Principal Independent Component Analysis with Multi-Reference

Jie Luo, Bo Hu, Xie-Ting Ling

Electronic Engineering Department
Fudan University
Shanghai 200433, P. R. CHINA
Email: xtling@fudan.edu.cn

Rui-Wen Liu

Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556, USA
Email: ruey-wen.liu.1@nd.edu

ABSTRACT

Conventional Blind Signal Separation algorithms do not adopt any asymmetric information of the input sources, thus the convergence point of a single output is always unpredictable. In this paper, a new Principal Independent Component Analysis concept is proposed, we try to extract the objective Independent Component directly without separating all the signals. A cumulant-based globally convergent algorithm is presented and the simulation results show a hopeful prospect of the Principal Independent Component Analysis in applications.

1. INTRODUCTION

During the past several years, Independent Component Analysis (ICA)[1][2] has begun to find a wide applicability in many diverse fields. Among them are signal detection, channel equalization and feature extraction[3]. Blind Signal Separation (BSS), which can be regarded as one of the classical applications of the ICA model, focuses on extracting all the Independent Components (ICs) from their linear combinations. Usually, BSS methods assume the IC sources and the mixing matrix are totally blind to the ICA network. Without introducing any prior information, the exact convergence point of a single output is theoretically undeterminable[4][5]. However, in some applications such as signal detection and noise cancellation, we may only be interested in one or two of the ICs. And picking out the desired signal from the separation results shows we do have some prior information of the IC sources, they are in fact not totally blind to us. In this paper, we try to develop a Principal Independent Component Analysis (PICA) approach that can extract such kind of useful information and get the objective signal without separating all the ICs.

2. PROBLEM DESCRIPTION AND PRINCIPAL INDEPENDENT COMPONENT ANALYSIS

The PICA network structure can be described by figure 1.

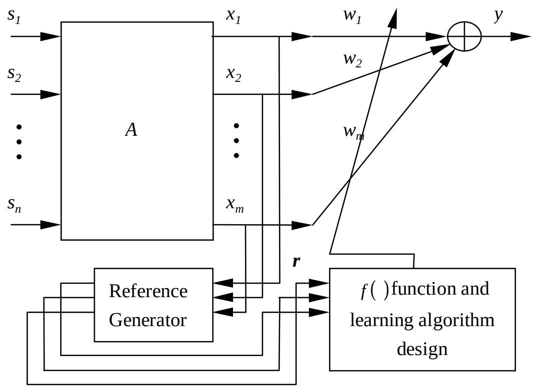


Figure 1. PICA network structure

$\boldsymbol{s} = [s_1, s_2, \cdots, s_n]^T$ is the source vector, the n complex-valued stationary non-gaussian ICs are assumed to be statistically independent from each other. A is a $n \times m$ complex-valued mixing matrix of full column rank. $\boldsymbol{x} = [x_1, x_2, \cdots, x_m]^T$ is the sampling vector, $\boldsymbol{w} = [w_1, w_2, \cdots, w_m]^T$ is the weight vector of the neural network and y is the output. As we have mentioned, without any prior information, conventional BSS methods can not give a globally convergent algorithm and the convergence point of the output is unpredictable. Here we will continue assuming that the exact value of the IC sources and the mixing matrix are blind to us. However, suppose we can get a set of reference signals $r_1 \quad r_2 \quad \cdots \quad r_v$, which can also be expressed as linear combinations of the ICs.

$$r_i = \boldsymbol{m}_i^T s = \sum_{l=1} m_{il} s_l \qquad (i = 1,2,\cdots,v) \qquad (1)$$

Notice in figure 1, the references are not generated directly from IC sources, hence the original "blind" assumption remains true to some extend. Now we would like to give some definitions to the useful functions that will appear in this paper.

Suppose the network output can be denoted by

$$y = \boldsymbol{w}^T \boldsymbol{x} = \boldsymbol{w}^T A\boldsymbol{s} = \boldsymbol{g}^T \boldsymbol{s} = \sum_l g_l s_l \qquad \left(\boldsymbol{g} = A^T \boldsymbol{w}\right) \tag{2}$$

The second order cumulant and fourth order cumulant of y are defined respectively by[6]

$$Cum(y:2) = E\{y^* y\} \tag{3}$$

$$Cum(y:4) = E\{y^* y y^* y\} - 2E^2\{y^* y\} - \left|E\{y^2\}\right|^2 \tag{4}$$

The fourth order cross cumulant between y, r_i is defined by

$$Cum(y, r_i:4) = E\{y^* y r_i^* r_i\} - E^2\{y^* y\}E^2\{r_i^* r_i\} - \left|E\{y r_i^*\}\right|^2 - \left|E\{y r_i\}\right|^2 \tag{5}$$

According to [6], we have

$$Cum(y:2) = \sum_l |g_l|^2 Cum(s_l:2) \tag{6}$$

$$Cum(y:4) = \sum_l |g_l|^4 Cum(s_l:4) \tag{7}$$

$$Cum(y, r_i:4) = \sum_l |g_l|^2 |m_{il}|^2 Cum(s_l:4) \tag{8}$$

Then, we define the "cross-non-gaussianity" between y and r_i as[7]

$$Ng(y, r_i) = \frac{Cum(y, r_i:4)}{Cum(y:2)Cum(r_i:2)} = \frac{\sum_l |g_l|^2 |m_{il}|^2 Cum(s_l:4)}{\left[\sum_l |g_l|^2 Cum(s_l:2)\right]\left[\sum_l |m_{il}|^2 Cum(s_l:2)\right]} \tag{9}$$

and the non-gaussian energy of s_i in r_j is defined by

$$Ng\_E(s_i \mid r_j) = \frac{|m_{ji}|^2 Cum(s_i:4)}{Cum(s_i:2)} \tag{10}$$

Now if we define a multi-variable linear function

$$f(x_1, x_2, \cdots x_v) = \sum_{i=1}^{v} a_i x_i \tag{11}$$

and design the object function of the network as

$$E(y) = \varphi \frac{f\left(Cum(y, r_1:4), Cum(y, r_2:4), \cdots, Cum(y, r_v:4)\right)}{Cum(y:2)} - [Cum(y:2) - 1]^2 \tag{12}$$

Here, φ is a sign variable, $\varphi = 1$ or $\varphi = -1$. We will see the following conclusion will hold

Remark 1 Without losing the generality, if the IC sources can be arranged and satisfy

$$\begin{aligned}
&\varphi f\left(Ng_E(s_1 \mid r_1), Ng_E(s_1 \mid r_2), \cdots, Ng_E(s_1 \mid r_v)\right) \\
&> \varphi f\left(Ng_E(s_2 \mid r_1), Ng_E(s_2 \mid r_2), \cdots, Ng_E(s_2 \mid r_v)\right) \\
&> \cdots \\
&> \varphi f\left(Ng_E(s_n \mid r_1), Ng_E(s_n \mid r_2), \cdots, Ng_E(s_n \mid r_v)\right)
\end{aligned} \tag{13}$$

then maximizing object function (12), the output of the network can finally be denoted by

$$y = g_1 s_1 \tag{14}$$

and $Cum(y:2) = 1$ will be satisfied.

Proof We can prove that on the final convergence point, g_1 will be the only none-zero component in vector $\boldsymbol{g}$. In fact, if there exists a $i \neq 1$, and $g_i \neq 0$, do a perturbation with $\sigma > 0$, let

$$|\tilde{g}_i|^2 = |g_i|^2 - \frac{\sigma}{Cum(s_i:2)} \tag{15}$$

$$|\tilde{g}_1|^2 = |g_1|^2 + \frac{\sigma}{Cum(s_1:2)} \tag{16}$$

we obtain

$$E(\tilde{y}) - E(y) = \frac{\varphi\sigma}{\left(\sum_l |g_l|^2 Cum(s_l:2)\right)} \sum_{l=1}^{v} \left[a_l Ng\_E(s_1 \mid r_l) - a_l Ng\_E(s_i \mid r_l)\right] > 0 \tag{17}$$

and the result shows (14) will be satisfied.

Meanwhile, if we have $y = g_1 s_1$ and $Cum(y:2) = |g_1|^2 Cum(s_1:2) \neq 1$, do perturbation with $2 > \sigma > 0$, let

$$|\tilde{g}_1|^2 = |g_1|^2 \left(1 - \sigma \frac{Cum(y:2) - 1}{Cum(y:2)}\right) \tag{18}$$

We get

$$E(\tilde{y}) - E(y) = \left(2\sigma - \sigma^2\right)\left(Cum(y:2) - 1\right)^2 > 0 \tag{19}$$

thus $Cum(y:2) \neq 1$ can not be a maxima of the network.

Proof completed.

The discrete gradient-based learning algorithm can be expressed by

$$\begin{aligned}
\boldsymbol{w}_{n+1} = \boldsymbol{w}_n &- \mu\varphi \sum_{l=1}^{v} \frac{a_l\left[\left\langle |y|^2 |r_l|^2 \right\rangle_n - \left|\left\langle y r_l \right\rangle_n\right|^2\right] y_n}{\left(\left\langle |y|^2 \right\rangle_n\right)^2} x_n^* \\
&+ \mu\varphi \sum_{l=1}^{v} \frac{a_l\left[\left|\left\langle y r_l^* \right\rangle_n\right|^2 + \left\langle |y|^2 \right\rangle_n |r_{l\,n}|^2\right] y_n}{\left(\left\langle |y|^2 \right\rangle_n\right)^2} x_n^* \\
&- \mu\varphi \sum_{l=1}^{v} \frac{a_l\left[\left\langle |y|^2 \right\rangle_n \left\langle y r_l \right\rangle_n r_{l\,n}^* + \left\langle |y|^2 \right\rangle_n \left\langle y r_l^* \right\rangle_n r_{l\,n}\right]}{\left(\left\langle |y|^2 \right\rangle_n\right)^2} x_n^* \\
&- 2\mu\left(\left\langle |y|^2 \right\rangle_n - 1\right) y_n \boldsymbol{x}_n^*
\end{aligned} \tag{20}$$

$$\left\langle |y|^2 |r_l|^2 \right\rangle_{n+1} = (1 - \delta_1)\left\langle |y|^2 |r_l|^2 \right\rangle_n + \delta_1 |y_n|^2 |r_{l\,n}|^2 \tag{21}$$

$$\left\langle |y|^2 \right\rangle_{n+1} = (1 - \delta_2)\left\langle |y|^2 \right\rangle_n + \delta_2 |y_n|^2 \tag{22}$$

$$\left\langle y r_l \right\rangle_{n+1} = (1 - \delta_3)\left\langle y r_l \right\rangle_n + \delta_3 y_n r_{l\,n} \tag{23}$$

$$\left\langle y r_l^* \right\rangle_{n+1} = (1 - \delta_4)\left\langle y r_l^* \right\rangle_n + \delta_4 y_n r_{l\,n}^* \tag{24}$$

Here μ is the learning step, and δ_1 δ_2 δ_3 δ_4 are the steps used for the online estimation of the high order moments respectively.

3. SIMULATION RESULTS

A base-band CDMA emulation system is shown in figure 2. The 3 IC source s_1 s_2 s_3 are assumed to be sub-gaussian 4×4 3×3 2×2 QAM signals respectively. The received signal *rec* is denoted by

$$rec = a_1 s_1 + a_2 s_2 + a_3 s_3 \tag{25}$$

And suppose after the demodulation for each user respectively, the final sampling signals yield

$$\begin{cases} x_1 = a_1 s_1 + \dfrac{a_2}{\lambda} s_2 + \dfrac{a_3}{\lambda} s_3 + n_1 \\ x_2 = \dfrac{a_1}{\lambda} s_1 + a_2 s_2 + \dfrac{a_3}{\lambda} s_3 + n_2 \\ x_3 = \dfrac{a_1}{\lambda} s_1 + \dfrac{a_2}{\lambda} s_2 + a_3 s_3 + n_3 \end{cases} \tag{26}$$

Here we use n_1 n_2 n_3 to simulate the additive noise, and use $\lambda > 1$ to simulate the attenuation of demodulation. Set $\varphi = -1$, and the reference signals are chosen as $r_1 = x_1$, $r_2 = x_2$ and $r_3 = x_3$.

In such an emulation system, we find out that the following equations can be satisfied.

$$\begin{cases} \varphi Ng\_E(s_1 \mid r_1) > \dfrac{\varphi}{3}\sum_{l=1}^{3} Ng\_E(s_1 \mid r_l) \\ \varphi Ng\_E(s_2 \mid r_1) < \dfrac{\varphi}{3}\sum_{l=1}^{3} Ng\_E(s_1 \mid r_l) \\ \varphi Ng\_E(s_3 \mid r_1) < \dfrac{\varphi}{3}\sum_{l=1}^{3} Ng\_E(s_1 \mid r_l) \end{cases} \tag{27}$$

Then if we set the object function as

$$E(y_1) = \varphi \frac{Cum(y_1, r_1 : 4) - \frac{1}{3}\sum_{l=1}^{3} Cum(y_1, r_l : 4)}{Cum(y_1 : 2)} - [Cum(y_1 : 2) - 1]^2 \tag{28}$$

According to remark 1, we will get $y_1 \to s_1$ by maximizing (28). Similarly, choosing the object functions of y_2, y_3 as

$$E(y_2) = \varphi \frac{Cum(y_2, r_2 : 4) - \frac{1}{3}\sum_{l=1}^{3} Cum(y_2, r_l : 4)}{Cum(y_2 : 2)} - [Cum(y_2 : 2) - 1]^2 \tag{29}$$

$$E(y_3) = \varphi \frac{Cum(y_3, r_3 : 4) - \frac{1}{3}\sum_{l=1}^{3} Cum(y_3, r_l : 4)}{Cum(y_3 : 2)} - [Cum(y_3 : 2) - 1]^2 \tag{30}$$

Maximizing these object functions, we can also get $y_2 \to s_2$, $y_3 \to s_3$ respectively. In our computer simulation, the SNR is set to 26db. Other variables are chosen as $a_1 = 6$, $a_2 = 1$, $a_3 = 1$ and $\lambda = 5$. In order to improve the convergence, we add a pre-whitening network after the sampling process. Similar to the pre-whitening network in [8], the outputs after pre-whitening will satisfy $p_i^* p_i = 1$ $(i = 1,2,3)$ and $p_i^* p_j = 0$ $(i, j = 1,2,3; i \neq j)$. Meanwhile, we define a set of normalized correlation functions

$$\theta_{ij} = \sqrt{\frac{\left|E\{y_i^* s_j\}\right|^2}{E\{|y_i|^2\}E\{|s_j|^2\}}} \quad (i, j = 1,2,3) \tag{31}$$

Obviously, if $y_i \to s_j$ can be satisfied, the correlation functions will yield $\theta_{ij} \to 1$ and $\theta_{ik} \to 0$ $(k \neq j)$. Thus the convergence of the network can partly be expressed by the convergence of these variables. Figure 3 (a) (c) (e) give the final results of the three outputs after 2100 iterations, and figure 3 (b) (d) (f) show the convergence process of the outputs respectively.

We should mention that, the only requirement here is $a_1 a_2 a_3 \neq 0$. Since $x_2 = 1.2 s_1 + s_2 + 0.2 s_3$, we can see the multi-user interference s_1 is even stronger than the user signal s_2. And though face such a hard situation, the PICA network can extract the object signal efficiently.

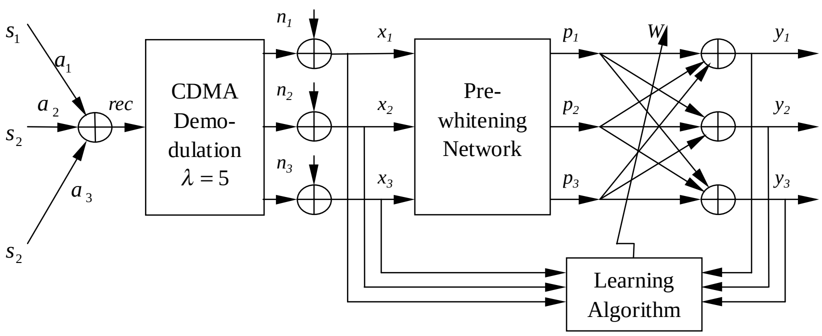
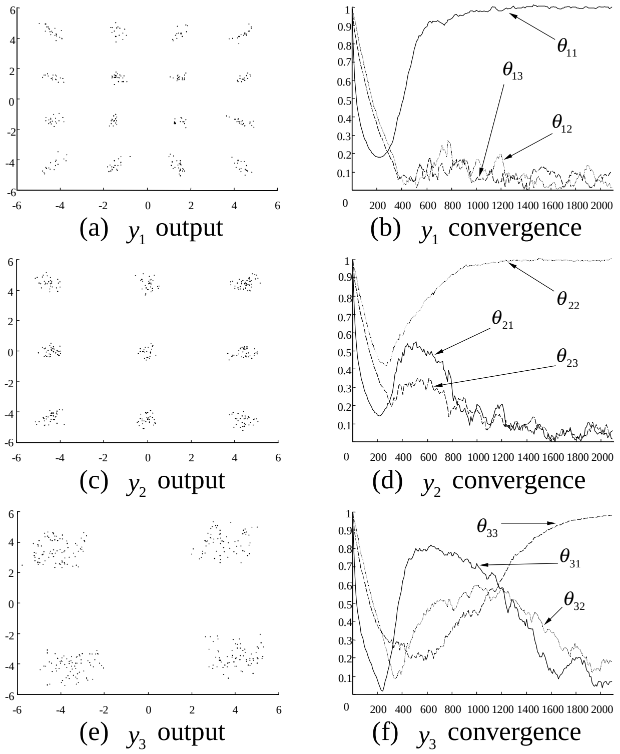


Figure 2. Base-band CDMA emulation system



(a) y_1 output (b) y_1 convergence

(c) y_2 output (d) y_2 convergence

(e) y_3 output (f) y_3 convergence

Figure 3. Simulation result of CDMA system (SNR=26db)

4. CONCLUSION

A globally convergent multi-reference PICA algorithm is proposed. Unlike conventional BSS methods, PICA network focuses its scope on extracting prior information and tracing the object signal directly. Compare with the multi-output BSS algorithms, the single-output PICA network is much simpler in computation complexity. Simulation result is given to show some outstanding feature of the PCIA idea in applications. Further research work is needed to improve the performance of the network.

REFERENCES

[1] C. Jutten, J. Herault, "Blind Separation of Sources, Part I: An Adaptive Algorithm Based on Neutomimetic Architecture", *Signal Processing*, vol. 24, pp. 1-10, 1991.

[2] P. Comon, "Independent Component Analysis - A New Concept?", *Signal Processing*, vol. 36, pp. 287-314, 1994.

[3] Aapo Hyvarinen, Erkki Oja, "Independent Component Analysis by General Non-linear Hebbian-like Learning Rules", *Signal Processing*, vol. 64, no. 3, 1998.

[4] R.W. Liu, "Blind signal separation: I-fundamental concepts", *Journal of Circuits and Systems*, vol. 1, No. 1, pp. 1-5, 1996.

[5] R.W. Liu, "Blind signal separation: II-blind signal separation", *Journal of Circuits and Systems*, vol. 1, No. 2, pp. 1-5, 1996.

[6] Jerry M. Mendel, "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications", *Proceeding of the IEEE*, vol. 79, No. 3, pp. 278-305, March, 1991.

[7] Jie Luo, Bo Hu, Xie-Ting Ling, "New Cumulant Based Algorithms for Independent Component Analysis and Principal Independent Component Analysis", submited to *IEEE Trans. on Neural Networks*. Can be found at http://www.bitsmart.com/rockey.

[8] Juha Karhunen, Liuyue Wang, Ricardo Vigario, "Nonlinear PCA Type Approaches for Source Separation and Independent Component Analysis", *IEEE ICNN'95*, Perth, Australia, pp. 995-1000, November, 1995.