA, respectively. Since OS-based optimization provides certain robustness against local extrema of any contrast function, this approach may be called *RobustICA* when used in the present context.

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 report.

After summarizing the basic signal model and mathematical notation in Section II, optimal criteria for ICA are reviewed in Section III. The FastICA algorithm is critically revisited in Section IV, whereas the novel RobustICA approach is presented in Section V. The regression-based procedure employed by deflationary methods is recalled in Section VI. Experimental results are reported in Section VII, and some conclusions drawn in Section VIII.

## II. Model and Notation

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

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

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. In light of model (1), 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 $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, and $(\cdot)^*$ 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 $\mathbb{C}$.

## III. Optimality Criteria

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

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

maximizes some optimality criterion or contrast function. A widely used contrast is the normalized kurtosis of the separator output [6], [11]:

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

This criterion is easily seen to be insensitive to scale, i.e., $\mathcal{K}(\lambda \mathbf{w}) = \mathcal{K}(\mathbf{w})$, $\forall \lambda \neq 0$. 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) [10]:

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

and the constant power (CP) [14], [17]:

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

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

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

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 $\mathbf{C}_x \stackrel{\text{def}}{=} \mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\}$ 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.

The maximization of $|\mathcal{K}(\mathbf{w})|$, resulting in the kurtosis maximization (KM) criterion, is applicable to any type of non-Gaussian sources. The minimization of the CM contrast $\mathcal{C}(\mathbf{w})$ is valid for sub-Gaussian (short-tailed, platykurtic) sources, which typically arise in digital communications. The minimization of $\mathcal{P}_r(\mathbf{w})$ specifically targets sources with $r$-ary PSK modulation.

## IV. 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:

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

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

$$\mathrm{E}\{(\mathbf{w}^{\mathrm{H}}\mathbf{x}\mathbf{x}^{\mathrm{H}}\mathbf{w})\mathbf{x}\mathbf{x}^{\mathrm{H}}\}\mathbf{w} = \lambda\mathbf{w} \tag{8}$$

where $\lambda$ 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 $\lambda$ is known, which is not the case here; $\lambda$ must be determined so as to satisfy the constraint, and thus unfortunately it depends again on $\mathbf{x}$ and $\mathbf{w}$. In [3], [7], $\lambda$ is arbitrarily set to a deterministic fixed value, which allows to spare computations. For this reason, as eventually pointed out in [8], the FastICA algorithm 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-ascent algorithm with a fixed step size, and is hence a particular case of the 4th-order technique presented in [6]. In the real-valued scenario, FastICA's update rule reads:

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

$$\mathbf{w}^{+} \leftarrow \mathbf{w}^{+}/\|\mathbf{w}^{+}\|. \tag{10}$$

The Hessian matrix approximation is somewhat fortunate in that, under the source statistical independence assumption, it theoretically endows the resulting algorithm with global cubic convergence. It is likely that the algorithms described in the next section inherit analogous convergence properties. Nevertheless, FastICA sometimes gets stuck at saddle points, particularly for short sample sizes [12].

## V. Optimal Step Size: RobustICA

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

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

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. However, for contrast functions such as the KM, CM and CP contrasts, $\mathcal{J}(\mathbf{w} + \mu\mathbf{g})$ is a low-degree rational function in $\mu$. As a result, the optimal step size $\mu_{\mathrm{opt}}$ can be found algebraically (in closed form) among the roots of a simple polynomial of degree at most 4.

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

Compute the OS polynomial coefficients

Extract the OS polynomial roots $\{\mu_k\}$

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

$$\mathbf{w}^{+} = \mathbf{w} + \mu_{\mathrm{opt}}\mathbf{g}. \tag{12}$$

The coefficients of the OS polynomial depend on the specific contrast function, and can be algebraically calculated at each iteration from the observed signal block and the current values of $\mathbf{w}$ and $\mathbf{g}$. To improve numerical conditioning in the determination of $\mu_{\mathrm{opt}}$, the normalized version of the gradient vector should be used in the above steps. The OS technique is readily extended to contrast-function minimization. The application of the OS methodology on the KM, CM and CP criteria result in the OS KM algorithm (OS-KMA), the OS CM algorithm (OS-CMA), and the OS CP algorithm (OS-CPA), respectively. The polynomials associated with the OS-CMA and the OS-KMA are of respective degree 3 and 4. The roots of such polynomials can be found with standard algebraic procedures such as Cardano's and Ferrari's formulas. As observed above, 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.

The OS technique in the blind and semi-blind equalization context is fully developed in [14], [15], [16]; details are omitted here for the sake of conciseness. A algorithm of this type was first proposed in a deflationary BSS context in [17]. 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.

The computational cost per iteration of FastICA and the two RobustICA techniques presented above is shown in Table I, where 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 $T$.

## VI. 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] or FastICA; see below), 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

$$\hat{\mathbf{h}}_k = \mathrm{E}\{\mathbf{x}\hat{s}_k^*\}/\mathrm{E}\{|\hat{s}_k|^2\}. \tag{13}$$

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

$$\mathbf{x} \leftarrow \mathbf{x} - \hat{\mathbf{x}}_k. \tag{14}$$

From the 'deflated' observations, the next source is estimated by running again the same extraction algorithm. The deflation procedure is repeated until no sources are left. In practice, the expectations in (13) are substituted by sample averages over the signal block, which accept efficient matrix-vector product formulations.

Deflation-based RobustICA can easily be extended to time-dispersive channels, which give rise to convolutive mixtures. In such scenarios, the deflationary OS-KMA would correspond to the OS version of the 4th-order technique of [6], where cost-function maximization is carried out through a somewhat heuristic adaptive step-size procedure (see [6, Table I]).

The deflation procedure described above is based on the linear regression of the estimated source signal vector on the observation matrix row space. By contrast, the deflation technique of [5] exploits a special parameterization of the extracting vectors as a function of certain rotation angles. Likewise, the FastICA algorithm [3], [7], [8] does not employ regression. Instead, after each iteration the algorithm projects the updated extracting vector onto the orthogonal subspace of the previously estimated vectors. Both the special parameterization of [5] and the Gram-Schmidt orthogonalization of [3], [7], [8] rely heavily on prewhitening. The regression-based deflation procedure is more general in that prewhitening is not required, but is still applicable if prewhitening is used. Regression is expected to provide improved quality indices based on the extracted signal estimates, whereas the other deflation techniques would yield more favourable performance in terms of the estimated mixing matrix.

## VII. Numerical Experiments

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 $T = 150$ samples. Since FastICA heavily relies on the whitening assumption, only real orthogonal mixtures are considered, 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. Isotropic additive white real Gaussian noise is present at the sensor output, with signal-to-noise ratio:

$$\mathrm{SNR} = \frac{\mathrm{trace}(\mathbf{H}\mathbf{H}^{\mathrm{T}})}{\sigma_v^2 L} = \frac{1}{\sigma_v^2}. \tag{15}$$

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

$$\mathrm{SMSE}_k = \mathrm{E}\{|\mathbf{s}_k - \hat{\mathbf{s}}_k|^2\} \tag{16}$$

and the extractor output bit error rate (BER) are used as separation quality indices. The minimum mean square error (MMSE) receiver, which jointly estimates the separating vectors assuming that all transmitted symbols are used for training, provides a bound of performance. 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. The three methods operate on identical sets of source, noise, mixing matrix, and initial separating vector realizations. Such sets are generated once before starting each computer experiment and re-used at each value of the independent parameter (e.g., the SNR). The following experiments evaluate the methods' performance under different initializations of the separating vector.

*Experiment 1: Single-tap initialization.* 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-CMA and OS-KMA attains the MMSE bound, whereas the first source by FastICA only slightly improves 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. Consequently, FastICA cannot get below the $10^{-2}$ average BER performance, as shown in Fig. 1(b).

The algorithms' computational burden, in terms of iterations and flops required for convergence, is displayed in Figs. 1(c)–(d), respectively. The three methods' flop count per iteration appears in Table I. 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 exceeds 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 this experiment, even though its cost per iteration (Table I) is less than twice and thrice lower than OS-CMA and OS-KMA, respectively. The three methods' performance as a function of complexity at a fixed SNR level is summarized by the plots of Fig. 1(e)–(f), which clearly illustrate RobustICA's higher efficiency. Note that the MMSE is not an iterative method, and so its cost is irrelevant here; its SMSE value is shown in Figs. 1(e)–(f) for reference only.

*Experiment 2: Canonical basis initialization*. The separating vector aiming to extract the $k$th source is initialized with the $k$th canonical basis vector of $\mathbb{R}^L$, $\mathbf{e}_k = [\underbrace{0, \dots, 0}_{(k-1)}, 1, \underbrace{0, \dots, 0}_{(L-k)}]^{\mathrm{T}}$, $k = 1, \dots, K$. The simulation set-up is otherwise as in the previous experiment. Fig. 2 shows that FastICA's extraction quality improves relative to the previous scenario, approaching RobustICA's performance and reaching a $10^{-3}$ average BER. However, the two RobustICA methods offer virtually the same results as in the previous scenario, and still provide a much more efficient performance than FastICA.

*Experiment 3: Random initialization*. The initial values of the separating vector taps are independently drawn from a normalized Gaussian variable, $\mathbf{w}_0(i) \equiv \mathcal{N}(0, 1)$, $i = 1, \dots, L$, yielding the results of Fig. 3 under the above general conditions. RobustICA's performance hardly changes compared to Experiments 1–2, whereas FastICA's lies between that of the previous experiments.

TABLE I

COMPUTATIONAL COMPLEXITY PER ITERATION OF THE ICA ALGORITHMS COMPARED IN THIS REPORT, FOR SIGNAL BLOCKS OF $T$ SAMPLES OBSERVED AT THE OUTPUT OF $L$ SENSORS, ASSUMING REAL-VALUED SOURCES AND MIXTURES. COMPLEXITY IS MEASURED IN FLOATING POINT OPERATIONS (FLOPS). A FLOP IS CONSIDERED AS A REAL PRODUCT. THE FIGURES IN THE SECOND ROW ARE FOR THE SIMULATION SCENARIO OF SECTION VII AND FIGS. 1–4.

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

The three methods' performance averaged over the extracted sources for Experiments 1–3 is summarized in Fig. 4. RobustICA's consistent behaviour contrasts with FastICA's sensitivity to initialization.

## VIII. CONCLUSIONS

This report has focused on the deflation approach to ICA. A novel class of ICA algorithms, gathered under the name of RobustICA, has been put forward. In this new approach, the optimal step-size (OS) of the gradient-ascent iteration is algebraically computed at each iteration in order to globally optimize the contrast function along the search direction. The OS methodology is not limited to the kurtosis criterion, but can also be used in conjunction with alternative contrasts. Numerical experiments have illustrated the superior performance and lower computational cost of RobustICA relative to FastICA, as well as its notably higher robustness to the initial value of the separating vector. Further work will consider the use of the OS strategy for simultaneous ICA, and its comparison with deflation techniques.

## REFERENCES


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

[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*, Wiley, 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, no. 1, pp. 59–83, July 1995.

[6] J. K. TUGNAIT, "Identification and deconvolution of multichannel non-Gaussian processes using higher order statistics and inverse filter criteria," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 658–672, Mar. 1997.

[7] A. HYVÄRINEN, "A family of fixed-point algorithms for independent component Analysis," in *Proc. ICASSP'97*, Munich, Germany, Apr. 20–24, 1997, pp. 3917–3920.

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

[9] P. COMON and E. MOREAU, "Improved contrast dedicated to blind separation in communications," in *Proc. ICASSP'97*, Munich, Germany, Apr. 20–24, 1997, pp. 3453–3456.

[10] 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.

[11] 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.

[12] P. TICHAVSKÝ, Z. KOLDOVSKÝ, and E. OJA, "Performance analysis of the FastICA algorithm and Cramér-Rao bounds for linear independent component analysis," *IEEE Transactions on Signal Processing*, to appear.

[13] P. COMON, "Independent component analysis, contrasts and convolutive mixtures," in *Proc. Second 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 Transactions on Signal Processing*, vol. 53, no. 11, pp. 4363–4375, Nov. 2005.

[15] V. ZARZOSO and P. COMON, "Blind channel equalization with algebraic optimal step size," in *Proc. EUSIPCO-2005, 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 (`http://www.i3s.unice.fr/%7Emh/RR/2004/RR-04.23-V.ZARZOSO.pdf`).

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

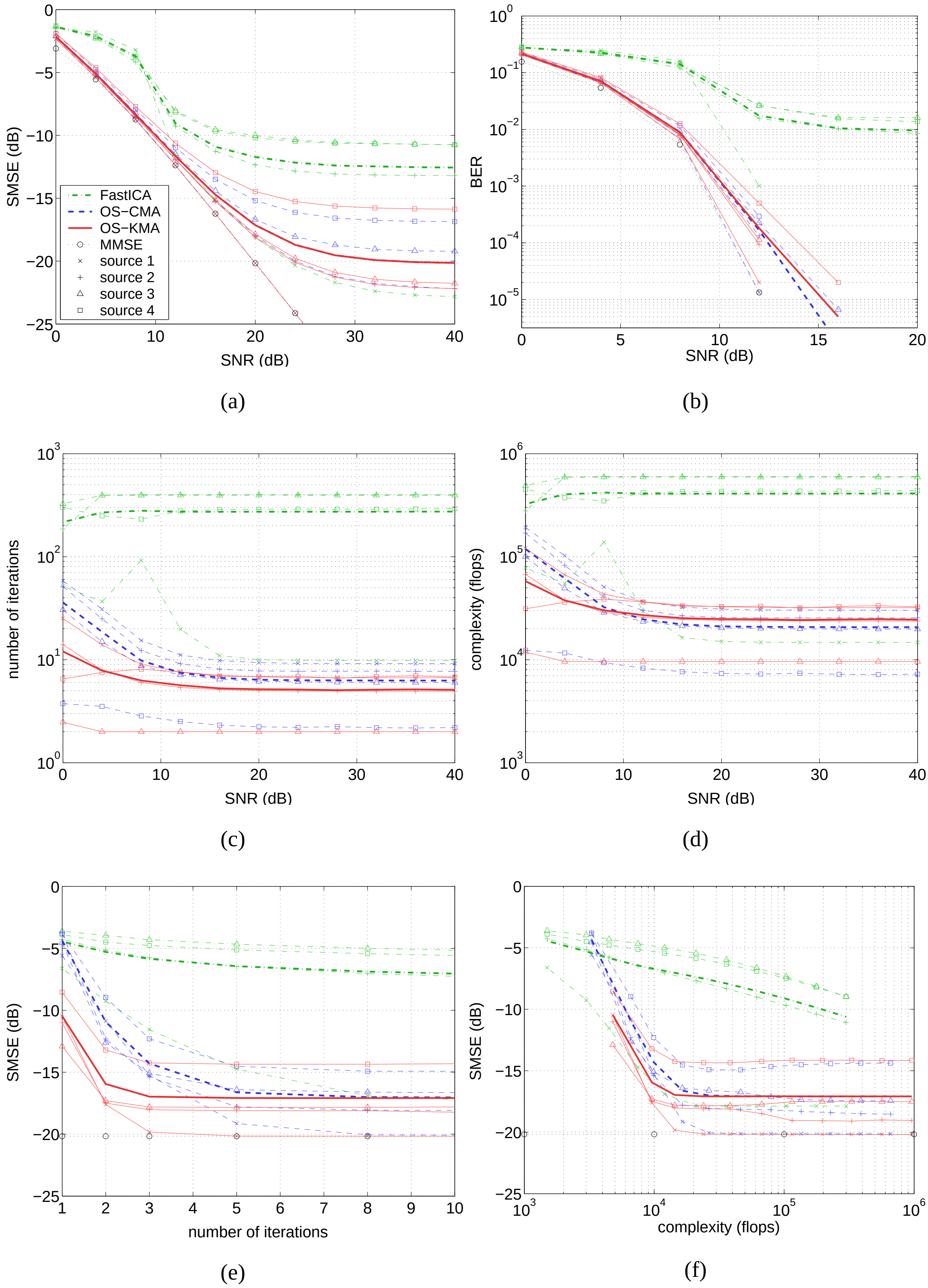


Fig. 1. Experiment 1: block deflationary ICA algorithms with single-tap initialization, $\mathbf{w}_0 = [0, 1, 0, 0]^{\mathrm{T}}$. (a) Source extraction quality, (b) separator-output bit error rate, (c) number of iterations for convergence, (d) computational complexity, (e) extraction quality against iteration count at 20-dB SNR, (f) extraction quality against computational cost at 20-dB SNR. Unmarked thick lines represent performance indices averaged over the 4 sources.

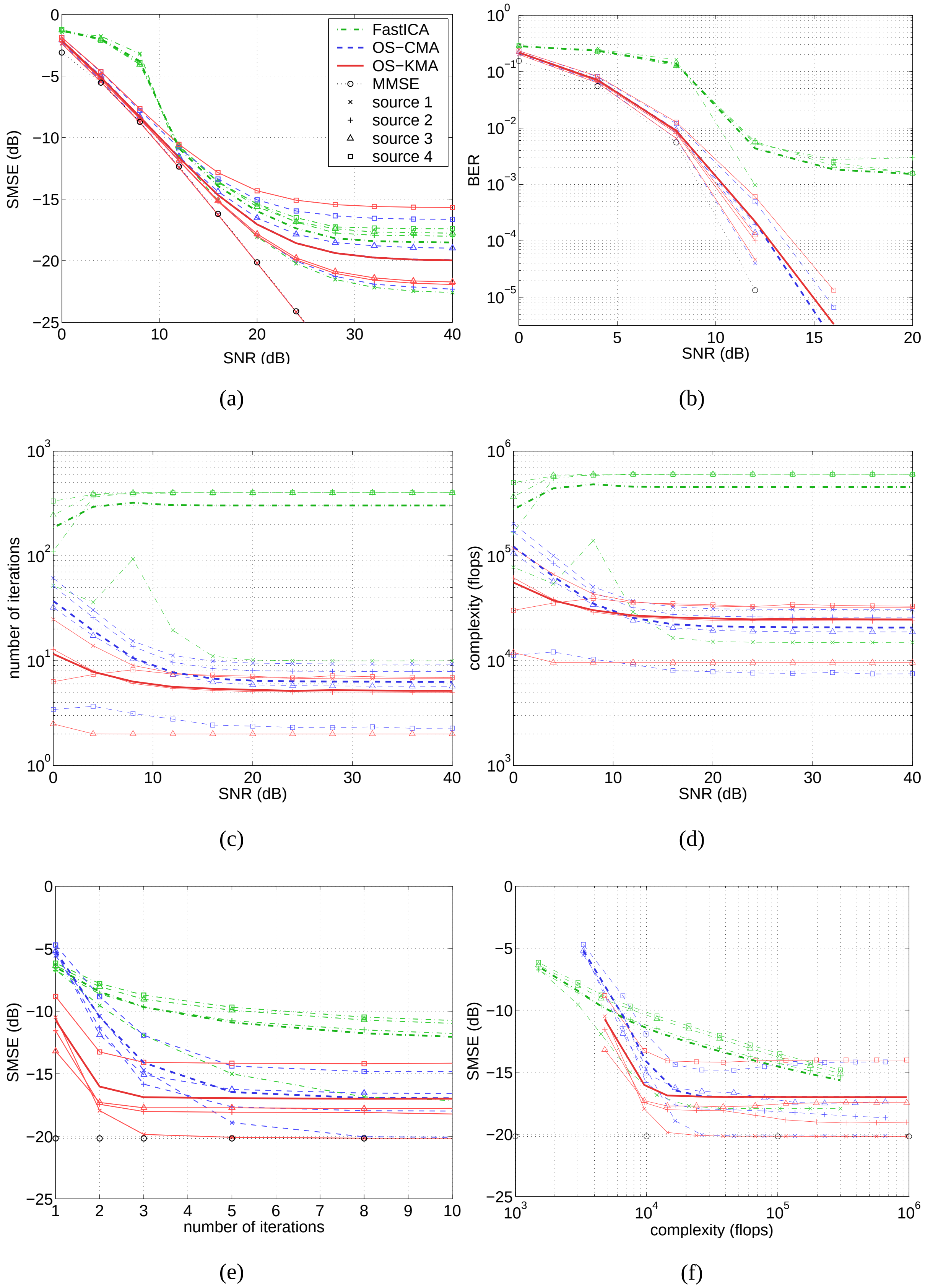


Fig. 2. Experiment 2: block deflationary ICA algorithms with canonical basis initialization, $\mathbf{w}_0 = \mathbf{e}_k$, $k = 1, \ldots, K$. (a) Source extraction quality, (b) separator-output bit error rate, (c) number of iterations for convergence, (d) computational complexity, (e) extraction quality against iteration count at 20-dB SNR, (f) extraction quality against computational cost at 20-dB SNR. Unmarked thick lines represent performance indices averaged over the 4 sources.

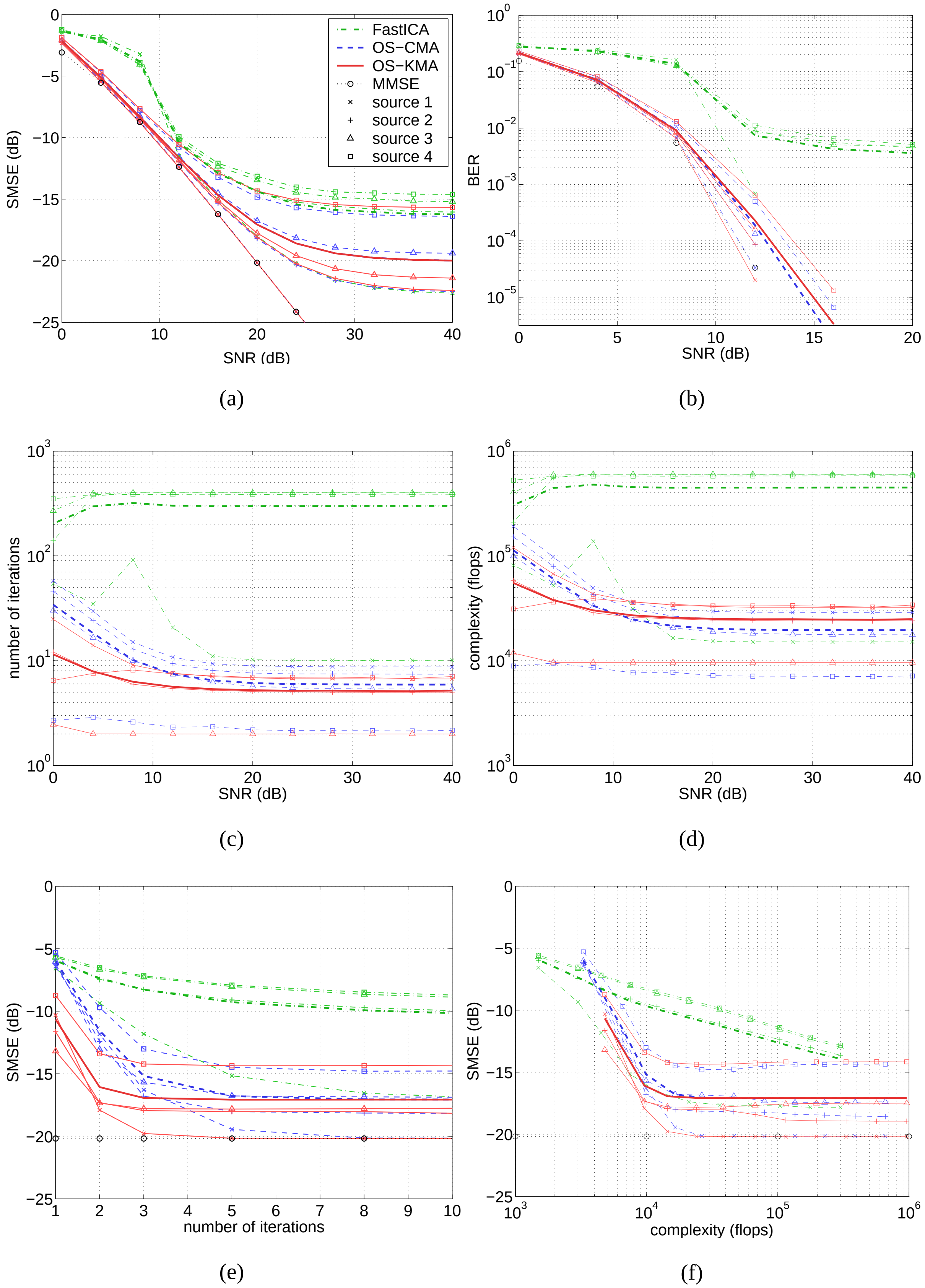


Fig. 3. Experiment 3: block deflationary ICA algorithms with random Gaussian initialization, $\mathbf{w}_0(i) \equiv \mathcal{N}(0,1)$, $i = 1, \ldots, L$. (a) Source extraction quality, (b) separator-output bit error rate, (c) number of iterations for convergence, (d) computational complexity, (e) extraction quality against iteration count at 20-dB SNR, (f) extraction quality against computational cost at 20-dB SNR. Unmarked thick lines represent performance indices averaged over the 4 sources.

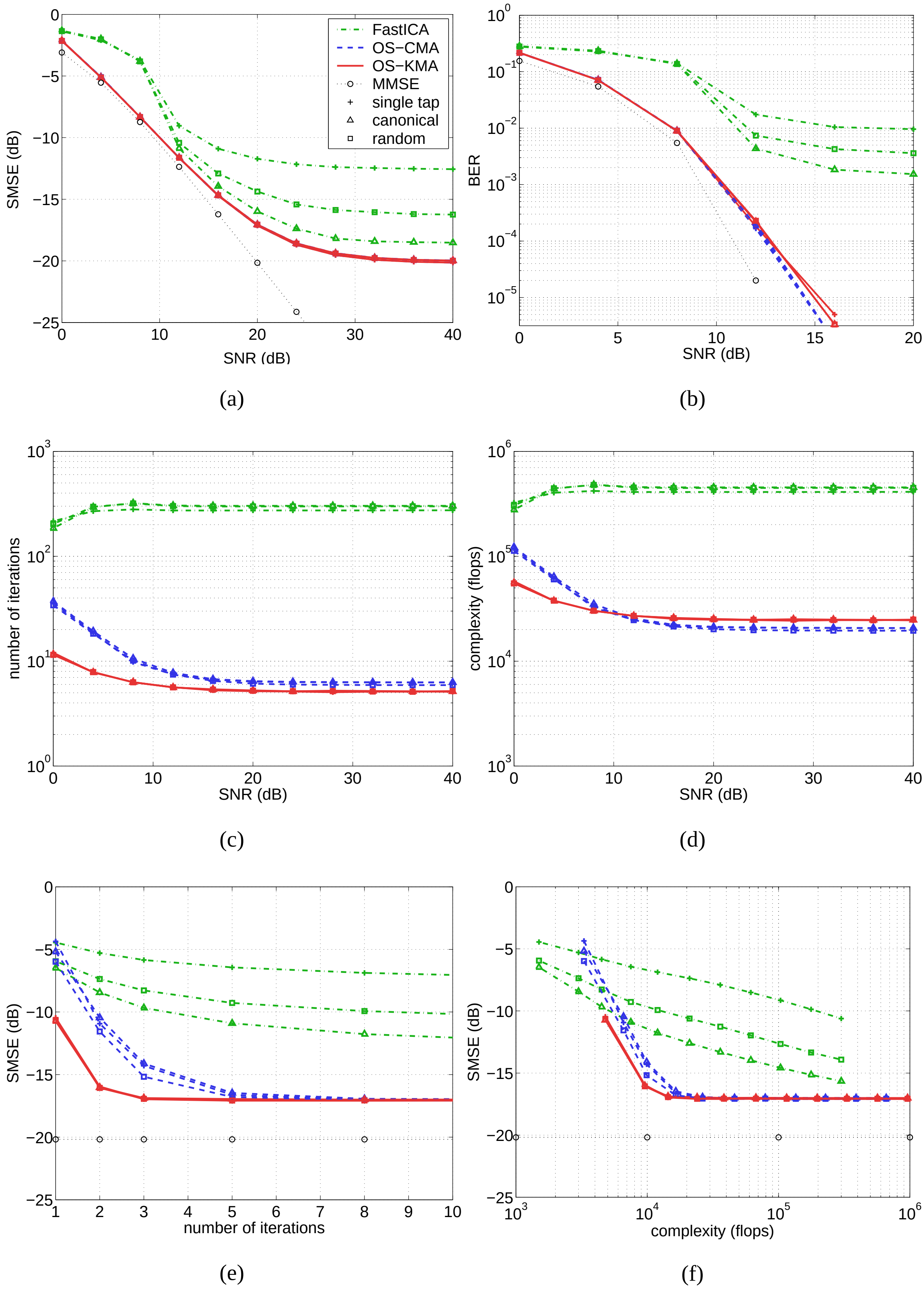


Fig. 4. Average performance of block deflationary ICA algorithms for the different separating vector initializations of Experiments 1–3. (a) Source extraction quality, (b) separator-output bit error rate, (c) number of iterations for convergence, (d) computational complexity, (e) extraction quality against iteration count at 20-dB SNR, (f) extraction quality against computational cost at 20-dB SNR.