Principal Independent Component Analysis

Jie Luo, Bo Hu, Xie-Ting Ling, and Ruey-Wen Liu

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. However, in most of the applications, we are usually interested in only one or two of the source signals and prior information is almost always available. In this paper, a principal independent component analysis (PICA) 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 simulation results are given to show the hopeful applicability of the PICA ideas.**

Index Terms—**Cumulants, globally convergent, high-order statistics, non-Gaussian energy, principal independent component analysis.**

I. Introduction

During the past several years, independent component analysis (ICA) [1]–[3] has begun to find a wide applicability in many diverse fields. Among them are signal detection, channel equalization, and feature extraction. Blind signal separation (BSS) [4], [6], [8], which can be regarded as one of the classical applications of the ICA model, focuses on extracting all the independent components (IC's) from their linear combinations. Many BSS algorithms are already well known. Among them are the H-J algorithm [6], [7], modified H-J algorithm [8], [9], the nonlinear PCA network [2], [3], and other cumulant-based approaches [4], [5]. BSS methods are called "blind" since they usually assume that the IC sources and the mixing matrix are totally unavailable to the ICA network [10]. Without introducing any prior information, the exact convergence point of a single output is theoretically unpredictable. However, in some applications such as signal detection and noise cancellation, we may not be interested in all the IC's simultaneously. Examining the signal processing process in applications, sometime we may come to the following questions. What will we do next to the BSS process? If we are not interested in all source signals, of course we would like to pick the desired signal out from the separation results. However, if absolutely no asymmetric information is available, how can we know which signal is the one we are looking for? Or, if we really can identify the source signals, why we do not use this prior information in the signal separation process to simplify the network? In fact, this is the key idea of the principal independent component analysis (PICA) methods [11]. By introducing some asymmetric information to the network, we now try to extract the objective signal directly without separating all the IC's. Especially in the simulation part of this paper, we will see that in most of the cases, limited prior information can do great help to simplify the network complexity.

Manuscript received April 30, 1998; revised November 5, 1998 and March 22, 1999.

J. Luo, B. Hu, and X.-T. Ling are with the Electronic Engineering Department, Fudan University, Shanghai 200433, China.

R.-W. Liu is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA.

Publisher Item Identifier S 1045-9227(99)05997-4.

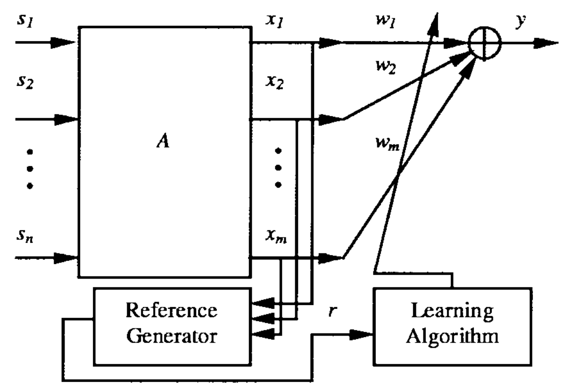

Fig. 1. PICA network with single reference.

This paper is organized as follows. In Section II, a basic model of PICA network is proposed. Thorough discussion to the convergence is given. And in Section III, we extend the PICA model to one with multireference. It can be seen that such a kind of extension makes the PICA methods flexible in applications. Especially from the simulation results given in Section IV, the feasible value of the PICA methods will become more and more clear.

II. Problem Description and the Basic PICA Structure

The basic PICA network can be described by Fig. 1.

Suppose we have n complex-valued non-Gaussian independently identically distributed (i.i.d.) source signals which can be denoted by $\boldsymbol{s} = [s_1, s_2, \cdots, s_n]^T$ in the vector form. A is a $m \times n (m \geq n)$ complex-valued mixing matrix of full comlumn rank. $\boldsymbol{x} = [x_1, x_2, \cdots, x_m]^T$ is the observed signal vector obtained from the receivers. $\boldsymbol{w} = [w_1, w_2, \cdots, w_m]^T$ is the weight vector of the neural network and y is the output. The relation between the vectors and the output can be described by

$$y = \boldsymbol{w}^T\boldsymbol{x} = \boldsymbol{w}^T A\boldsymbol{s} = \boldsymbol{g}^T\boldsymbol{s} = \sum_i g_i s_i \qquad (\boldsymbol{g} = A^T\boldsymbol{w}). \quad (1)$$

As we have mentioned in the introduction part, without any prior information, the convergence point of the output is theoretically 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 reference signal r, which can also be expressed as linear combinations of the IC's

$$r = \boldsymbol{m}^T\boldsymbol{s} = \sum_i m_i s_i. \quad (2)$$

Some ideas about the reference generator will be shown in the simulation part. Nevertheless, since the reference generator will vary greatly in different applications, we will not go into detail about it now, we just assume arbitrarily that a reference signal is available.

The second-order cumulant and fourth-order cumulant of y are defined, respectively, by

$$Cum(y:2) = E\{y^* y\} \tag{3}$$
$$Cum(y:4) = E\{y^* y y^* y\} - 2E^2\{y^* y\} - |E\{y^2\}|^2 \tag{4}$$

where $E\{\}$ denote the expectation and y^* is the conjugate transposition of y. The fourth-order cross cumulant between y, r is defined by

$$Cum(y, r:4) = E\{y^* y^* r\} - 2E^2\{y^* r\} - |E\{yr\}|^2. \tag{5}$$

According to [13], if the sources are i.i.d. signals, we have

$$Cum(y:2) = \sum_l |g_l|^2 Cum(s_l:2) \tag{6}$$
$$Cum(y:4) = \sum_l |g_l|^4 Cum(s_l:4) \tag{7}$$
$$Cum(y, r:4) = \sum_l |g_l|^2 |m_l|^2 Cum(s_l:4). \tag{8}$$

Then we define the "cross-non-Gaussianity" between y and r as

$$Ng(y,r) = \frac{Cum(y,r:4)}{Cum(y:2)Cum(r:2)} = \frac{\sum_l |g_l|^2 |m_l|^2 Cum(s_l:4)}{\sum_l |g_l|^2 Cum(s_l:2) \sum_l |m_l|^2 Cum(s_l:2)}. \tag{9}$$

Obviously, for any arbitrary variable $a \neq 0$ and $\beta \neq 0$, we will have

$$Ng(y,r) = Ng(\alpha y, \beta r). \tag{10}$$

The "non-Gaussian energy" of s_i in r is defined by

$$Ng\_E(s_i|r) = \frac{|m_i|^2 Cum(s_i:4)}{Cum(s_i:2)}. \tag{11}$$

Unlike conventional concept on "energy," we should mention that, for super-Gaussian source s_i (which satisfies $Cum(s_i:4) > 0$), $Ng(s_i|r)$ will always be nonnegtive, for Gaussian source s_i (which satisfies $Cum(s_i:4) = 0$), $Ng(s_i|r)$ will always be zero, while for sub-Gaussian source s_i (which satisfies $Cum(s_i:4) < 0$), $Ng(s_i|r)$ will always take an nonpositive value.

If for any arbitrary $i \neq j$, we have

$$Ng\_E(s_i|r) \neq Ng\_E(s_j|r). \tag{12}$$

Then the source signals can be arranged by their non-Gaussian energy in r. We still assume there is no Gaussian IC. Without loss of generality, suppose we have

$$\begin{aligned} Ng\_E(s_1|r) > \cdots > Ng\_E(s_p|r) > 0 > Ng\_E(s_{p+1}|r) \\ > \cdots > Ng\_E(s_n|r). \end{aligned} \tag{13}$$

According to their non-Gaussian energy value and their Gaussian type, we define $s1$ to be the "principal super-Gaussian IC in r," define s_p to be the "minor super-Gaussian IC in r." Similarly, we call s_{p+1} to be the "minor sub-Gaussian IC in r" and sn to be the "principal sub-Gaussian IC in r."

Then, given the cost function of the neural network as

$$E(y) = Ng(y,r) - (Cum(y:2) - 1)^2. \tag{14}$$

Proposition 1: Given (13) with respect to IC's and the reference signal r, by maximizing the cost function (14), the output of the network can finally be denoted by

$$y = g_1 s_1 \tag{15}$$

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

Proof: Of course, from the Proposition 1 we can see there will be one and only one point of the cost function that can satisfy all the requirements. In fact, none of the other points can be maxima of the cost function.

First, if there exist a $j \neq 1$, which makes $g_j \neq 0$, we do a perturbation with $\sigma > 0$, let

$$|\tilde{g}_j|^2 = |g_j|^2 - \frac{\sigma}{Cum(s_j:2)} \tag{16}$$
$$|\tilde{g}_1|^2 = |g_1|^2 + \frac{\sigma}{Cum(s_1:2).} \tag{17}$$

We get

$$E(\tilde{y}) - E(y) = \frac{\sigma\left(\frac{|m_1|^2 Cum(s_1:4)}{Cum(s_1:2)} - \frac{|m_j|^2 Cum(s_j:4)}{Cum(s_j:2)}\right)}{\left(\sum_l |g_l|^2 Cum(s_l:2)\right)\left(\sum_l |m_l|^2 Cum(s_l:2)\right)} > 0 \tag{18}$$

which means only g_1 can be nonzero.

Second, for any point with $Cum(y:2) \neq 1$, do a perturbation with $2 > \sigma > 0$, let

$$|\tilde{g}_i|^2 = |g_i|^2 \left(1 - \sigma \frac{Cum(y:2) - 1}{Cum(y:2)}\right) \qquad (i = 1, 2, \cdots, n) \tag{19}$$

we get

$$E(\tilde{y}) - E(y) = (2\sigma - \sigma^2)(Cum(y:2) - 1)^2 > 0. \tag{20}$$

Thus we can see proposition 1 will hold.

III. Extended PICA Network with Multireference

In part II, in order to provide some asymmetric information, we assumed arbitrarily that a reference signal is available. However, to most of the cases, it is not so easy to obtain the asymmetric information in such a simple form. In this part, we will extend the PICA network to a more flexible form. The multireference PICA network can be described by Fig. 2.

Here we assume v reference signals are available. All the v references can be expressed by the linear combinations of the IC's

$$r = \boldsymbol{m}_j^T \boldsymbol{s} = \sum_i m_{ji} s_i \qquad (j = 1, 2, \cdots, v). \tag{21}$$

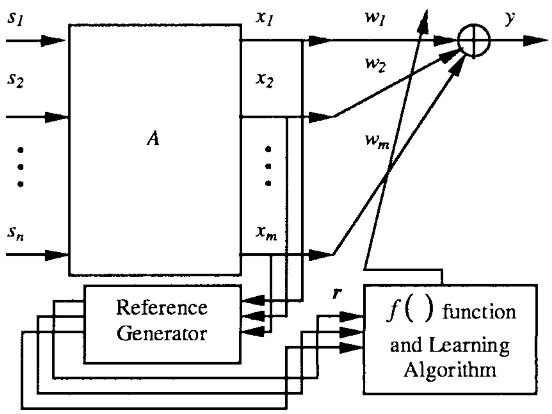

Fig. 2. PICA network with multireference.

Moreover, we define a multivariable linear function

$$f(x_1, x_2, \cdots, x_r) = \sum_{i=1}^{v} a_i x_i \tag{22}$$

with respect to variables $x_1, x_2, \cdots, x_r$. And the object function of the network is designed as shown in (23) at the bottom of the page. If we suppose for any $i \neq j$

$$\begin{aligned} &f(Ng\_E(s_i|r_1), Ng\_E(s_i|r_2), \cdots, Ng\_E(s_i|r_v)) \\ &\quad \neq f(Ng\_E(s_j|r_1), Ng\_E(s_j|r_2), \cdots, Ng\_E(s_j|r_v)). \end{aligned} \tag{24}$$

then we have

Proposition 2: Given (24), if the IC sources can be arranged by

$$\begin{aligned} &f(Ng\_E(s_1|r_1), Ng\_E(s_1|r_2), \cdots Ng\_E(s_1|r_v)) \\ &\quad > f(Ng\_E(s_2|r_1), Ng\_E(s_2|r_2), \cdots, Ng\_E(s_2|r_v)) \\ &\quad > \cdots \\ &\quad > f(Ng\_E(s_n|r_1), Ng\_E(s_n|r_2), \cdots, Ng\_E(s_n|r_v)) \end{aligned} \tag{25}$$

maximizing the cost function (23), the output of the network can finally be denoted by

$$y = g_1 s_1 \tag{26}$$

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

Proof: In fact, similar to that of proposition 1, the proof of this proposition is quite simple, too.

For any $j \neq 1$, if $g_j \neq 0$, do a perturbation with $\sigma > 0$ let

$$|\tilde{g}_j|^2 = |g_j|^2 - \frac{\sigma}{Cum(s_j:2)} \tag{27}$$

$$|\tilde{g}_1|^2 = |g_1|^2 - \frac{\sigma}{Cum(s_1:2)}. \tag{28}$$

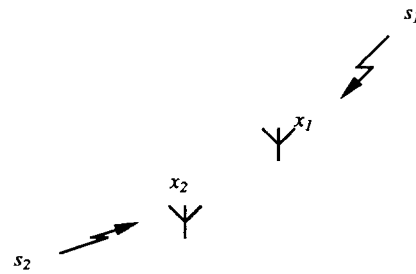

Fig. 3. The geographical asymmetric information.

We get

$$\begin{aligned} &E(\tilde{y}) - E(y) \\ &\quad = \frac{\sigma f(Ng\_E(s_i|r_1), Ng\_E(s_i|r_2), \cdots, Ng\_E(s_i|r_1))}{\left(\sum_l |g_l|^2 Cum(s_l:2)\right)} \\ &\qquad - \frac{\sigma f(Ng\_E(s_j|r_1), Ng\_E(s_j|r_2), \cdots, Ng\_E(s_j|r_1))}{\left(\sum_l |g_l|^2 Cum(s_l:2)\right)} \end{aligned} \tag{29}$$

while for any $Cum(y:2) \neq 1$, do perturbation with $2 > \sigma > 0$, let

$$|\tilde{g}_i|^2 = |g_i|^2 \left(1 - \sigma \frac{Cum(y:2) - 1}{Cum(y:2)}\right) \qquad (i = 1, 2, \cdots, n) \tag{30}$$

we obtain

$$E(\tilde{y}) = E(y) = (2\sigma - \sigma^2)(Cum(y:2) - 1)^2 > 0. \tag{31}$$

Proof completed.

Comparing with the basic PICA model, multireference PICA network gives us more flexibility to extract the asymmetric information of the IC source. In the next part, we will give some examples to show the powerful feature of the $f(\,)$ function in applications.

IV. Simulation Results

In the first experiment, we suppose there are two sub-Gaussian IC sources. The receivers and the IC sources are shown in Fig. 3.

Suppose the only prior information in hand is that receiver x_1 is relatively closer to IC source s_1 than receiver x_2, while it is relatively further to s_2 than x_2. In other words, if x_1, x_2 can be expressed by

$$\begin{aligned} x_1 &= m_{11} s_1 + m_{12} s_2 \\ x_2 &= m_{21} s_1 + m_{22} s_2. \end{aligned} \tag{32}$$

then the prior information here is $|m_{11}|^2 > |m_{21}|^2$ and $|m_{12}|^2 < |m_{22}|^2$. Now we simply choose the reference signals

$$E(y) = \frac{f(Cum(y, r_1:4), Cum(y, r_2:4), \cdots, Cum(y, r_v:4))}{Cum(y:2)} - [Cum(y:2) - 1]^2 \tag{23}$$

as $r_1 = x_1$ and $r_2 = x_2$, the $f(\,)$ function is set to be

$$f(x_1, x_2) = -(x_1 - x_2). \tag{33}$$

Notice here both the two IC's are sub-Gaussian, we have $Cum(s_1:4) < 0, Cum(s_1:4) < 0$. According to proposition 2, since

$$\begin{aligned} f(Ng\_E(s_1|r_1), &Ng\_E(s_1|r_2)) \\ &= (|m_{11}|^2 - |m_{21}|^2)Cum(s_1:4) \\ &> 0 \\ &> -(|m_{12}|^2 - |m_{22}|^2)Cum(s_2:4) \\ &= f(Ng\_E(s_2|r_1), Ng\_E(s_2|r_2)) \end{aligned} \tag{34}$$

by maximizing the following cost function:

$$\begin{aligned} E(y) = &\frac{-[Cum(y, r_1:4) - Cum(y, r_2:4)]}{Cum(y:2)} \\ &- [Cum(y:2) - 1]^2 \end{aligned} \tag{35}$$

the network output will converge to IC s_1.

In the computer simulation, s_1 is a sub-Gaussian QAM signal while s_2 is a 3 × 3 sub-Gaussian QAM signal of the same distribution. The mixing matrix is randomly chosen as

$$A = \begin{bmatrix} -2.3124 - 1.7902i & 1.7101 + 1.3347i \\ 1.2672 + 1.0838i & -2.4063 - 2.329li \end{bmatrix}. \tag{36}$$

We use gradient method and use the similar approach as that in [14] to estimate the high-order moments of the signals. In order to describe the convergence of the network, we use the correlation coefficients defined by

$$\theta_i = \sqrt{\frac{|E\{y^* s_i\}|^2}{E\{|y|^2\}E\{|s_i|^2\}}}. \tag{37}$$

Obviously, if $y \to g_1 s_1$ can be satisfied, $\theta_1 \to 1$ and $\theta_2 \to 0$ will be held true. The weight vector of the network is set to be one initially. And Fig. 4(a) shows the output constellation after 900 iterations while Fig. 4(b) gives the convergence of the output presented by the covariance functions.

In the second experiment, we try to show a more skillful design of the $f(\,)$ function in PICA network. Suppose we have a base-band CDMA emulation system, shown in Fig. 5.

The received signal rec is denoted by linear combination of three sub-Gaussian QAM IC's

$$rec = a_1 s_1 + a_2 s_2 + a_3 s_3. \tag{38}$$

And suppose after the demodulation for each user respectively, the final sampling signal yields

$$\begin{aligned} r_1 &= a_1 s_1 + \frac{a_2}{\lambda} s_2 + \frac{a_3}{\lambda} s_3 + n_1 \\ r_2 &= \frac{a_1}{\lambda} s_1 + a_2 s_2 + \frac{a_3}{\lambda} s_3 + n_2 \\ r_3 &= \frac{a_1}{\lambda} s_1 + \frac{a_2}{\lambda} s_2 + a_3 s_3 + n_3. \end{aligned} \tag{39}$$

Here we use a single variable $\lambda > 1$ to simulate the attenuation of demodulation. n_1, n_2, n_3 are additive white Gaussian noises. The reference signals are set to be r_1, r_2, and r_3. Since

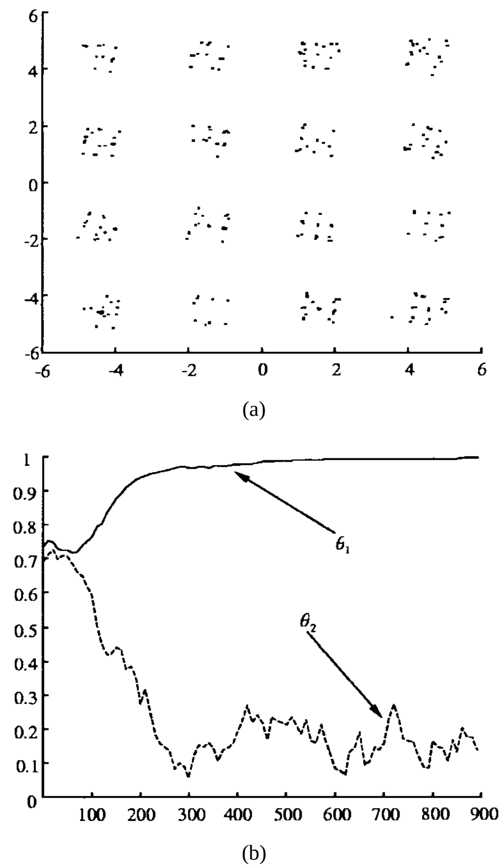

Fig. 4. Simulation of signal tracing. (a) Output constellation after 900 iterations. (b) Convergence of the output presented by the covariance functions.

s_1 is not attenuated only in x_1, the prior information can be expressed by

$$\begin{aligned} -Ng\_E(s_1|r_1) = &-|a_1|^2 Cum(s_1:2)Ng(s_1) \\ &> \tfrac{1}{3}\sum_{i=1}^{3} Ng\_E(s_1|r_i). \end{aligned} \tag{40}$$

Then if we set the cost function to be

$$\begin{aligned} E(y_1) = &\frac{Cum(y_1, r_1:4) - \frac{1}{3}\sum_{i=1}^{3} Cum(y_1, r_i:4)}{Cum(y_1:2)} \\ &- [Cum(y_1:2) - 1]^2. \end{aligned} \tag{41}$$

According to Proposition 2, we will get $y_1 \to s_1$ by maximizing (41). Similarly, by maximizing the following cost functions:

$$\begin{aligned} E(y_2) = &\frac{Cum(y_2, r_2:4) - \frac{1}{3}\sum_{i=1}^{3} Cum(y_2, r_i:4)}{Cum(y_2:2)} \\ &- [Cum(y_2:2) - 1]^2 \end{aligned} \tag{42}$$

$$\begin{aligned} E(y_3) = &\frac{Cum(y_3, r_3:4) - \frac{1}{3}\sum_{i=1}^{3} Cum(y_3, r_i:4)}{Cum(y_3:2)} \\ &- [Cum(y_3:2) - 1]^2 \end{aligned} \tag{43}$$

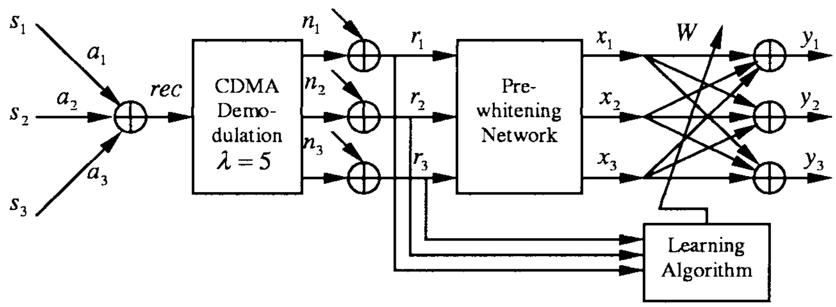

Fig. 5. Base-band CDMA emulation system.

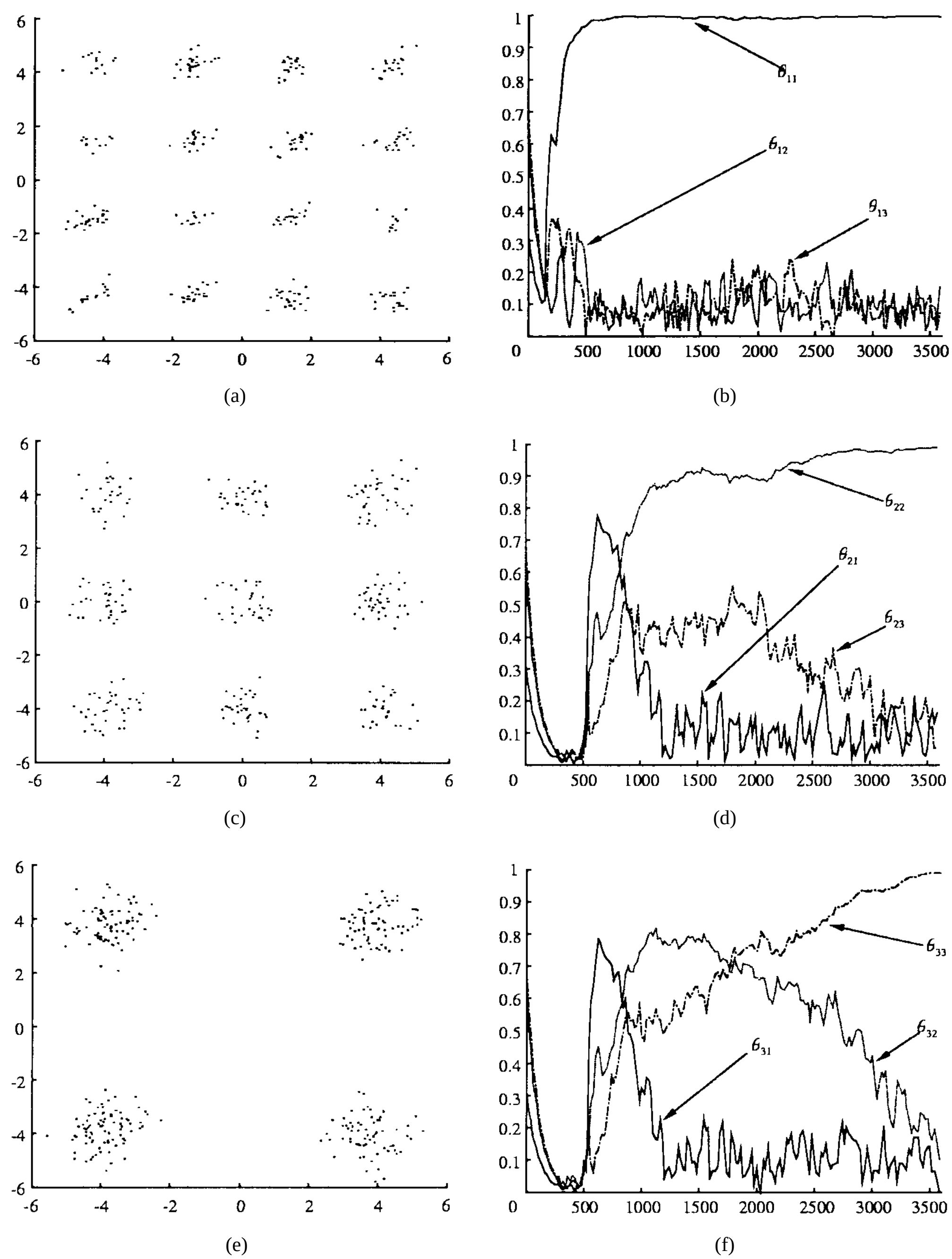

Fig. 6. Base-band CDMA "near-far" resistance using PICA network (SNR = 14 dB). (a) y_1 output constellation. (b) y_1 convergence. (c) y_2 output constellation. (d) y_2 convergence. (e) y_3 output constellation. (f) y_3 convergence.

we can get $y_2 \rightarrow s_2, y_3 \rightarrow s_3$, respectively. In order to improve the convergence, we add a prewhitening process before the PICA network and make

$$E(x_i^* x_j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{44}$$

to be held true.

In our experiment, s_1, s_2, s_3 are 4 × 4, 3 × 3, 2 ×2 2 sub-Gaussian sources, respectively. We set $a_1 = 6, a_2 = 1, a_3 = 1, \lambda = 5$. The SNR of the prewhitening input is set to be 14 db. And the correlation coefficients are given by

$$\theta_{ij} = \sqrt{\frac{|E\{y_i^* s_j\}|^2}{E\{|y_i|^2\}E\{|s_j|^2\}}}. \tag{45}$$

Fig. 6(a), (c), (e) shows the output constellations after 3600 iterations while (b), (d), (f) gives the convergence of the three outputs, respectively.

Here we should mention that, according to Fig. 5, $r_3 = 1.2s_1 + 0.2s_2 + s_3$. We can see the interuser interference s_1 is even larger than the user signal s_3 itself, which means a very serious near-far problem exists. In addtion, our receiver have only a low SNR of 14 dB. Though facing such a hard situation, the PICA network can still extract the objet signal efficiently.

V. Conclusion

A new concept of PICA is proposed. Unlike conventional BSS methods, PICA network focuses its scope on extracting prior information and tracing the object signal directly. Compare with the multioutput BSS algorithms, the single-output PICA network is much simpler in computation complexity. Especially the multireference extension makes the PICA method flexible and powerful in applications.

References

[1] P. Comon, “Independent component analysis, A new concept?,” *Signal Processing*, vol. 36, pp. 287–314, 1994.

[2] E. Oja, “The nonlinear PCA learning rule and signal separation—Mathematical analysis,” Helsinki Univ. Technol., Rep. A26, Aug. 1995.

[3] E. Oja, J. Karhunen, L. Wang, and R. Vigario, “Principal and independent components in neural networks—Recent developments,” in *Proc. VII Italian Wkshp. Neural Nets WIRN’95*, May 18–20, 1995, Vietri sul Mare, Italy, 1995.

[4] J.-F. Cardoso, S. Bose, and B. Friedlander, “On optimal source separation based on second- and fourth-order cumulants,” in *Proc. IEEE SSAP Wkshp.*, Corfou, 1996.

[5] J.-F. Cardoso, “Multidimensional independent component analysis,” in *Proc. ICASSP’98*, Seattle, WA.

[6] C. Jutten and J. Herault, “Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture,” *Signal Processing*, vol. 24, pp. 1–20, 1991.

[7] P. Comon, C. Jutten, and J. Herault, “Blind separation of source, Part II: Problems statement,” *Signal Processing*, vol. 24, pp. 11–20, 1991.

[8] A. Cichocki and R. Unbehauen, “Robust neural networks with on-line learning for blind identification and blind separation of sources,” *IEEE Trans. Circuits and Syst. I*, vol. 43, Nov. 1996.

[9] S. Amari, T.-P. Chen, and A. Cichocki, “Stability analysis of learning algorithms for blind source separation,” *Neural Networks*, vol. 10, no. 8, pp. 1345–1351, Nov. 1997.

[10] R. W. Liu, “Blind signal separation: I-fundamental concepts,” *J. Circuits Syst.*, vol. 1, no. 1, pp. 1–5, 1996.

[11] J. Luo, B. Hu, X.-T. Ling, and R.-W. Liu, “Principal independent component analysis with multireference,” in *IEEE ICA’99*, Jan. 11–15, 1999, Aussois, France, to be published.

[12] A. Hyvarinen and E. Oja, “Independent component analysis by general nonlinear Hebbian-like learning rules,” *Signal Processing*, vol. 64. no. 3, 1998, to be published.

[13] J. M. Mendel, “Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications,” *Proc. IEEE*, vol. 79, Mar. 1991.

[14] O. Shalvi and E. Weinstein, “New criteria for blind deconvolution of nonminimum phase systems (channels),” *IEEE Trans. Inform. Theory*, vol. 36, Mar. 1990.