A Robust Complex FastICA Algorithm Using the Huber M-Estimator Cost Function

Jih-Cheng Chao[1] and Scott C. Douglas[2]

[1] Semiconductor Group, Texas Instruments, Dallas, Texas 75243, USA
[2] Department of Electrical Engineering, Southern Methodist University, Dallas, Texas 75275, USA

Abstract. In this paper, we propose to use the Huber M-estimator cost function as a contrast function within the complex FastICA algorithm of Bingham and Hyvarinen for the blind separation of mixtures of independent, non-Gaussian, and proper complex-valued signals. Sufficient and necessary conditions for the local stability of the complex-circular FastICA algorithm for an arbitrary cost are provided. A local stability analysis shows that the algorithm based on the Huber M-estimator cost has behavior that is largely independent of the cost function's threshold parameter for mixtures of non-Gaussian signals. Simulations demonstrate the ability of the proposed algorithm to separate mixtures of various complex-valued sources with performance that meets or exceeds that obtained by the FastICA algorithm using kurtosis-based and other contrast functions.

1 Introduction

In complex-valued blind source separation (BSS), one possesses a set of measured signal vectors

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \boldsymbol{\nu}(k), \tag{1}$$

where $\mathbf{A}$ is an arbitrary complex-valued $(m \times m)$ mixing matrix, such that $\mathbf{A} = \mathbf{A}_R + j\mathbf{A}_I$, $\mathbf{s}(k) = [s_1(k) \ \cdots \ s_m(k)]^T$ is a complex-valued signal of sources, and $s_i(k) = s_{R,i}(k) + js_{I,i}(k)$, where $j = \sqrt{-1}$, and $\boldsymbol{\nu}(k)$ contains circular Gaussian uncorrelated noise. In most treatments of the complex-valued BSS task, the $\{s_i(k)\}$ are assumed to be statistically-independent, and $\mathbf{A}$ is full rank. The goal is to obtain a separating matrix $\mathbf{B}$ such that

$$\mathbf{y}(k) = \mathbf{B}\mathbf{x}(k) \tag{2}$$

contains estimates of the source signals. In independent component analysis (ICA), the linear model in (1) may not hold, yet the goal is to produce signal features in $\mathbf{y}(k)$ that are as independent as possible.

One of the most-popular procedures for complex-valued BSS is the complex circular FastICA algorithm in [1]. This algorithm first prewhitens the mixtures $\mathbf{x}(k)$ to obtain $\mathbf{v}(k) = \mathbf{P}\mathbf{x}(k)$ such that $E\{\mathbf{v}(k)\mathbf{v}^H(k)\} = \mathbf{I}$, after which the

M.E. Davies et al. (Eds.): ICA 2007, LNCS 4666, pp. 152–160, 2007.

rows of a unitary separation matrix $\mathbf{W}$ are adapted sequentially such that $\mathbf{y}(k) = \mathbf{W}\mathbf{v}(k)$ contains the separated sources. For mixtures of sources that are proper, such that $E\{s_i^2(k)\} = 0$ for all i, this algorithm appears to separate such complex mixtures given enough snapshots N for an appropriate choice of algorithm nonlinearity. Several algorithm nonlinearities are suggested as possible candidates, although little work has been performed to determine the suitability of these choices for general complex-valued source signals. More recently, several researchers have explored the structure of the complex-valued BSS task for mixtures of non-circular sources, such that $E\{s_i^2(k)\} \neq 0$ [2]–[4]. In what follows, we limit our discussion to the complex-circular source distribution case, as several practical applications involve mixtures of complex-circular sources.

In this paper, we extend our recent work on employing the Huber M-estimator cost function from robust statistics as a FastICA algorithm contrast [5] to the complex-valued BSS task for mixtures of proper sources ($E\{s_i^2(k)\} = 0$). We provide the complete form of the local stability condition for the complex-circular FastICA algorithm omitted in [1]. We then propose a single-parameter nonlinearity for the algorithm and show through both theory and simulations that the algorithm's performance is largely independent of the cost function's threshold parameter for many source distributions, making it a robust choice for separating complex-valued mixtures with unknown circularly-symmetric source p.d.f.'s. Simulations comparing various contrast choices for the complex circular FastICA algorithm show that ours based on the Huber M-estimator cost often works better than others based on kurtosis maximization or heuristic choice.

2 Complex Circular FastICA Algorithm

We first give the general form of the single-unit FastICA algorithm for extracting one non-Gaussian-distributed proper source from an m-dimensional complex linear mixture [1] and study its local stability properties. The algorithm assumes that the source mixtures have been prewhitened by a linear transformation $\mathbf{P}$ where $\mathbf{v}(k) = \mathbf{P}\mathbf{x}(k)$ contains uncorrelated entries, such that the sample covariance of $\mathbf{v}(k)$ is the identity matrix. For the vector $\mathbf{w}_t = [w_{1t} \; \cdots \; w_{mt}]^T$, the complex circular FastICA update is

$$y_t(k) = \mathbf{w}_t^T \mathbf{v}(k) \tag{3}$$

$$\widetilde{\mathbf{w}}_t = E\{y_t(k) g(|y_t(k)|^2)\mathbf{v}^*(k)\} - E\{g(|y_t(k)|^2) + |y_t(k)|^2 g'(|y_t(k)|^2)\}\mathbf{w}_t \tag{4}$$

$$\mathbf{w}_{t+1} = \frac{\widetilde{\mathbf{w}}_t}{\sqrt{\widetilde{\mathbf{w}}_t^H \widetilde{\mathbf{w}}_t}}, \tag{5}$$

where $y_t(k)$ is the estimated source at time k and algorithm iteration t, $g(u)$ is a real-valued nonlinearity, $g'(u) = dg(u)/du$, and the expectations in (4) are computed using N-sample averages. This algorithm is formulated in [1] as the solution to the following optimization problem:

$$\text{maximize} \;\; \left|E\{G(|y_t(k)|^2)\} - E\{G(|n|^2)\}\right|^2 \tag{6}$$

$$\text{such that} \;\; E\{|y_t(k)|^2\} = 1, \tag{7}$$

where n has a circularly-symmetric unit-variance Gaussian distribution and $G(u)$ is a real-valued even-symmetric but otherwise "arbitrary non-linear contrast function" [1] producing $g(u) = dG(u)/du$. The criterion in (6) is described as the square of a simple estimate of the negentropy of $y_t(k)$. Several cost functions are suggested as possible choices for $G(u)$, including $G(u) = \sqrt{a_1 + u}$ for $a_1 \approx 0.1$, $G(u) = \log(a_2 + u)$ for $a_2 \approx 0.1$, and the kurtosis-based $G(u) = 0.5u^2$, although no verification of (9) for the first two choices of $G(u)$ and any well-known non-Gaussian distributions has been given.

In [1], the authors give the following necessary condition for the above algorithm to be locally-stable at a separating solution, where s_i possesses the distribution of the source extracted in $y_t(k)$:

$$(E\{g(|s_i|^2) + |s_i|^2 g'(|s_i|^2) - |s_i|^2 g(|s_i|^2)\}) \neq 0. \tag{8}$$

This condition is not sufficient, however, for local stability of the algorithm, as the curvature of the cost function has not been considered in [1]. Although omitted for brevity, we can show that the necessary and sufficient local stability conditions for the algorithm about a separating solution are

$$\begin{aligned} &[E\{g(|s_i|^2) + |s_i|^2 g'(|s_i|^2) - |s_i|^2 g(|s_i|^2)\}] \\ &\times [E\{G(|s_i|^2)\} - E\{G(|n|^2)\}] < 0. \end{aligned} \tag{9}$$

This result can be compared to that for the real-valued FastICA algorithm in [6], which shows a somewhat-different relationship. Thus, it is necessary and sufficient for the two real-valued quantities on the left-hand-side of the inequality in (9) to be non-zero and have different signs for the complex circular FastICA algorithm to be locally-stable.

3 A Huber M-Estimator Cost Function for the Complex Circular FastICA Algorithm

In [5], a novel single-parameter cost function based on the Huber M-estimator cost in robust statistics [7] was proposed for the real-valued FastICA algorithm. Unlike most other cost functions, the one chosen in [5] has certain nice practical and analytical properties. In particular, it is possible to show that there always exists a nonlinearity parameter for the cost function such that two sufficient conditions for local stability of the algorithm are met. We now extend this work to design a novel cost function for the complex-circular FastICA algorithm.

As the algorithm in [1] implicitly assumes mixtures of proper source signals, we propose to choose $G(|y_t(k)|^2)$ such that the amplitude of $y_t(k)$ is maximized according to the Huber M-estimator cost. Thus, we have

$$G(u) = \begin{cases} \dfrac{u}{2} & u < \theta^2 \\ \theta u^{1/2} - \dfrac{\theta^2}{2} & u \geq \theta^2 \end{cases} \tag{10}$$

where $\theta > 0$ is a threshold parameter designed to trade off the parameter estimation quality with the estimate's robustness to outliers and lack of prior distributional knowledge. The corresponding algorithm nonlinearities are

$$g(u) \equiv \frac{\partial G(u)}{\partial u} = \begin{cases} \frac{1}{2} & u < \theta^2 \\ \frac{\theta}{2} u^{-1/2} & u \geq \theta^2 \end{cases} \tag{11}$$

$$g'(u) \equiv \frac{dg(u)}{du} = \begin{cases} 0 & u < \theta^2 \\ -\frac{\theta}{4} u^{-3/2} & u \geq \theta^2 . \end{cases} \tag{12}$$

After some simplification, we can implement the circular complex FastICA update using the above nonlinearities as

$$\widetilde{\mathbf{w}}_t = 2E\{y_t(k) h_\theta(|y_t(k)|) \mathbf{v}^*(k)\} - E\{t_\theta(|y_t(k)|) + h_\theta(|y_t(k)|)\} \mathbf{w}_t \tag{13}$$

$$h_\theta(u) = \begin{cases} 1 & u < \theta \\ \frac{\theta}{u} & u \geq \theta \end{cases}, \quad t_\theta(u) = \begin{cases} 1 & u < \theta \\ 0 & u \geq \theta . \end{cases} \tag{14}$$

The functions $h_\theta(u)$ and $t_\theta(u)$ depend on the threshold parameter θ, and the choice of this nonlinearity will be considered in the next section. Table 1 lists a short MATLAB script for implementing the multiple-unit version of this algorithm, in which the QR decomposition is used for signal deflation.

Table 1. Complex circular FastICA algorithm with Huber M-estimator cost

```
%--------------------------------------------------------------------
[N,m]=size(x); R = (1/N)*(x'*x); v = x/chol(0.5*(R+R')); W = eye(m);
for i=1:iter
    y = v*W;
    absy = abs(y);
    t = (absy<theta);
    h = t + theta*(1-t)./absy;
    W = 2*(v'*(y.*h)) - W*diag(sum(t+h));
    [W,T] = qr(W);
end
%--------------------------------------------------------------------
```

4 On the Local Stability of the Huber M-Estimator Cost for FastICA

Given the new stability condition in (9), what can be said about the circularly-symmetric Huber M-estimator cost function when it is used in the complex FastICA algorithm? The following two theorems, proven in the Appendix, illustrate two properties about this cost. These theorems make statements about the p.d.f. of $u = |s_i|^2$, the squared amplitude of the extracted source. The theorems are non-trivial extensions of the theorems presented in [5].

Theorem 1: *Let $g(u)$ and $g'(u)$ have the forms in (11) and (12), respectively. Then, so long as the random variable u is not exponentially-distributed, there always exists a value of θ such that*

$$E\{g(u)\} + E\{ug'(u)\} - E\{ug(u)\} \neq 0. \tag{15}$$

Theorem 2: *Let $G(u)$ have the form in (10). Then, so long as the random variable u is not exponentially-distributed, there always exists a value of θ such that*

$$E\{G(u)\} - E\{G(|n|^2)\} \neq 0. \tag{16}$$

Note that if s_i is unit-variance circular Gaussian, the p.d.f. of $u = |s_i|^2$ is exponential ($p(u) = e^{-u}$ for $u \geq 0$). Taken together, these two theorems *do not* ensure (9) for all non-Gaussian proper source distributions. They suggest, however, that the design range for θ could be significant for many distributions. We substantiate this claim through the analysis below and by simulations in the next section. These results are significant because, to our knowledge, few if any statements about the stability of a specific non-kurtosis-based cost function within the complex FastICA algorithm have been given in the scientific literature. Moreover, it is unlikely that such results could be easily found given the complexity of the integrals for other $g(y)$ choices (*e.g.* $g(y) = 0.5(a_1 + y)^{-1/2}$ for $a_1 \approx 1$).

We have evaluated the range of θ values for which (9) is satisfied for five well-known zero-mean, unit-power, non-Gaussian distributions: 4-QAM-$\{\pm 1\} + j\{\pm 1\}$, 16-QAM-$\{\pm\frac{1}{\sqrt{10}} \pm \frac{3}{\sqrt{10}}\} + j\{\pm\frac{1}{\sqrt{10}} \pm \frac{3}{\sqrt{10}}\}$, 64-QAM-$\{\pm\frac{1}{\sqrt{42}} \pm \frac{3}{\sqrt{42}} \pm \frac{5}{\sqrt{42}} \pm \frac{7}{\sqrt{42}}\} + j\{\pm\frac{1}{\sqrt{42}} \pm \frac{3}{\sqrt{42}} \pm \frac{5}{\sqrt{42}} \pm \frac{7}{\sqrt{42}}\}$, the uniform amplitude circular distribution such that $|s_i|$ is equally probable for $0 \leq |s_i| \leq \sqrt{2}$ and is zero otherwise, and the exponential amplitude distribution in which $|s_i|$ is exponentially-distribution with $E\{|s_i|^2\} = 1$. For all of these five distributions, the Huber M-estimator cost produces an algorithm that is locally-stable for θ in the range $[0, |s_{max}|)$, where s_{max} is the maximum possible value of $s_i(k)$ admitted by the source p.d.f. Thus, any positive value of θ that places part of the nonlinear portion of $g(u)$ within the range of $|s_i(k)|^2$ often results in a locally-convergent algorithm. Again, this evaluation does not guarantee that the chosen cost function will always work, but it suggests that one does not need to design specific values of θ to achieve separation.

In practice, one may not know what θ value to choose to obtain separation of a particular source mixture. As was suggested in [5] in the real-valued case, we recommend that one *randomize* the value of θ over a range of positive values during coefficient adaptation. The main observed effect using such randomization is a slight slowdown in convergence speed.

5 Simulations

We now explore the performance of the FastICA algorithm with various cost function via simulations. In these simulations, $m = 15$-source mixtures were

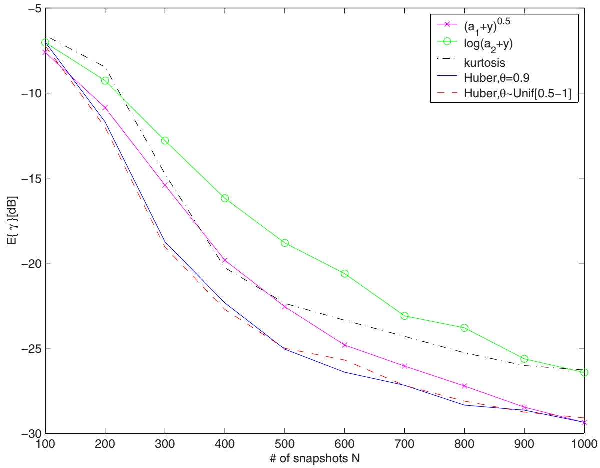

Fig. 1. $E\{\gamma\}$ vs. number of snapshots N for the various algorithms in the simulation example

generated consisting of three 4-QAM, three 16-QAM, three 64-QAM, three uniform and three exponential amplitude circular-distributed independent sources, and a random mixing matrix. The multi-unit FastICA procedure was applied to this data for numbers of snapshots ranging from $N = 100$ to $N = 5000$ and for different θ values. The performance factor computed is the separation cost

$$\gamma = \frac{1}{2m}\left(\sum_{i=1}^{m}\sum_{l=1}^{m}\frac{|c_{il}|^2}{\max\limits_{1\le i\le m}|c_{il}|^2}+\frac{|c_{il}|^2}{\max\limits_{1\le l\le m}|c_{li}|^2}\right)-1 \quad (17)$$

with $\mathbf{C} = \mathbf{WPA}$ as obtained at convergence of the algorithm. One hundred iterations were averaged to obtain each data point shown.

Fig. 1 compares the performance of FastICA with the Huber cost function and $\theta = 0.9$ and with the Huber cost function and a uniformly-randomized θ in the range $0.5 \le \theta \le 1$ at each iteration with three other versions of FastICA – using $G(y) = \sqrt{a_1 + y}$ or $g(y) = \frac{1}{2\sqrt{a_1+y}}$ for $a_1 \approx 0.1$, $G(y) = \log(a_2 + y)$ or $g(y) = \frac{1}{a_2+y}$ for $a_2 \approx 0.1$, and the kurtosis-based choice $G(y) = 0.5y^2$ or $g(y) = y$. As can be seen, the Huber cost function-based versions outperform the algorithms based on previously-proposed contrast functions. More significantly, our algorithm version with a randomized threshold parameter θ provides good separation performance across all sample sizes; performance deviations were less than ±1dB from the algorithm with a fixed $\theta = 0.9$ value.

Fig. 2 illustrates the performance sensitivity of the FastICA algorithm with Huber M-estimator cost to the value of θ for these signal mixtures. As can be seen, the algorithm performs well for values of θ satisfying $0.1 \le \theta \le 1$, and its performance degrades gracefully for higher θ values.

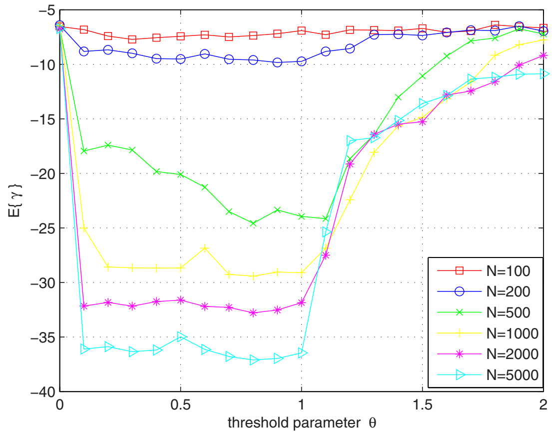

Fig. 2. $E\{\gamma\}$ vs. θ for the FastICA algorithm with Huber M-estimator cost in the simulation example

6 Conclusions

In many blind source separation and independent component analysis algorithms, the cost function used to measure signal independence is a design parameter. In this paper, we have considered Huber's single-parameter M-estimator cost function for use within the complex-valued FastICA algorithm for proper source mixtures. The algorithm obtained is computationally-simple, and the procedure works well for a wide range of threshold parameters θ. The reasons for the algorithm's robust behavior for a wide range of the threshold parameter is indicated through a stability analysis.

References

1. Bingham, E., Hyvarinen, A.: A fast fixed-point algorithm for independent component analysis of complex valued signals. Int J. Neural Systems 10(1), 1–8 (2000)
2. De Lathauwer, L., De Moor, B.: On the blind separation of non-circular sources. In: Proc. EUSIPCO-02, Toulouse, France (September 2002)
3. Novey, M., Adali, T.: ICA by maximization of nongaussianity using complex functions. In: Proc. IEEE Workshop Machine Learning for Signal Processing, Mystic, CT, September 2005, pp. 21–26. IEEE Computer Society Press, Los Alamitos (2005)
4. Eriksson, J., Koivunen, V.: Complex random vectors and ICA models: Identifiability, uniqueness and separability. IEEE Trans. Inform. Theory 52, 1017–1029 (2006)
5. Chao, J., Douglas, S.C.: A simple and robust fastICA algorithm using the Huber M-estimator cost function. In: Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing, Toulouse, France, vol. 5, pp. 685–688 (May 2006)
6. Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Networks 10, 626–634 (1999)
7. Huber, P.: Robust Statistics. Wiley, New York (1981)

7 Appendix

Proof of Theorem 1: Assume without loss of generality that u is unit variance. Consider the terms on the left-hand-side of (15) for the nonlinearities in (11) and (12), and define $f_1(\theta) = 2(E\{g(u)\} + E\{ug'(u)\} - E\{ug(u)\})$. Then, we obtain

$$f_1(\theta) = \int_{\theta^2}^{\infty} u^{-1/2}(u^{3/2} - \theta u - u^{1/2} + \frac{\theta}{2})p(u)du. \tag{18}$$

For Eq. (15) not to hold, $f_1(\theta) = 0$ for all possible values of θ. Suppose that the slightly-more-general condition

$$f_1(\theta) = c_1\theta + c_2 \tag{19}$$

is true, where c_1 and c_2 are unknown constants. Such a condition justified when $p(u)$ is smooth, as $f_1(\theta)$ an then be modeled by a polynomial approximation - see the comment below. Then,

$$\frac{\partial f_1(\theta)}{\partial \theta} = p(\theta^2)\theta + \int_{\theta^2}^{\infty} (\frac{1}{2}u^{-1/2} - u^{1/2})p(u)du = c_1 \tag{20}$$

$$\frac{\partial^2 f_1(\theta)}{\partial \theta^2} = p'(\theta^2) + p(\theta^2) = 0, \tag{21}$$

which yields the relationship

$$p'(u) = -p(u). \tag{22}$$

The only distribution $p(u)$ satisfying (22) is the exponential distribution, i.e. $p(u) = e^{-u}$ for $u \geq 0$. Thus, the theorem follows. Note that if s_i is circular Gaussian-distributed, $|s_i|^2$ has an exponential distribution, although other distributions for s_i could lead to an exponential distribution for $|s_i|^2$.

Proof of Theorem 2: Substituting (10) into the left-hand-side of (16), defining $f_2(\theta) = 2E\{G(u) - G(|n|^2)\}$, and simplifying yields the expression

$$f_2(\theta) = -\int_{\theta^2}^{\infty} (u^{1/2} - \theta)^2[p(u) - p_n(u)]du, \tag{23}$$

where $p_n(u) = e^{-u}$ for $u \geq 0$. For Eq. (16) not to hold, $f_2(\theta) = 0$ for all possible values of θ. Suppose that the slightly-more-general condition

$$f_2(\theta) = c_1\theta + c_2 \tag{24}$$

is true, where c_1 and c_2 are unknown constants. Then,

$$\frac{\partial f_2(\theta)}{\partial \theta} = 2\int_{\theta^2}^{\infty} (u^{1/2} - \theta)[p(u) - p_n(u)]du \ = \ c_1 \tag{25}$$

$$\frac{\partial^2 f_2(\theta)}{\partial \theta^2} = -2\int_{\theta^2}^{\infty} [p(u) - p_n(u)]du \ = \ 0. \tag{26}$$

For (24) to hold for all $\theta > 0$, we must have

$$p(u) = p_n(u), \tag{27}$$

which results in $c_1 = 0$, $c_2 = 0$, and finally $f_2(\theta) = 0$. Thus, the theorem follows.

Comment: In both of the above proofs, $f_i(\theta)$ is a continuous function of θ given a continuous smooth amplitude-squared distribution $p(u)$. Thus, we can express $f_i(\theta)$ as a polynomial function of θ with coefficients c_i. Now, for the condition $f_i(\theta) = 0$, we must have all $c_i = 0$. Clearly, it is impossible that $c_0 = 0$ and all c_i not equal to 0 for $i > 0$ and the condition $f_i(\theta) = 0$, because any change in θ would make $f_i(\theta)$ not equal to zero. Hence, $f_i(\theta)$ defines only one function and therefore only one distribution $p(u)$ has $f_i(\theta) = 0$. In both of the above proofs, the exponential distribution yields $f_i(\theta) = 0$.