

# A STATISTICAL CONVERGENCE ANALYSIS OF THE FASTICA ALGORITHM FOR TWO-SOURCE MIXTURES

Scott C. Douglas

Department of Electrical Engineering  
Southern Methodist University  
Dallas, Texas 75275 USA

## ABSTRACT

While the FastICA algorithm is a popular procedure for independent component analysis (ICA) and blind source separation, its average convergence behavior has yet to be studied. This paper provides several statistical convergence analyses of the kurtosis-based FastICA algorithm for two-source noiseless mixtures. We derive explicit and approximate expressions for the evolutions of the average value and the p.d.f. of the inter-channel interference (ICI) under arbitrary and uniform priors for the initial separating system vector. Our results support the observation in [1]: this version of the FastICA algorithm reduces the average ICI by 1/3 or 4.77dB at each iteration, independent of the source distributions and initial system state. Simulations verify the analytical results.

## 1. INTRODUCTION

The FastICA algorithm of Hyvarinen and Oja is perhaps the most popular block-based technique for blind source separation and independent component analysis (ICA) [2, 3]. The FastICA algorithm employs an approximation to a normalized measure of entropy to extract independent, arbitrarily-distributed sources from a linear mixture. Prewhitening is commonly used as a pre-processing step to aid in the extraction of different sources from the mixture. The procedure is popular because (i) it is simple, (ii) it separates arbitrarily-distributed non-Gaussian sources, (iii) no step size is needed, (iv) it has an effective stopping rule, and (v) it converges quickly – typically, between five and twenty iterations per separation unit are adequate. A number of research communities have adopted FastICA for data analysis tasks.

Good algorithms are deserving of further study, and the FastICA algorithm has been studied in several ways. Its convergence is at least locally-quadratic about a separating solution and cubic when a kurtosis-based measure is employed [3]. FastICA convergence in the undermodeled case is described in [4]. The global fixed-point behavior of two-source FastICA with symmetric orthogonalization is elucidated in [5], and the algorithm's local stationary points for general  $m$ -source mixtures is given in [1]. Cramer-Rao lower bounds for the algorithm are given in [6] for finite sample sizes.

While the above studies explain various performance attributes of the FastICA algorithm, the question still remains:

*Why does the FastICA algorithm converge so fast?* It is known that the kurtosis-based version converges cubically locally [2, 3], but its global and average convergence speed is unknown. A curious observation was made in [1]: the transient behavior of the average inter-channel interference (ICI) at iteration  $t$  is well-described by the “(1/3)rd Rule”

$$E\{ICI_t\} \approx \left(\frac{1}{3}\right) E\{ICI_{t-1}\}. \quad (1)$$

That is, the average ICI decreases by 1/3 or 4.77dB at each iteration, independent of the source distributions, the number of sources, and the distribution of the initial system vector. To our knowledge, no one has shown why this behavior occurs.

This paper contains several analytical justifications of (1) for the FastICA algorithm operating on noiseless non-Gaussian signal mixtures and a kurtosis-based contrast. We consider the simplest possible situation – a single-unit FastICA procedure applied to a noiseless two-source mixture. It is often possible in simple cases to exactly describe quantitative algorithm behaviors that qualitatively describe how the algorithms perform in more-realistic scenarios; moreover, the two-source separation task is solved at the  $(m - 1)$ th stage of the  $m$ -dimensional FastICA algorithm. We derive the following results:

1. An exact expression for the p.d.f. of  $ICI_t$  given an arbitrary p.d.f. for  $ICI_0$  for equal-kurtosis sources;
2. An exponentially-tight expression for  $E\{ICI_t\}$  for an arbitrary (smooth) p.d.f. for  $ICI_0$  for equal-kurtosis sources; and
3. Exact and limiting expressions for  $E\{ICI_t\}$  for a uniformly-distributed look direction for the initial separating weight vector for *arbitrary-kurtosis* sources.

In every case above, the “(1/3)rd Rule” fits these analytical predictions of FastICA algorithm behavior. These results are the first steps towards characterizing FastICA algorithm behavior for arbitrary  $m$ -source mixtures and contrasts.

For notational simplicity, we now describe the FastICA algorithm in the two-source case; general descriptions can be found in [3]. Let  $\mathbf{s}(k) = [s_1(k) \ s_2(k)]^T$ , where  $s_1(k)$  and  $s_2(k)$  are statistically-independent with zero means, unit variances, and non-zero (normalized) kurtoses  $\kappa_i = E\{s_i^4(k)\}/$

$(E\{s_i^2(k)\})^2 - 3$ ,  $i \in \{1, 2\}$ . Two-dimensional mixtures  $\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k)$ ,  $1 \leq k \leq N$  are measured. These mixtures are first prewhitened by a matrix  $\mathbf{P}$  such that  $\mathbf{v}(k) = \mathbf{P}\mathbf{x}(k)$  has uncorrelated elements with unit variances, where  $\mathbf{P}$  is calculated from the sample covariance matrix of  $\mathbf{x}(k)$ . Then, a two-tap separating system vector  $\mathbf{w}_t = [w_{1,t} \ w_{2,t}]^T$  is adjusted according to the FastICA update

$$y_t(k) = \mathbf{w}_t^T \mathbf{v}(k) \quad (2)$$

$$\tilde{\mathbf{w}}_t = E\{y_t^3(k)\mathbf{v}(k)\} - E\{y_t^2(k)\}\mathbf{w}_t \quad (3)$$

$$\mathbf{w}_{t+1} = \frac{\tilde{\mathbf{w}}_t}{\sqrt{\tilde{\mathbf{w}}_t^T \tilde{\mathbf{w}}_t}} \quad (4)$$

where data averages are used to compute the expectations in (3). In this paper, infinite data records are assumed, such that the expectations in (3) are exact. In later sections, we consider the equivalent expressions for (2)–(4) in the combined coefficient space  $\mathbf{c}_t = \mathbf{A}^T \mathbf{P}^T \mathbf{w}$ , in which the parametrization

$$\mathbf{c}_t = [\cos(\theta_t) \ \sin(\theta_t)]^T \quad (5)$$

for a two-coefficient system is particularly useful, where  $\theta_t$  is effectively “adjusted” by the FastICA procedure. The goal of source separation is to adjust  $\theta_t$  such that  $\theta_t \rightarrow \theta_{opt} \in \{0, \pi/2, \pi, 3\pi/2\}$ . We only consider  $0 \leq \theta_t \leq \pi/2$  as the behaviors in the other three quadrants are identical.

To characterize the convergence performance, we shall use the scale-free *inter-channel interference* (ICI) measure

$$ICI_t = \frac{\left(\sum_{i=1}^m c_{i,t}^2\right) - \max_{1 \leq j \leq m} c_{j,t}^2}{\max_{1 \leq j \leq m} c_{j,t}^2}. \quad (6)$$

Because  $ICI_t$  depends on the squared values  $\{c_{i,t}^2\}$  of the combined system coefficients and due to the nature of the data-averaged form of the FastICA update, the evolution of  $ICI_t$  does not depend on the signs of the kurtoses of the sources. Without loss of generality, assume  $\kappa_i > 0$ , or equivalently, replace  $\kappa_i$  by  $|\kappa_i|$  in the analysis equations.

## 2. STUDYING THE AVERAGE CONVERGENCE PERFORMANCE OF THE FASTICA ALGORITHM

Consider the behavior of the FastICA algorithm for infinite data under a *coefficient-stochastic* setting, in which  $\mathbf{w}_0$  or  $\mathbf{c}_0$  or  $\theta_0 = \arctan(c_{2,0}/c_{1,0})$  is unknown and assumed random. The average performance of the algorithm thus depends on how this initial condition affects the evolution of the ICI performance measure in (6). This analysis technique is well-known in the adaptive filtering community and has led to a number of convergence results for various algorithms [7].

As can be shown from [2] (see also [1]), the evolutionary behavior of the FastICA algorithm for infinite data is

$$c_{i,t} = \frac{\kappa_i c_{i,t-1}^3}{\sqrt{\kappa_1^3 c_{1,t-1}^6 + \kappa_2 c_{2,t-1}^6}} \quad (7)$$

for  $i \in \{1, 2\}$ . Thus, we have

$$\frac{c_{2,t}^2}{c_{1,t}^2} = \left(\frac{\kappa_2}{\kappa_1}\right)^{3^t-1} \left(\frac{c_{2,0}^2}{c_{1,0}^2}\right)^{3^t} = \frac{\kappa_1}{\kappa_2} \left(\sqrt{\frac{\kappa_2}{\kappa_1}} \tan(\theta_0)\right)^{2(3^t)} \quad (8)$$

An expression for  $ICI_t$  in (6) can be obtained from (8) as  $ICI_t = \min[(c_{2,t}^2/c_{1,t}^2), (c_{1,t}^2/c_{2,t}^2)]$ , which translates into

$$ICI_t = \begin{cases} \frac{\kappa_1}{\kappa_2} \left(\sqrt{\frac{\kappa_2}{\kappa_1}} \tan(\theta_0)\right)^{2(3^t)}, & 0 \leq \theta \leq \alpha_t \\ \frac{\kappa_2}{\kappa_1} \left(\sqrt{\frac{\kappa_1}{\kappa_2}} \cot(\theta_0)\right)^{2(3^t)}, & \alpha_t < \theta \leq \pi/2 \end{cases} \quad (9)$$

where

$$a = \sqrt{\frac{\kappa_1}{\kappa_2}} \quad \text{and} \quad \alpha_t = \arctan\left(a \cdot (a^{-1})^{\frac{1}{3^t}}\right). \quad (10)$$

From (9), we see that  $ICI_t$  depends on  $\theta_0$ ,  $\kappa_2/\kappa_1$ , and  $t$ . For a given distribution  $\bar{p}_0(\theta)$  on  $\theta_0$ , (9) induces a distribution  $p_t(u)$  of the ICI at iteration  $t$ . Consider the following cases:

1. *Equal-(Magnitude)-Kurtosis Sources*: For  $\kappa_1 = \kappa_2$ ,  $\alpha_t = \pi/4$  for all  $t$ . If  $\theta_0$  is distributed symmetrically about  $\theta_0 = \pi/4$ ,  $ICI_t$  obeys the *scalar evolutionary equation*

$$ICI_t = (ICI_{t-1})^3 = (ICI_0)^{3^t}. \quad (11)$$

The behavior of  $ICI_t$  is completely determined by the distribution  $p_0(u)$  of  $ICI_0$ .

2. *Arbitrary-(Magnitude)-Kurtosis Sources*: Here, it is easiest to explore the distribution of  $ICI_t$  and its moments, such as  $E\{ICI_t\}$ , using the p.d.f.  $\bar{p}_0(\theta)$  directly.

In either case, the distributions  $p_0(u)$  or  $\bar{p}_0(\theta)$  represent our uncertainty about the the rows of the mixing matrix. Setting  $p_0(u) = \delta(u)$ ,  $\bar{p}_0(\theta) = \delta(\theta)$  or  $\bar{p}_0(\theta - \pi/2) = \delta(\theta)$  would make  $E\{ICI_t\} = 0$  for all  $t$ .

## 3. THE P.D.F. OF THE INTER-CHANNEL INTERFERENCE FOR EQUAL-KURTOSIS SOURCES

For equal-kurtosis sources,  $ICI_t$  is related to  $ICI_0$  in (11) through an invertible function. Thus,  $p_t(u)$  can be computed from  $p_0(u)$ , as the following theorem and corollary show.

**Theorem 1:** *Let the initial inter-channel interference  $ICI_0$  have p.d.f.  $p_0(u)$ . Then, the p.d.f.  $p_t(u)$  of  $ICI_t$  is*

$$p_t(u) = \left(\frac{1}{3}\right)^t p_0\left(u^{\frac{1}{3^t}}\right) \frac{1}{u^{1-\frac{1}{3^t}}}. \quad (12)$$

*Corollary 1.1:* *Let the initial separating weight vector have an angular distribution  $\bar{p}_0(\theta)$  over  $0 \leq \theta \leq \pi/4$ . Then,*

$$p_t(u) = \left(\frac{1}{3}\right)^t \bar{p}_0(\arctan(u^{\frac{1}{2 \cdot 3^t}})) \frac{1}{2} \frac{1}{1+u^{\frac{1}{3^t}}} \frac{1}{u^{1-\frac{1}{2 \cdot 3^t}}} \quad (13)$$

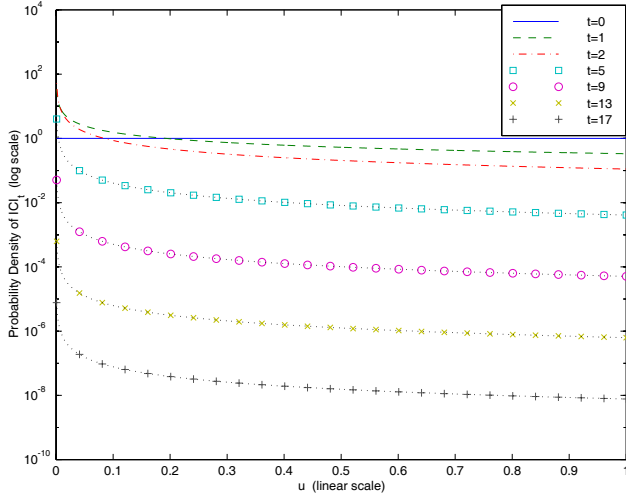


Fig. 1: Probability density functions  $p_t(u)$  for a uniform-[0, 1] p.d.f. for  $ICI_0$ , equal-kurtosis source mixtures.

*Proof:* Let  $u_t = g_t(u_0)$ , where  $u_0$  is an r.v. with p.d.f.  $p_0(u)$ . Then, the p.d.f. of  $u_t$  is

$$p_t(u) = p_0(g_t^{-1}(u)) \frac{dg_t^{-1}(u)}{du}, \quad (14)$$

where  $u_0 = g_t^{-1}(u_t)$  is the inverse function of  $u_t = g_t(u_0) = u_0^{3^t}$ . Since  $g_t^{-1}(u) = u^{(1/3^t)}$ , we obtain the p.d.f. in (12). For the corollary, consider the distribution of  $p_0(u)$  that results from  $\mathbf{c}_0 = [\cos(\theta_0) \sin(\theta_0)]^T$  having a p.d.f.  $\bar{p}_0(\theta)$  on the unit arc  $\theta_0 \in [0, \pi/4]$ . Using  $ICI_0 = \tan^2(\theta_0)$  and (14),

$$p_0(u) = \bar{p}_0(\arctan(u^{1/2})) \frac{1}{2} \frac{1}{1+u} u^{-1/2}. \quad (15)$$

Substituting this expression into (12), we obtain (13).

*Implications:* In the equal-kurtosis case, the p.d.f. of  $ICI_t$  quickly approaches the function

$$p_t(u) \sim \left(\frac{1}{3}\right)^t \frac{K_t}{u}, \quad \lim_{t \rightarrow \infty} K_t = \lim_{u \rightarrow 1^-} p_0(u). \quad (16)$$

The p.d.f.  $ICI_0$  only appears to affect that of  $ICI_t$  through its limiting value near  $ICI_0 = 1$ , as  $p_0(u^{(1/3^t)})$  quickly approaches  $\lim_{u \rightarrow 1^-} p_0(u)$  for all  $0 < u \leq 1$  and values of  $t$  greater than 4. Even for small values of  $t$ , the p.d.f. of  $ICI_t$  is highly-skewed towards zero, indicating fast convergence of the ICI. Moreover, if  $\lim_{u \rightarrow 1^-} p_0(u) = 0$ , then convergence is clearly faster than linear.

Plots of  $ICI_t$  are now used to justify the above claims. Shown in Figure 2 are the p.d.f.'s  $p_t(u)$  of  $ICI_t$  for different values of  $t$  where  $ICI_0$  is uniformly-distributed over  $[0, 1]$ . For values of  $t \geq 5$ , the shape of the p.d.f. is close to  $(1/3)^t$  with an amplitude that decreases as  $(1/3)^t$ .

The ICI distribution for even moderate values of  $t$  is relatively insensitive to the p.d.f. of  $ICI_0$ . Consider a p.d.f. for

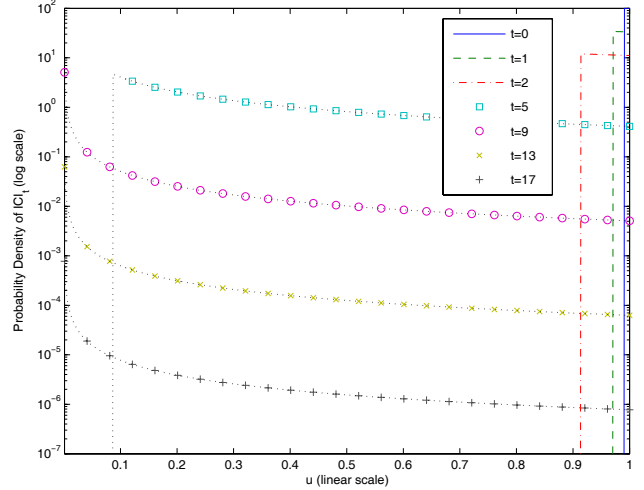


Fig. 2: Probability density functions  $p_t(u)$  for a challenging p.d.f. for  $ICI_0$ , equal-kurtosis source mixtures.

$ICI_0$  that is highly-skewed towards the saddle point  $\theta_0 = \pi/4$  or  $ICI_0 = 1$ , as given by

$$p_0(u) = \begin{cases} 100 & 0.99 \leq u \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

Using (12), we can plot  $ICI_t$  for different  $t$ . Shown in Figure 4 are the results. After a few iterations, the ICI is distributed nearly over all of  $[0, 1]$  with a shape that is  $\sim 1/u$ .

#### 4. THE AVERAGE INTER-CHANNEL INTERFERENCE FOR EQUAL-KURTOSIS SOURCES

We now evaluate  $E\{ICI_t\}$  for equal-kurtosis sources. If the p.d.f. of  $\theta_0$  is known, we can obtain an exact result. The behavior of  $p_t(u)$  observed in the last section, however, suggests that a general approximate expression for  $E\{ICI_t\}$  could be found. The following theorem gives this expression.

**Theorem 2:** Suppose the p.d.f.  $p_0(u)$  of  $ICI_0$  is non-zero and well-behaved at  $ICI_0 = 1$ , such that the left-sided derivatives

$$p_0^{(i)}(u) = \lim_{\Delta \rightarrow 0} \frac{p_0^{(i-1)}(u - \Delta) - p_0^{(i-1)}(u)}{\Delta} \quad (18)$$

exist and are finite at  $u = 1$  for  $i \in \{1, 2\}$ . Then, the average ICI at iteration  $t$  obeys

$$E\{ICI_t\} = E\{\widehat{ICI}_t\} + \mathcal{O}\left(\left[\frac{1}{9}\right]^t\right) \quad (19)$$

$$E\{\widehat{ICI}_t\} = \left(\frac{1}{3}\right)^t p_0(1). \quad (20)$$

*Proof:* Consider the average value of  $ICI_t$  in (11) over  $p_0(u)$ , as given by

$$E\{ICI_t\} = \int_0^1 \zeta^{3^t} p_0(\zeta) d\zeta = \int_0^\infty e^{-(3^t+1)x} p_0(e^{-x}) dx, \quad (21)$$

where we have used the transformation  $\zeta = e^{-x}$ . Integrating by parts using  $u = p_0(e^{-x})$  and  $dv = e^{-(3^t+1)x} dx$  produces

$$E\{ICI_t\} = \frac{p_0(1)}{3^t+1} - \frac{1}{3^t+1} \int_0^\infty e^{-(3^t+2)x} p_0^{(1)}(e^{-x}) dx. \quad (22)$$

The integral on the right-hand-side of (22) can also be integrated by parts, such that (21) becomes

$$E\{ICI_t\} = \frac{p_0(1)}{3^t+1} + \frac{1}{(3^t+1)(3^t+2)} \left[ p_0^{(1)}(1) - \int_0^\infty e^{-3^{t+1}x} p_0^{(2)}(e^{-x}) dx \right]. \quad (23)$$

The results of the theorem follow directly from (23).

*Implications:* The value of the p.d.f. of  $ICI_0$  near  $ICI_0 = 1$  determines the limiting behavior of  $E\{ICI_t\}$ . In practice, the average ICI follows the “(1/3)rd Rule” because one cannot guarantee that  $\mathbf{w}_0$  is bounded away from a saddle point.

Although (19)–(20) applies for an arbitrary p.d.f. for  $ICI_0$ , we can confirm the “(1/3)rd Rule” for a reasonable p.d.f. choice. Assume that  $\mathbf{w}_0$  is uniformly-distributed on the unit circle, or  $\bar{p}_0(\theta) \sim \text{Unif}[0, \pi/4]$ . Then we have the following result.

**Theorem 3:** For a uniform prior on the angular direction of  $\mathbf{w}_0$ , the average ICI at iteration  $t$  for  $\kappa_1 = \kappa_2$  is exactly

$$E\{ICI_t\} = -1 + \frac{4}{\pi} \sum_{j=0}^{3^t-1} \frac{1}{2j+1} (-1)^j. \quad (24)$$

*Proof:* Using the parametrization (5), we can express

$$E\{ICI_t\} = \frac{4}{\pi} \int_0^{\pi/4} \tan^{2(3^t)}(\theta) d\theta. \quad (25)$$

By recursive application of the formula

$$\int \tan^i(\theta) d\theta = \frac{\tan^{i-1}(\theta)}{i-1} - \int \tan^{i-2}(\theta) d\theta, \quad (26)$$

one can show that  $E\{ICI_t\}$  has the series expansion in (24).

*Implications:* Eqn. (24) is easily evaluated numerically, and we are most-often interested in small  $t$  values (e.g.  $0 < t \leq 15$ ). Numerical evaluation of this formula shows that  $E\{ICI_0\} = -1 + 4/\pi$  and

$$E\{ICI_t\} \approx E\{\widehat{ICI}_t\} = \frac{1}{\pi} \left(\frac{1}{3}\right)^t, \quad (27)$$

for large  $t$ , which is what (20) predicts in this case (i.e.  $\bar{p}_0(\theta) = 4/\pi$  results in  $p_0(1) = 1/\pi$  from (13)).

## 5. THE AVERAGE INTER-CHANNEL INTERFERENCE FOR ARBITRARY SOURCES

In practice, the sources being separated by the FastICA algorithm will have different distributions and therefore different

kurtosis (magnitudes). Suppose that  $\mathbf{w}_0$  is again uniformly-distributed on the unit circle, representing a uniform prior for the coefficient vector. Then, we have the following result.

**Theorem 4:** For a uniform distribution on the separation system direction, a limiting expression for the average ICI for a mixture of two sources with kurtoses  $\kappa_1$  and  $\kappa_2$  is

$$E\{\widehat{ICI}_t\} = \frac{2}{\pi \left( \sqrt{\frac{\kappa_1}{\kappa_2}} + \sqrt{\frac{\kappa_2}{\kappa_1}} \right)} \left(\frac{1}{3}\right)^t \quad (28)$$

*Proof:* Our starting point is (9). Letting  $\bar{p}_0(\theta) = 2/\pi$  for  $0 \leq \theta \leq \pi/2$ ,

$$E\{ICI_t\} = \frac{2}{\pi} \left\{ \int_0^{\alpha_t} a^2 (a^{-1} \tan \theta)^{2(3^t)} d\theta + a^{-2} \int_{\alpha_t}^{\pi/2} (a \cot \theta)^{2(3^t)} d\theta \right\} \quad (29)$$

Define  $\alpha_t = \tan^{-1}(a_t) = a \cdot (a^{-1})^{\frac{1}{3^t}}$ . Using  $\cot \theta = \tan(\frac{\pi}{2} - \theta)$  and  $\frac{\pi}{2} - \arctan(a_t) = \arctan(a_t^{-1})$ , (29) is

$$E\{ICI_t\} = \frac{2}{\pi} \left\{ \int_0^{\arctan(a_t)} a^2 (a^{-1} \tan \theta)^{2(3^t)} d\theta + a^{-2} \int_0^{\arctan(a_t^{-1})} (a \tan \theta)^{2(3^t)} d\theta \right\} \quad (30)$$

By the change of variables  $u = b^{-1} \tan \theta$ , one can show for any positive constants  $b$  and  $b_t = b \cdot (b^{-1})^{1/(3^t)}$

$$\int_0^{\arctan(b_t)} (b^{-1} \tan \theta)^{2(3^t)} d\theta = \int_0^{(b^{-1})^{1/(3^t)}} \frac{u^{2(3^t)}}{b^{-1} + bu^2} du. \quad (31)$$

As  $t \rightarrow \infty$ , the area underneath the integrand in (31) is concentrated at  $u = (b^{-1})^{1/(3^t)}$ , such that we may approximate

$$\begin{aligned} & \int_0^{(b^{-1})^{1/(3^t)}} \frac{u^{2(3^t)}}{b^{-1} + bu^2} du \\ & \approx \frac{1}{b^{-1} + b \cdot (b^{-2})^{1/(3^t)}} \int_0^{(b^{-1})^{1/(3^t)}} u^{2(3^t)} du \quad (32) \\ & = \left( \frac{1}{b^{-1} (b)^{\frac{1}{3^t}} + b (b^{-1})^{\frac{1}{3^t}}} \right) \frac{b^{-2}}{2(3^t) + 1}. \quad (33) \end{aligned}$$

Using (33) to approximate (29) yields

$$E\{\widehat{ICI}_t\} = \frac{1}{\pi} \left( \frac{2}{a^{-1} (a)^{\frac{1}{3^t}} + a (a^{-1})^{\frac{1}{3^t}}} \right) \frac{2}{2(3^t) + 1} \quad (34)$$

As  $t \rightarrow \infty$ , we find that both  $a^{1/(3^t)}$  and  $(a^{-1})^{1/(3^t)}$  quickly tend to unity, and  $2(3^t) \gg 1$ . Thus, (34) simplifies to (28).

*Implications:* Several points can be emphasized.

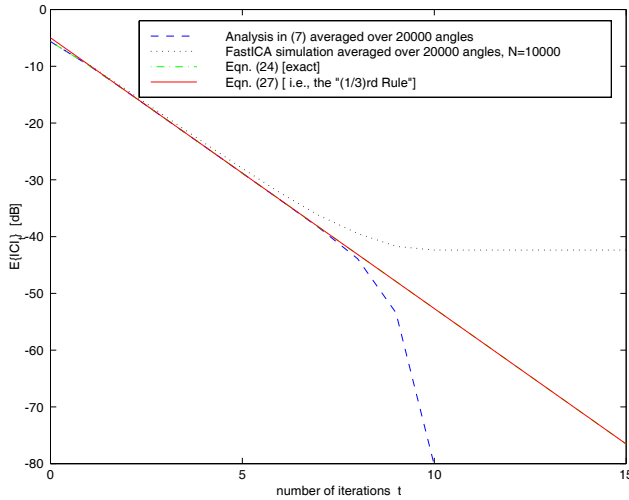


Fig. 3: Evolutions of  $E\{ICI_t\}$  from predictions and simulations, equal-kurtosis source mixture.

1. Eqn. (28) generalizes (27) for arbitrary-kurtosis mixtures. It also follows the “(1/3)rd Rule,” *i.e.* the asymptotic convergence rate of the FastICA algorithm is *unaffected* by the kurtosis magnitudes.

2. The FastICA algorithm is more likely to extract large-magnitude-kurtosis sources than smaller-magnitude-kurtosis ones due to the form of the integral in the proof in (29), in which  $\alpha_t > \pi/4$  when  $\kappa_1 > \kappa_2$ .

3. The average ICI is uniformly lower when the kurtosis magnitudes of the sources differ.

4. If one of the sources in a two-source mixture has a zero kurtosis, the FastICA algorithm theoretically provides one-step convergence to  $E\{ICI_1\} = 0$  with infinite data. In practice, numerical effects limit its convergence speed.

## 6. SIMULATIONS

MATLAB simulations are used to check the results of previous sections. In each case, theoretical expressions of  $E\{ICI_t\}$  were compared to two ensemble averages of  $ICI_t$  in (6), in which  $\mathbf{c}_t$  was generated (i) using the analysis equation in (7) and (ii) using the FastICA algorithm applied to  $N = 10000$  snapshots of mixtures of binary and/or uniformly-distributed sources. In both cases,  $M$  different initial  $\mathbf{c}_0$  vectors spaced uniformly on the unit arc over  $[\pi/(4M), \pi/2(1 - 1/(2M))]$  have been employed.

Fig. 3 shows plots of  $E\{ICI_t\}$  as predicted by (24) and (27) along with ensemble averages of  $ICI_t$  using the analysis equation and FastICA, respectively, for equal-kurtosis binary source mixtures, where  $M = 20000$ . All of the expressions are similar up to  $t = 7$ . For  $t > 7$ , the averaged value of  $E\{ICI_t\}$  computed by the theoretical analysis in (7) begins to converge cubically because  $M < \infty$ . The FastICA algorithm, however, exhibits a limiting  $E\{ICI_t\}$  value because of finite data records (*e.g.*  $N < \infty$ ), and thus the (deterministic)

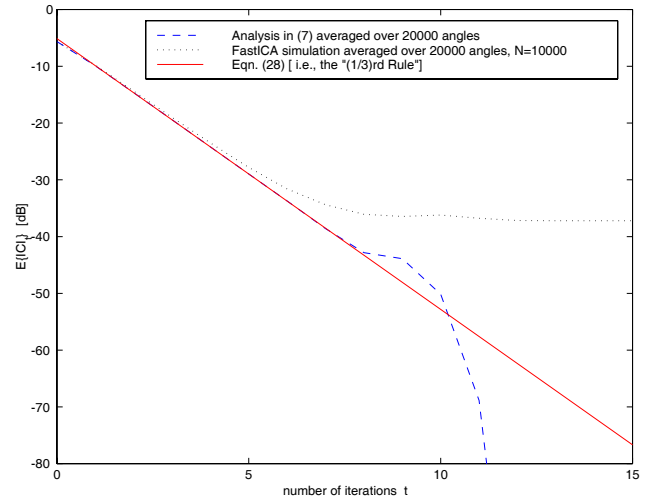


Fig. 4: Evolutions of  $E\{ICI_t\}$  from predictions and simulations, arbitrary-kurtosis source mixture.

cubic convergence of FastICA is not observed on average.

Fig. 4 considers the above experiment for mixtures of binary and uniformly-distributed sources, where  $|\kappa_1| = 2$  and  $|\kappa_2| = 6/5$ . The prediction in (28) accurately describes the behavior of the FastICA algorithm during its convergence period in this situation, and finite-sample effects again limit the FastICA algorithm’s steady-state performance.

## 7. CONCLUSIONS

We have carefully analyzed the transient convergence behavior of the popular FastICA algorithm for noiseless two-dimensional equal- and arbitrary-kurtosis source mixtures. The results confirm the empirical “(1/3)rd Rule” observed in [1], indicating that this algorithm’s fast convergence behavior is independent of the source distributions, the number of sources, and the distribution of the initial system vector. Extensions of these results to the general  $m$ -source separation case can be found in [8].

**Acknowledgement:** The author would like to thank Zhijian Yuan and Erkki Oja for various helpful discussions.

## 8. REFERENCES

- [1] S.C. Douglas, “On the convergence behavior of the FastICA algorithm,” *Proc. Fourth Symp. Indep. Compon. Anal. Blind Signal Separation*, Kyoto, Japan, pp. 409-414, Apr. 2003.
- [2] A. Hyvärinen and E. Oja, “A fast fixed-point algorithm for independent component analysis,” *Neural Computation*, vol. 9, no. 7, pp. 1483-1492, Oct. 1997.
- [3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis* (New York: Wiley, 2001).
- [4] P.A. Regaliá and E. Kofidis, “Monotonic convergence of fixed-point algorithms for ICA,” *IEEE Trans. Neural Networks*, vol. 14, pp. 943-949, July 2003.
- [5] E. Oja, “Convergence of the symmetrical FastICA algorithm,” *Proc. 9th Int. Conf. Neural Inform. Processing*, Singapore, Nov. 2002.
- [6] Z. Koldovsky, P. Tichavsky, and E. Oja, “Cramer-Rao lower bound for linear independent component analysis,” *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Philadelphia, PA, vol. 3, pp. 581-584, Mar. 2005.
- [7] W.A. Gardner, “Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique,” *Signal Processing*, vol. 6, no. 2, pp. 113-133, Apr. 1984.
- [8] S.C. Douglas, Z. Yuan, and E. Oja, “Average convergence behavior of the FastICA algorithm for blind source separation,” *Proc. 6th Int. Conf. Indep. Compon. Anal. Blind Signal Separation*, Charleston, SC, Mar. 2006 (in press).