Analysis of ICA techniques in terms of Failure percentage and Average CPU Time for Real World BSS Task

Naveen Dubey[#1], Rajesh Mehra[*2]

[#]ME Scholar, Department of Electronics & Communication Engineering, National Institute of Technical Teachers Training and Research, Chandigarh, India, 160019

Address Including Country Name

Abstract — *Taking assumption the hidden sources are statistically independent, Independent component analysis technique separates these sources from a linear mixture of audio signals, communication signals generated by equally spaced independent audio sources. Since, in Audio applications source exhibit non - dependence. Mutual information minimization corresponds to minimization of entropy, that ensures quality of separation and exploits non-Gaussianity, non-circularity and sample dependence simultaneously. In this paper these properties are exploited with the help of Cramer-Rao lower bound, modified convex divergence based ICA, Fast ICA and JADE. The performance of these techniques are examined with the help of a number of example and a comparative analysis presented in term of failure percentage and average CPU time taken for execution.*

Keywords— *Independent Component Analysis, Blind Source Separation, Convex Divergence, Independence, non-Gaussianity,.*

I. INTRODUCTION

Independent Component analysis is a very promising technique for blind source separation. ICA estimates original source signal from a convolutive mixture with some very simple ambiguities like, permutation and scaling ambiguity. Audio Source Separation in blind scenario considered as a standard source separation problem that is similar to identification of a linear system and a inversion problem of mixing parameters [1]. Here primary assumption is sources are stationary and no noise component is involved. Assuming there are 'n' number of audio sources and 'm' microphones are there for recording. The mixture signals $X(t)= [x_1(t), ... ,x_m(t)]^T$ observed at any time instant 't' and source signals are $S(t) = [s_1(t), ... , s_n(t)]^T$. If we consider A(t) as mixing matrix then, it can be formulated as;

$$X(t) = A(t)\otimes S(t) \qquad (1)$$

There are many version of ICA algorithms are available for handling the source separation problem, the joint approximate diagonalization of eigenmatrices (JADE) [2]. Convex divergence ICA [3], complex Fast ICA [4], weight initialized ICA [5] , The Gaussian entropy rate minimization Algorithm [6], Kurtosis Maximization technique [7]. Convolutive bounded component analysis algorithm [8].

In this paper basic fundamentals of three algorithms Cramer-Rao Lower Bound based entropy rate bound minimization (ERBM), Modified convex divergence ICA and Weight initialized fast ICA is discussed. CRLB is a derived general form of Fisher information matrix (FIM) for exploiting the non-Gaussianity, non-circularity and sample dependence [9]. In this technique in place of entropy rate minimization which is equivalent to minimization of mutual information rate, a semi-parametric method based entropy rate bound is minimized. In weight initialized Fast ICA a sophisticated procedure is used for weight initialization as well unlike general Fast ICA technique . In this technique mean removal (centring) is applied over randomly initialized weight matrix along with mean removal of observed data. The modified convex divergence ICA is an optimized alternative of C-ICA, in this technique weight initialization is included unlike C-ICA and a modified convex divergence measure is included for weight updation. Natural gradient descent based unsupervised learning technique is implemented.

Performance of these ICA techniques are analyzed by examples of audio mixtures and comparative analysis presented for Failure percentage and Average time required to CPU. Rest of this paper is organized as follows; section II comprises introduction of ERBM using CRLB , Section III describes weight initialized Fast-ICA, Section IV consist of detailed description of modified convex divergence based ICA, Simulation results and discussion are included in section V followed by conclusion drawn in section VI.

II. ENTROPY RATE BOUND MINIMIZATION

Entropy rate bound minimization is a technique with ability to minimize the ill effects of model mismatch. Entropy rate bound minimization is a technique that exploits non-Gaussianity, no-circularity and sample dependence jointly. Here we are considering a random vector containing real part and imaginary part, V = VR+ j VI, where VR and VI are real and imaginary part of the vector respectively. Considering the vector have zero mean E[V] = 0. The probability density function and Entropy rate of 'V' is formulated as;

p(V) ≅ p (VR, VI) (2)

Hr(V) ≅ limit T→∞ E{-log p (VR, VI) }/ T (3)

These definitions are valid for assumption of stationary condition. An unifying framework that can include three diversity non-Gaussianity, non-circularity and sample dependence can be jointly included by generalization of mutual information for the decomposition of independent components. This can be expressed as;

$$\Gamma \text{ (x1, . . ., xN)}=\sum_{k=1}^{N} H(x_K) - \log(|W|^2) - H(V) \qquad (4)$$

Where H(xK) is entropy rate of the kth process and entropy rate H(V) is the observation vector value of 'V' with respect to fixed W. So, the cost function can be written in terms of 'W'

$$\tau_r(W) = \sum_{k=1}^{N} H(x_K) - \log(|W|^2) \qquad (5)$$

The Cramer-Rao lower bound for weight W can be derived by analysis of Fisher information matrix (FIM) by incorporating the condition for which FIM is non-singular and CRLB is finite. For the entropy rate estimation, here we consider existence of a whitening filter parameter 'l' such that u(t) = lH x(i+1)(t)= pH x(i+1)+ qH x*(i+1)(t), where p = [p0 . . . . . pi]T, q= [q0 . . . .,qi]T and 'l'=[pT qT] . The optimized whitening filter coefficient can be estimated by;

$$\min_{l} H(u), s.t. \left\| p_0 \right|^2 - \left| q_0 \right|^2 \right| = 1 \qquad (6)$$

Equation (6) ensures that input and output must have same entropy [10]. Now using the cost function of mutual information rate from equation (5) an entropy rate estimator can be derived. In this technique instead of minimizing the cost function with respect to unmixing matrix W, decoupling technique can be implemented for minimizing the cost function with respect to row vectors of unmixing matrix [11]. Further optimization achieved by applying gradient based updation rule [12].Entropy rate bound minimization can be achieved as follows;

1. Load input data vector 'V'
2. Whiten data 'V'
3. Initialize unmixing matrix W using entropy rate bound with gradient based updation.
4. Calculate cost function Γr (W) using equation(5).
5. Initialize do-while loop for a fixed number of iteration for
   Y = W. V
6. Initialize a nested do-while loop and initialize whitening coefficient 'l'

If lite == 0;
'redefine l
End.

7. Calculate gradient ∂Γr (wk) /∂wk , where wk is kth row vector;
8. Update $w_k^{new} \leftarrow w_k - \partial\tau_k(w_k) / \partial w_k$;
9. Normalize the estimated row vector
   $w_k^{new} = w_k^{new} / \left\| w_k^{new} \right\|$;
10. Evaluate cost function Γk (WNew) using equation (5).
11. If Γr (WNew) < Γr (W)

'then
update weight matrix and cost function
increase size of step t= 1.5 t
else
if N- trace|W, WNew| < 10-7N then
break; [13]
end
decrease step size t= 0.5 t

end of loop;

In above mention technique updation of whitening filter coefficient consumes most of the CPU time but computationally lighter version of this technique is available in literature.

III. WEIGHT INITIALIZED FAST ICA

Fast ICA is one of the very popular ICA techniques to solve blind source separation problem. Fast ICA is based on negentropy which is different from entropy rate bound minimization method discussed in previous section. Key objective of blind source separation techniques is ensuring non-Gaussianity of estimated signals, for that in Fast ICA processes negentropy used as an important tool which is strongly dependent on entropy. Entropy can be defined for any random variable 'V' [13]

$$H(V) = -\sum_i p(V = v_i) \ln p(V = v_i), \qquad (7)$$

Negentropy of variable 'V' can be defined as

J(V) = H(Vgauss) –H(V) (8)

The Fast-ICA procedure focused on negentropy maximization and the unmixing matrix 'W' weight updation is as follows;

$$w^{new} = w - \frac{E\{Vg(w^T V)\} - E\{g'(w^T V)w\}}{E\{g'(w^T V)\} - E\{V^T V g'(w^T V)\}} \qquad (9)$$

Where, g(.) is a derivative of non-quadratic function G(.) .

The weight initialized Fast-ICA incorporated with a modified weight initialization technique, which is as follows;

Step1: initialize a random weight matrix
W← rand (m,n)
Step2: Calculate E[w]
Step3: update the unmixing matrix

WNew←w-E[w]
The weight initialization improves the convergence speed of algorithms.

IV. DIVERGENCE BASED ICA

ICA algorithms are intended to minimize the mutual information between observed variables, to achieve this objective a valid contrast function or divergence measure is necessity. The mutual information content between two continuous variables v1 and v2 can be defined using fundamentals of Shannon's entropy H[.] [14] and relative entropy values can be formulated by using joint probability p(v1, v2) and marginal probability p(v1), p(v2) . By incorporating joint probability and marginal probability function various divergence measure are defined. By introducing joint and marginal probability in Euclidian distance Euclidian Divergence (E-Div) and with Cauchy-Schwartz inequality, Cauchy-Schwartz divergence (CS-Div) is defined [15].The characteristics' DE(v1, v2)>=0 and DCS(v1, v2)>=0 holds true if an only if v1 and v2 are independent . Amari [16] introduced defined a divergence measure of dependence Dα(v1, v2, α) in which by substitution of α= -1 Kullback-Lieber (KL) Divergence is derived [17]. There are many more divergence measure introduced, f-divergence [18], JS-Div based on Jensens' inequality [19].
Chien and Hsieh [3] proposed a divergence based on convex function to derive a sophisticated contrast function for independent component analysis by incorporating joint probability distribution function p(v1, v2) and multiplication of marginal probability distributions p(v1,) and p(v2) into a convex function along with certain combination weight. Performance of this technique is dependent on a convexity parameter α and by varying the α, quality of separation in terms of signal to interference ratio of separated signal and speed of convergence can be controlled. A modified convex divergence and alternate of convex divergence ICA is proposed in [20]. The modified divergence function is as follows;

$$D_{NC}(v_1,v_2,\alpha)=\iint \begin{matrix}\{\lambda f(p(v_1,v_2))+(1-\lambda)f(p(v_1)p(y)) \\ -f(\lambda p(v_1,v_2)+(1-\lambda)p(v_1)p(y))\}dv_1dv_2\end{matrix}$$

$$=\iint\{\frac{\alpha^2}{1-\alpha^4}\left[\frac{1+\alpha}{2}+\frac{1-\alpha}{2}p(v_1,v_2)-(p(v_1,v_2))^{1-\alpha/2}\right]$$
$$+(\frac{\alpha^3}{1-\alpha^4})\left[\frac{1+\alpha}{2}+\frac{1-\alpha}{2}p(v_1)p(v_2)-(p(v_1)p(v_2)^{1-\alpha/2}\right]$$
$$-\frac{\alpha}{1-\alpha^2}\{\left[\frac{1+\alpha}{2}+\frac{1-\alpha}{2}\left(\frac{\alpha}{1+\alpha^2}p(v_1,v_2)+\frac{1-\alpha+\alpha^2}{1+\alpha^2}p(v_1)p(v_2)\right)\right]$$
$$-[\frac{\alpha}{1+\alpha^2}p(v_1,v_2)+\frac{1-\alpha+\alpha^2}{1+\alpha^2}p(v_1)p(v_2)]^{1-\alpha/2}\}\}dv_1dv_2 \quad (10)$$

Equation (10) is derived from substituting a modified combination weight λ=α/1+α2for improving the convergence speed of convex divergence ICA and function from equation (11).

$$f(n)=\frac{\alpha}{1-\alpha^2}\left[\frac{1+\alpha}{2}+\frac{1-\alpha}{2}n-n^{(1-\alpha)/2}\right] \quad (11)$$

Where DNC is modified divergence measure and α is convexity factor. To solve problem of blind source separation modified convex divergence based ICA techniques is a follows.
Algorithms initialized with initialization of some perquisites like learning rate, convexity factors, and required number of iterations. Next step is data pre-processing; which includes centring and whitening of data. Centring is a simple procedure of mean remove from data [V] = [V] - E[V] and whitening is performed using eigenmatrices V ← φ D-1/2φT X ; Where φ and D denotes eigenmatrix and eigenvalue of matrix E[VVT]. Unmixing is estimated simply following the fundamentals of unsupervised learning and this process starts with weight initialization by using the weight initialization techniques discussed in section III. For weight updation modified divergence measure is incorporated in scaled natural gradient based learning [3].

$$W^{(t+1)}=W^{(t)}-\eta\frac{c^{(t)}}{d^{(t)}}\frac{\partial D_{NC}(\chi,W^{(t)},\alpha)}{\partial W^{(t)}}W^{(t)T}W^{(t)} \quad (12)$$

Unlike entropy rate bound minimization in this technique kurtosis value is considered for define the stopping criteria. According to central limit theorem mixture of more than one non-Gaussian signal is Gaussian in nature. Profile of separated signal is always super-Gaussian in nature irrespective to mixing environment [22]. Kurtosis of super-Gaussian signals is positive. In this algorithm a positive kurtosis value considered as stopping criteria.
In continuation of above mentioned procedure the estimated unmixing matrix 'W' is normalized WNEW←W/||W|| , mean shifting Esti_W ← W+ η* W and whitening W ← Esti_W* Whiten Matrix, is followed by normalization. Final separated source signals are estimated by using;

S = V*W (13)

V. PERFORMANCE ANALYSIS AND RESULT DISCUSSION

In order to analyze performance of ICA algorithms a Gaussian ARMA (u,v) process can modeled. An MA process is a special case of an ARMA process when u=0. To evaluate the effectiveness of algorithms for practical applications and performance of algorithms are compared for complex blind source separation task. Quaderature Amplitude Modulated data generated as artificial mixture. In this process fourth order MA model with randomly generated coefficients which are driven by QAM sources with order 2m for the mth source. This method is adapted for the sake of non-Gaussianity, non-circularity and sample-dependence. It is pretty evident that except odd order sources all other sources are second order-circular. This experiment is performed with various sample size, the number of sources is fixed to 12 for equal number of even and odd order sources. Sample size is fixed to 1500.In this experiment performance of source separation algorithms are evaluated in terms of normalized value of interference to signal ratio, failure percentage and average CPU time.

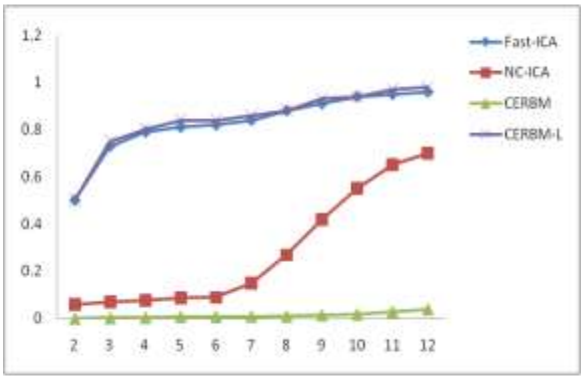

Fig 1 Normalized Interference to signal ratio as a function of number of sources.

Performance of algorithms is in form of normalized interference to signal ration shown in figure 1. Results are evident that complex entropy rate bound minimization technique incorporate all three kind of diversity and performs better than other competitors. One more observation can be made that performance of algorithms degraded with increment in number of sources.

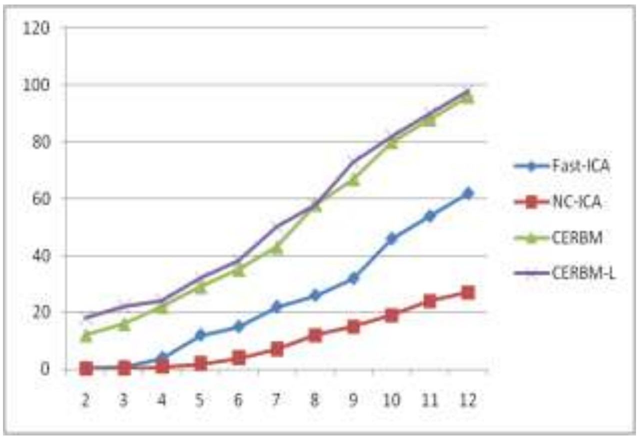

Fig 2 Average CPU time consumed as a function of number of sources.

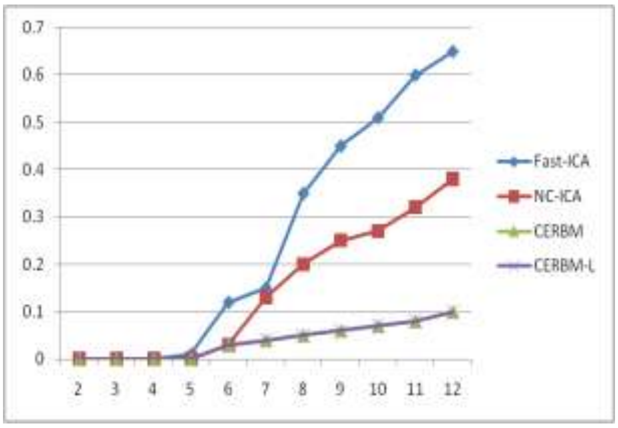

Fig 3 Failure percentage function of number of sources.

Average time consumption by CPU is higher for CERBM and CERBM-L algorithms because it needs more time for calculation of whitening filter coefficients but somewhere CERBM is faster than CERBM-L. Average CPU time increases with increment in number of sources and NC-ICA is faster than other competitor, which is key objective of modified convex divergence algorithm. Percentage failure of CERBM is very low but performance gets affected due to increment in number of sources.

VI. CONCLUSION

In this paper a comparative analysis of ICA algorithms is done for interference to signal ratio, average CPU time and percentage of failure. A general form of Fisher information matrix and Cramer-Rao lower bound is established for inclusion of three type of diversity non-Gaussianity, non-circularity and sample dependence. Results are evident that Entropy rate bound minimization exploits all three diversities. CERBM exhibits better performance in blind source separation task but consumes much CPU time because of time consumption in coefficient estimation of whitening filter and NC-ICA consumes minimum CPU time. Percentage of failure rises with increment in number of sources

REFERENCES:

[1] Vincent Emmanuel, Bertin Nancy, G.Remi, Bimbot Frederic, "From Blind to Guided Audio Source Separation", IEEE Signal Proc. Magazine, pp. 107-115, May 2014

[2] J. F. Cardoso and A Souloumiac, "Blind beamforming for non-Gaussian signals", Proc. Inst. Electri. Eng. F-Radar Signal processing, vol. 140, No. 6, pp. 362-370, Nov 1998.

[3] Jen-Tzung Chien, Hsin-Lung Hsieh, "Convex Divergence ICA for Blind Source Separation", IEEE Transc. On Audio, Speech and Language Prosc., Vol. 20, No. 01, pp. 302-312.

[4] E. Bingham and A. Hyvarinen, "A fast fixed point algorithm for independent component analysis of complex valued signals" Int, J. Neural syst. Vol. 10, No. 01, pp. 1-8, Feb 2000.

[5] Ritesh K Yadav, R. Mehra, N. Dubey, "Blind Audio Source Separation using Weight Initialized Independent Component Analysis", in proc. Of IEEE conference on Next Generation Computing, UPES, Dehradun, India, Sept. 2015.

[6] T. Adali, T.Kim, and V. Calhoun, "Independent component analysis by complex nonlinearities", In proc. Of Int. conf. Acoust. Signal processing, 2004, Vol. 5, pp. 525-528.

[7] H. Li and T. Adali, "A class of complex ICA algorithm based on the kurtosis cost function", IEEE Transc. Neural Netw. Vol. 19, No. 3, pp. 408-420, Mar. 2008

[8] X.-L. Li, T. Adali, "Complex Independent Component Analysis vy entropy bound minimization" IEEE Transc. Circuit Syst. , Vol. 57, No. 7, pp. 1417-1430, Jul 2010.

[9] B. Loesch and B. Yang, "Cramer-Rao Bound for circular complex independent component analysis", IEEE Transc. Signal Proc., Vol. 61, No. 2, pp. 365-379, Jan 2013.

[10] X-L Li, T. Adali, "Blind separation of noncircular correlated source using Gaussian entropy rate," IEEE Transc. Signal Proc., Vol. 59, No. 6, pp. 2969-2975, Jun 2011.

[11] X-L Li and X D Zang, "Non orthogonal joint diagonalization free of degenerate solution", IEEE Transc. Signal Processing, Vol. 55, No. 10, pp. 1803-1814, May 2007.

[12] M. Anderson, X-L Li, P Rodriguez, T. Adali, " An effective coupling method for matrix optimization and its application to ICA problem," in porc. Of IEEE int. conf. Acous. Sig, Processing. March. 2012.

[13] T. M. Cover and J. A Thomas, "Elements of information theory", publisher John Wiley and sons, Hoboken pp; 640, ISBN 0471748811, 1991.

[14] C. E. Shannon, "A mathematical theory of communication," Bell Syst.Tech. J., vol. 27, pp. 379–423, 1948.

[15] D. Xu, J. C. Principe, J. Fisher, III, and H.-C. Wu, "A novel measurefor independent component analysis (ICA)," in Proc. Int. Conf. Acoust.,Speech, Signal Process., 1998, vol. 2, pp. 1161–1164.

[16] S. Amari, Differential-Geometrical Methods in Statistics. New York: Springer-Verlag, 1985.

[17] J. Zhang, "Divergence function, duality, and convex analysis," Neural Comput., vol. 16, pp. 159–195, 2004.

[18] Csiszár and P. C. Shields, "Information theory and statistics: A tutorial," Foundations Trends Commun. Inf. Theory, vol. 1, no. 4, pp.417–528, 2004.

[19] J. Lin, "Divergence measures based on the Shannon entropy," IEEETrans. Inf. Theory, vol. 37, no. 1, pp. 145–151, 1991.

[20] Naveen Dubey, Rajesh Mehra, "Unsupervised Learning Based Modified C-ICA for Audio Source Separation in Blind Scenario", Accepted in Int. J. Information Technology and Computer Sc.

[21] N. Dubey, R. Mehra, "Blind Audio Source Separation (BASS): An Unsupervised Approach", Edit-2015, Int. J. of Electrical and Electronics Eng. Vol.2 Spl issue. 01 , pp. 29-33, 2015

[22] Naveen Dubey, Rajesh Mehra, "Blind Audio Source Separation in Time Domain Using ICA Decomposition", IJCA, Vol 132, No. 6, pp.48-56, 2015

[23] Shaminder Kaur , Rajesh Mehra, "FPGA Implimentation of OFDM Transceiver using FFT Algorithms", IJEST, Vol.04, No. 04, pp. 1532-1537, April 2012

Naveen Dubey is currently associated with Electronics and Communication Engineering Department of RKG Institute of Technology for Women, Ghaziabad, India since 2008. He is ME-Scholar at National Institute of Technical Teachers' Training & Research, Chandigarh, India and received Bachelor of Technology from UP Technical University, Lucknow, India in 2008. Received best performer award in creativity and innovation from IIT-Delhi. He guided 20 UG projects and presented 2 projects in Department of Science and Technology, Government of India. He published and presented more than 10 papers in International conferences and journals. His research areas are Digital Signal Processing, Neural Networks and EM Fields. His research project includes Blind audio source separation.

Dr. Rajesh Mehra: Dr. Mehra is currently associated with Electronics and Communication Engineering Department of National Institute of Technical Teachers' Training & Research, Chandigarh, India since 1996. He has received his Doctor of Philosophy in Engineering and Technology from Panjab University, Chandigarh, India in 2015. Dr. Mehra received his Master of Engineering from Panjab Univeristy, Chandigarh, India in 2008 and Bachelor of Technology from NIT, Jalandhar, India in 1994. Dr. Mehra has 20 years of academic and industry experience. He has more than 250 papers in his credit which are published in refereed International Journals and Conferences. Dr. Mehra has 55 ME thesis in his credit. He has also authored one book on PLC & SCADA. His research areas are Advanced Digital Signal Processing, VLSI Design, FPGA System Design, Embedded System Design, and Wireless & Mobile Communication. Dr. Mehra is member of IEEE and ISTE.