Discussion on Nonlinear Functions of the Blind Source Separation[1)]

XIE Sheng-Li HE Zhao-Shui ZHANG Jin-Long FU Yu-Li

(*School of Electronics & Information Engineering, South China University of Technology, Guangzhou* 510640)

(E-mail: adshlxie@scut.edu.cn)

Abstract This paper proposes a new algorithm of blind source separation (BSS). The algorithm can overcome the difficulty known as "the sensors are less than the source signals" and works effectively when the sensors are less. Then, the paper discusses the nonlinear functions used in the new algorithm. A uniform nonlinear function is proposed and some criterion are given to choose its parameters. Finally, some simulations are presented to show the effectness of the algorithm and the correctness of the criterion.

Key words Blind source separation (BSS), nonlinear function, stability

1 Introduction

The BSS problem can be explained by the following mathematical model

$$\boldsymbol{x}(t) = A\boldsymbol{s}(t) \tag{1}$$

$$\boldsymbol{u}(t) = W\boldsymbol{x}(t) \tag{2}$$

where $\boldsymbol{x}(t) = (x_1(t), \cdots, x_m(t))^{\mathrm{T}}$ is the observed signal vector, $\boldsymbol{s}(t) = (s_1(t), \cdots, s_n(t))^{\mathrm{T}}$ is the source signal vector to be separated. $A = (a_{ij})_{m\times n}$ is an unknown mixing matrix. BSS purposes to get the separated signal $\boldsymbol{u}(t) = (u_1(t), u_2(t), \cdots, u_n(t))^{\mathrm{T}}$ from the observed signal, by adjusting the separating matrix W, so as to

$$\boldsymbol{u}(t) = WA\boldsymbol{s}(t) = PD\boldsymbol{s}(t) \tag{3}$$

where P is a permutation matrix and D is a diagonal matrix. In other words, the separated signal would be different from the source signal in amplitude and the order of the arrangement, but it is of uniform waveform as the source signal.

BSS is the hot spot in the fields of neural networks (NN) and signal processing[1]. There are many applications or potential applications, for example, echo cancellation, voice enhancement, medical signal processing (EEG, ECG, MEG *etc.*), sonar, wireless communication, optical communication, image processing, and so on. Among these, R. Crsstescu and J. Joutsensalo applied BSS to the blind detection of CDMA wireless communication successfully.

The adaptive BSS algorithms are one important type of the BSS algorithms appeared recent years. This type of algorithms has the following unit form[2]

$$W(k+1) = W(k) + \mu(I - G(\boldsymbol{u}(k))\boldsymbol{u}(k)^{\mathrm{T}})W(k) \tag{4}$$

where

$$g_i(u_i) = -\frac{\mathrm{d}\log p_i(u_i)}{\mathrm{d}u_i} = -p_i'(u_i)/p_i(u_i) \tag{5}$$

$p_i(s_i)$ denotes the probability density function (pdf) of the source signal s_i, and we denote $G(u) = (g_1(u_1), \cdots, g_n(u_n))^{\mathrm{T}}$. Since the pdfs are unknown, the different nonlinear functions cause different

1) Supported by Science Foundation of Guangdong Province for Program of Research Team (04205783), National Natural Science Foundation of P. R. China (60274006), the National Natural Science Foundation of P. R. China for Excellent Youth (60325310), and the Trans-Century Training Programme Foundation for the Talents by the State Education Commission

Received April 14, 2004; in revised form July 27, 2005

BSS algorithms, for example, Gaussian kernel based nonlinear functions[3], information theory based nonlinear functions[4], the fourth-order cumulant based nonlinear functions[5], neural network based nonlinear functions[6,7], adaptive processing based nonlinear functions, and switching based nonlinear functions[9∼11]. Although the algorithms are partially successful, we found that mixtures of some signals could not be separated. For this reason, [2] gave a statistical algorithm with the unified form. However, for selecting the nonlinear functions, there is no criterion proposed.

On the other hand, most of the existing algorithms are based on the condition "the sensors are not less than the source signals". This means the mixing matrix is of full column rank. In this case, the signals could be separated simultaneously. When the sensors are less than the number of source signals, *i.e.*, the mixing matrix is not of full column rank, BSS is changed into sequential extraction, which is difficult.

To solve the problem, we improve the above algorithm and propose an algorithm that separates the source signals one by one. This algorithm can work when the sensors are less than the source signals. Then, we study the nonlinear functions in the algorithm in detail. A uniform nonlinear function with parameters is proposed. By some analysis of stability and the fourth-order cumulant, we present some criteria for choosing the parameters in the nonlinear functions. At last, some simulations are given to demonstrate the availability of the algorithm and the validity of the criteria for choosing the parameters.

2 A new algorithm of BSS

When the source signals cannot be entirely separated, the simultaneous separation algorithm (4) will fail. So we suggest some algorithm to separate the source signals sequentially. We change the optimal problem into another optimal problem with a constraint[7]

$$\begin{cases} \max G(\boldsymbol{u}) \\ \|\boldsymbol{w}\|^2 = 1 \end{cases} \tag{6}$$

Also, the optimal problem (6) can be changed into

$$\max P(w) = \max\{G(w) + \lambda(\|w\| - 1)^2\} \tag{7}$$

By using stochastic gradient descent method, we obtain a new learning algorithm. From

$$\nabla \boldsymbol{w} = \frac{\partial P(w,\beta)}{\partial w} = (\boldsymbol{x} g(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x})) + 2\lambda w \tag{8}$$

where

$$g(\boldsymbol{u}) = -\frac{\mathrm{d}\log p(u)}{\mathrm{d}\boldsymbol{u}} = -p'(u)/p(u) \tag{9}$$

the new algorithm is

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) + \alpha \boldsymbol{x}(t) g(\boldsymbol{w}^{\mathrm{T}}(t)\boldsymbol{x}(t)) + 2\alpha\lambda\boldsymbol{w}(t) \tag{10}$$

For the sake of enhancing the stability of the algorithm, we set normalization for every iterative step. The algorithm is

$$\begin{cases} \boldsymbol{w}^{new}(t+1) = \boldsymbol{w}(t) + \alpha(t)[\boldsymbol{x}(t) g(\boldsymbol{w}(t)^{\mathrm{T}}\boldsymbol{x}(t)) + 2\lambda\boldsymbol{w}(t)] \\ \boldsymbol{w}(t+1) = \dfrac{\boldsymbol{w}^{new(t+1)}}{\|\boldsymbol{w}^{new}(t+1)\|} \end{cases} \tag{11}$$

The procedure is given as follows.

1) Set the initial vector $\boldsymbol{w}(0)$;

2) Set the step length $\alpha(t)$ and the nonlinear function $g(\bullet)$. Take iteration of (10) or (11) until $\boldsymbol{w}$ converges;

3) Substituting $\boldsymbol{w}$ into $u(t) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}(t)$, we get the separated signals. Using the method proposed in [1], we eliminate the separated signal $u(t)$ from the observed signal $\boldsymbol{x}(t)$. Repeat steps 1)∼3) until every source signal is separated.

3 Discussion of the nonlinear functions

Theoretically, the nonlinear functions are related to the pdfs of the source signals. The relations can be expressed by (7). However, since the pdfs are unknown, the nonlinear functions could not be determined easily. Most of the algorithms (see [4, 5]) select the nonlinear functions experientially. To Choose the nonlinear functions theoretically, we discuss the problem as follows.

3.1 Choosing the nonlinear function

We choose the following nonlinear function in our algorithm

$$g(u) = au - bu^3 \tag{12}$$

Many nonlinear functions in the existing algorithms can be approximated by the function (12) by adjusting the parameters a and b. For example, we have

1) the nonlinear function proposed in [4], based on minimal mutual information.

$$g(u) = \frac{3}{4}u^{11} + \frac{25}{4}u^9 - \frac{14}{3}u^7 - \frac{47}{4}u^5 + \frac{29}{4}u^3 \tag{13}$$

2) the nonlinear function led by the fourth-order cumulant in [5]

$$g(u) \approx u^3 - \frac{m_4}{m_2}u \tag{14}$$

3) the nonlinear function comes from the neural network method, given by [6]

$$g(u) = \begin{cases} \tanh(au) = au - \dfrac{1}{3}a^3u^3, & \sup er - Gaussian \\ u^3, & sub - Gaussian \end{cases} \tag{15}$$

4) the nonlinear function based on a switching in [9]

$$g(u) = \begin{cases} u + \tanh(u) = 2u - \frac{1}{3}^3, & \sup er - Gaussian \\ u - \tanh(u) = \frac{1}{3}u^3, & sub - Gaussian \end{cases} \tag{16}$$

5) the adaptive process based nonlinear function proposed in [8]

$$g(u) = \alpha u^3 + (1-\alpha)\tanh(u) = \alpha u^3 + (1-\alpha)(u - \frac{1}{3}u^3) = \frac{4\alpha - 1}{3}u^3 + (1-\alpha)u \tag{17}$$

The curves of the nonlinear functions are demonstrated in Fig. 1. By change of the parameters, the nonlinear function (12) can cover the nonlinear functions of the existing algorithms.

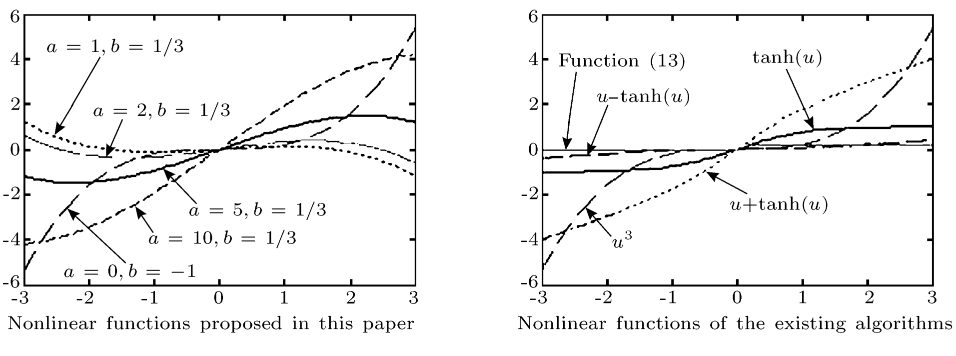


Fig. 1 Nonlinear functions

3.2 Distribution of the source signals and the pdfs of the output signals

Signals can be categorized to Gaussian signals, super-Gaussian signals and sub-Gaussian signals. They can be approximately described by the generalized Gaussian distribution density model (see [11]):

$$p(u, \alpha) = \frac{\alpha}{2\lambda\Gamma(1/\alpha)} e^{-|\frac{u}{\lambda}|^\alpha} \tag{18}$$

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\mathrm{d}t$, while the parameters α and λ determine the shape and the variance of the density function respectively. As parameter α increases/decreases, the signal is more sub-Gaussian/super-Gaussian. If $\alpha \to \infty$ the signal is approximating to a uniform distribution.

In addition, according to (9), we get the pdf of output signal:

$$p(u) = \frac{1}{2\beta}e^{-(\frac{au^2}{2}-\frac{bu^4}{4})} \tag{19}$$

where β is used to control $\int_{-\infty}^{\infty} p(u)\mathrm{d}u = 1$. The pdf is changed with parameters a and b. It is shown in Fig. 2 that when parameter a becomes larger, the signal becomes more super-Gaussian and the signal is sub-Gaussian when $a = 0$. Also, we can see that the generalized Gaussian distribution can imply three kinds of signals with respect to Gaussian statistical property of signal and is very representative. Therefore, the generalized Gaussian distribution can be used to separate sub-Gaussian signals and super-Gaussian signals as well.

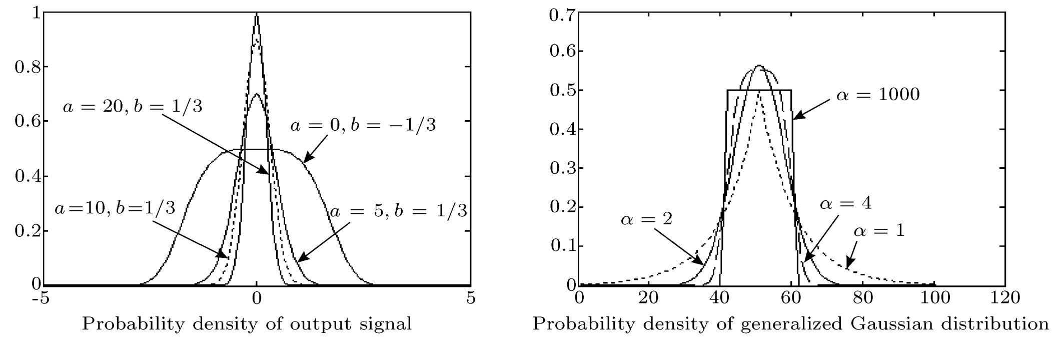


Fig. 2 Probability density

3.3 Local stability analysis

Although the nonlinear function and the pdf satisfy (9) theoretically, in practice it is not required that (9) hold accurately, and the nonlinear function can guarantee stability of the algorithm when errors exist in some sense and the algorithm can still separate signals correctly (see [2]). In the discussion of local stability about the nonlinear function, many algorithms require the following inequality holds:

$$E\{g'(u)\}E\{u^2\} - E\{g(u)\}E\{u\} > 0 \tag{20}$$

Considering the nonlinear function in this paper, we have

$$\begin{aligned} E\{g'(u)\}E\{u^2\} - E\{g(u)\}E\{u\} = E(a-3bu^2)E(u^2) - E\{(au-bu^3)u\} = \\ bE(u^4) - 3bE^2(u^2) = b\cdot kurt(u) \end{aligned} \tag{21}$$

and then,

$$\begin{cases} b > 0, & kurt(u) > 0 \\ b < 0, & kurt(u) < 0 \end{cases} \tag{22}$$

3.4 Parameters of the nonlinear function

Now we discuss the parameters. From the above local stability analysis, we can see that it is required that inequality $b > 0$ holds when the source signal is sub-Gaussian and it is required that inequality $b < 0$ holds when the source signal is super-Gaussian. Firstly, we discuss the fourth-order cumulant of the generalized Gaussian distribution in order to choose parameter a.

$$kurt(\alpha) = \frac{\Gamma(5/\alpha)\Gamma(1/\alpha)}{\Gamma^2(3/\alpha)} - 3 \tag{23}$$

It is not difficult to get $\lim_{\alpha\to 0} kurt(\alpha) = \infty$, and $\lim_{\alpha\to\infty}(\alpha) = -1.2$. The fourth-order cumulant changes with parameter a, as shown in Fig. 3 and Table 1.

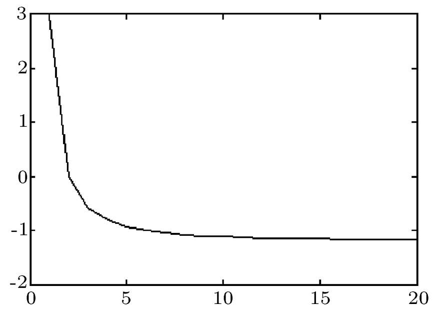


Fig. 3 The *kurt* VS parameter *a*

Table 1 *Kurt* of generalized Gaussian signals

α	∞	10000	100	50	20	8	5	3	2
kurt	−1.2	−1.2	−1.1989	−1.1956	−1.1756	−1.0766	−0.9299	−0.5816	0
α	1.8	1.7	1.5	1.4	1.2	1	0.5	0.3	0.1
kurt	0.2324	0.3788	0.7620	1.0179	1.7435	3	22.2	170.9691	∞

To study the monotonicity of the four-order cumulant, we compute the derivative of the fourth-order cumulant, *i.e.*,

$$kurt'(\alpha) = \lim_{\Delta\alpha \to 0} \frac{kurt(\alpha + \Delta\alpha) - kurt(\alpha)}{\Delta\alpha} \tag{24}$$

From Fig. 3, we can see that four-order cumulant $kurt(\alpha)$ is a decreasing function with respect to α, and that when the signal is sub-Gaussian, *i.e.*, $\alpha > 2$ the fourth-order cumulant changes very slowly, and its slope approximates to 0. However, when the signal is super-Gaussian, *i.e.*, $\alpha < 2$, $kurt(\alpha)$ changes with parameter α remarkably. From the discussion of section 3.2, we know that the Gaussian statistical property of signal is controlled by parameter a and the Gaussian measure of signal is also controlled by parameter a. So we choose a as follows.

$$a = d\ kurt'(\alpha) \tag{25}$$

where d is a constant. Obviously, $a = 0$ when signal is a sub-Gaussian and the nonlinear function is $g(u) = u^3$, which is consistent with the choice of the general algorithm; For super-Gaussian signal, the signal is more Gaussian with the increasing of parameter a. And the signal is more Gaussian, the derivative (slope) of the fourth-order cumulant is larger. Accordingly, we should choose large parameter a for signal with large fourth-order cumulant.

4 Experiments

In order to check up the validity of the algorithm, we demonstrate the selections of parameters a and b using image signals, and voice signals, respectively. The simulation results are as follows.

The fourth-order cumulants of above signals are shown in Table 2.

Table 2 *Kurts* of source signals

Figure	4		5		6	
Source signal	s_1	s_2	s_1	s_2	s_1	s_2
kurt	−0.7492	−0.9085	−0.7492	12.0477	30.1449	21.6491

From Table 2, we can see two signals in Fig. 4 are sub-Gaussian. Fig. 4 shows that BSS of "c" is successful using the criteria in this paper while "d" and "e" fail. In Fig. 5, a super-Gaussian signal and a sub-Gaussian signal are included. During the procedure of BSS, the source signal with a larger absolute value of the fourth-order cumulant is separated at first. The "e" in Fig. 5 is successful while "c" and "d" fail. The "c" does not follow the criteria and "d" chooses the wrong parameter a. With regard to Fig. 6, two source signals are super-Gaussian, so "c" fails while "d", "e", "f" and "g" succeed.

Further more, parameter a suits to separate source signal s_2 and it is firstly separated in "d", "e" and "f" while in "g" parameter a suits to separate source signal s_1 and s_2 is separated firstly. Obviously, we can see that if the fourth-order cumulant is larger, parameter a should be larger. But parameter a cannot be arbitrarily large. The failure of "h" is an example.

a) Original image b) Mixed image c) a=0, b=−1/3 d) a=1, b=1/3 e) a=2, b=1/3

Fig. 4 BSS of two sub-Gaussian signals mixed linearly

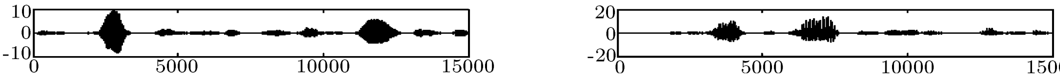
a) Original image b) Mixed image c) a=0, b=−1/3 d) a=1, b=1/3 e) a=4, b=1

Fig. 5 BSS of a sub-Gaussian and a super-Gaussian signals mixed linearly

a) Source signals

b) Mixed signals

c) Separated signals when $a = 0, b = -1/3$



d) Separated signals when $a = 1, b = 1/3$

e) Separated signals when $a = 2, b = 1/3$

f) Separated signals when $a = 5, b = 1/3$

g) Separated signals when $a = 8, b = 1/3$

h) Separated signals when $a = 15, b = 1/3$

Fig. 6 BSS that two super-Gaussian signals are linearly mixed

In addition, we have done the simulation when mixed signals are less than the number of source signals. Computation shows that four voice signals are all super-Gaussian. Three mixed signals are obtained. We choose parameters such that $a = 2, b = -1/3$.

The simulations are as following Fig. 7.

a1) Source signal $s_1(t)$ a2) Source signal $s_2(t)$

a3) Source signal $s_3(t)$ a4) Source signal $s_4(t)$

b1) Mixture $x_1(t)$ b2) Mixture $x_2(t)$

b3) Mixture $x_3(t)$ c1) Separated signal $u_1(t)$

c2) Separated signal $u_2(t)$ c3) Separated signal $u_3(t)$

Fig. 7 The simulations

From the above simulation, we can see that two signals are successfully separated, and it shows that the algorithm is robust.

5 Conclusion

This paper presents a BSS algorithm that can work when sensors are less than the number of source signals. Then, the nonlinear functions of BSS algorithms are discussed in detail. An uniform nonlinear function with parameters a and b is proposed. Combining with stability analysis of the algorithm, we give the criteria for choosing parameter b, and propose the criteria for choosing parameter a by the fourth-order cumulant of signal. Finally, we demonstrate how to choose parameters a and b, and experiments show the algorithm is valid and the criteria for choosing parameters is correct.

References

1 Zhang Jin-long, Xie Sheng-Li, He Zhao-Shui. Separability theory for blind signal separation. *Acta Automatia Sinica*, 2004, **30**(3): 337～344

2 Lin Qiu-Hua, Yin Fu-Liang. Unified form of several adaptive algorithms for blind source separation. *Journal of Dalian University of Technology*, 2002, **42**(4): 498～503

3 He Zhen-Ya, Yang Lu-Xi, Liu Ju, Lu Zi-yi, He Chen, Shi Yu-Hui. Blind source separation using clustering-based multivariate density estimation algorithm. *IEEE Transactions on Signal Processing*, 2000, **48**(2): 575～579

4 Amari S, Cichocki A, Yang H H. A new algorithm for blind source separation. Advances in Neural Information Processing 8 (Proc.NIPS′95), Cambridge, MA: MIT Press, 1996. 757～763

5 Cichocki A, Thawonmas R, Amari S. Sequential blind signal extraction in order specified by stochastic properties. *Electronics Letters*, 1997, **33**(1-2): 64～65

6 Andrzej Cichocki, Wlodzimierz Kasprzak. Local adaptive learning algorithms for blind separation of natural images. *Neural Network World*, 1996, **6**(4): 515～523

7 Erkki Oja. From neural learning to independent components. Neurocomputing, 1998, 22: 187～199

8 Ihm B C, Park D J, Kwon Y H. Blind separation for mixtures of sub-Gaussian and super-Gaussian sources, Circuits and Systems, 2000. In: Proceedings of ISCAS 2000 Geneva: The 2000 IEEE International Symposium, 2000. **3**: 738～741

9 Lee T W, Girolami M, Sejnowski T. Independent component analysis using an extended Informax algorithm for mixed sub-Gaussian and super-Gaussian sources. *Neural Computation*, 1999, **11**(2): 609～633

10 Cao Jian-Ting, Murata N. A stable and robust ICA algorithm based on t-distribution and generalized Gaussian distribution models. Neural Networks for Signal Processing IX, 1999. In: Proceedings of the 1999 IEEE Signal Processing Society Workshop. IEEE Press, 1999. 283～292

11 Choi S, Cichocki A, Amari S. Flexible independent Component Analysis. Neural Networks for Signal Processing VIII, 1998. In: Proceedings of the 1998 IEEE Signal Processing Society Workshop. New York: IEEE Press, 1998. 83～92

XIE Sheng-Li Professor in the Institute of Radio and Automation at South China University of Technology. He is also a senior member of IEEE. His research interests include nonlinear learning control, robotic systems, adaptive acoustic echo cancellation, blind signal processing, and image processing.

HE Zhao-Shui Ph. D. candidate in the Institute of Radio and Automation at South China University of Technology. His research interests include blind source separation.

ZHANG Jin-Long Received his master and Ph. D. degrees in 2000 and 2003, respectively. His research interests include blind signal processing and optimization of wireless networks.

FU Yu-Li Received his Ph. D. degree in system engineering from Huazhong University of Science & Technology, China, in 2000. From 2001 to 2002, he worked with Department of Statistics & Modeling Science, Strathclyde University, Scotland, as a postdoctoral researcher of Royal Society UK. Now he is an associate professor with the School of Electronics & Information Engineering. His research interests include blind signal processing, channel equalization, and learning theory and artificial neural networks.