A Functional Link Network With Higher Order Statistics for Signal Enhancement

Bor-Shyh Lin, Bor-Shing Lin, Fok-Ching Chong, and Feipei Lai

Abstract—A functional link network with higher order statistics is introduced for signal enhancement. The proposed scheme uses the mean-square error (MSE) between higher order statistics of desired signals and filtered output as the learning criterion for training weights in the functional link network. This is motivated by the fact that higher order statistics have a natural tolerance to Gaussian and symmetrically distributed non-Gaussian noises. Results show that the performance of functional link network with higher order statistics is less sensitive to the selection of learning rates than the conventional functional link network and adaptive line enhancement. It is also demonstrated that it can enhance signal more effectively under different noise levels for stationary and nonstationary Gaussian noises.

Index Terms—Functional link network, higher order statistics, signal enhancement.

I. INTRODUCTION

Signal enhancement is an important technique of statistics signal processing with direct application in many fields, such as engineering, biomedical, and econometric models. Adaptive filtering techniques are widely used in signal enhancement problem [1]-[3]. Signal enhancement can be considered as a mapping from the noisy input space to the noise-free output space. The original scheme of adaptive filtering for signal enhancement was proposed by Widrow et al. [1] in 1975. However, in many applications, the signal of interest (SOI) is nonlinear. It is difficult to adapt to a nonlinear signal using a linear model. To overcome this issue, nonlinear filters are developed. Neural networks are considered as an alternative for nonlinear signal enhancement. Based on multilayer perceptron (MLP) architecture, it needs far fewer filter weights and a small amount of training data to perform an approximation of a nonlinear function [4]-[8]. Gandhi and Ramamurti employed a three-layer neural network for signal detecting in non-Gaussian noise [7]. However, it is just suitable for those problems that do not have time restrictions because of the slow learning rate of MLP.

In 1989, Pao proposed the functional link network (FLN) [8]. The FLN is a universal approximator [9], [10]. It has a similar structure to the three-layer MLP, except that instead of employing enhancement nodes in the hidden layer, the network between input space and enhancement nodes is referred to as functional links. The output of the FLN is a linear sum of enhancement nodes. The mapping between the input nodes and enhancement nodes in the FLN is fixed. Therefore, the learning algorithm of the FLN only updates those weights that contribute to the output. Thus, the amount of learning in the FLN is significantly reduced. However, the mean-square error (MSE) between desired and filtered output signals is commonly employed as the learning criterion of both the MLP and the FLN for training weights. Its learning is directly influenced by additive noise in desired signals.

Recently higher order statistics (or cumulants) techniques were designed for signal enhancement [11]-[14]. Higher order statistics can

Manuscript received January 10, 2006; revised February 18, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Sergios Theodoridis.

B.-Shyh Lin, B.-Shing Lin and F.-C. Chong are with the Institute of Electrical Engineering, National Taiwan University, Taipei, Taiwan, R.O.C. (e-mail: borshyhlin@ntu.edu.tw).

F. Lai is with the Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, R.O.C.

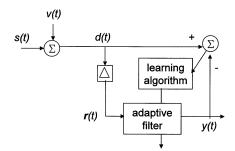


Fig. 1. Basic scheme for signal enhancement.

provide the nature of suppressing Gaussian noise process of unknown spectral characteristics in deterministic signals [14]. In this study, we make use of this advantage of higher order statistics in developing learning algorithm of the FLN. In the proposed method, the MSE between higher order cumulants of desired signals and filtered outputs is used as the learning criterion for training weights. By reducing the influence of additive noises in desired signals on learning, the performance of the FLN for signal enhancement can be effectively improved.

II. PROBLEM DEFINITION

The basic scheme of signal enhancement, shown in Fig. 1, was derived by Widrow *et al.* in 1975 [1]–[3]. It is used to detect a periodic signal buried in a broadband noise background. Let d(t) denote the measurement of the primary sensor, satisfying

$$d(t) = s(t) + v(t) \tag{1}$$

where s(t) and v(t), respectively, denote SOI and additive Gaussian uncorrelated noises at iteration t. In this scheme, an adaptive filter is treated as a noise-free function. Here, the measured signal of the primary sensor is used as the desired signal. The delayed version of the desired input d(t) is commonly used as reference signals r(t), i.e.,

$$r(t) = d(t - \Delta) \tag{2}$$

where the delay Δ is named the prediction depth and has to be capable to remove the correlation of the noises respectively in desired signals and reference signals. The unit of sampling period is used as the prediction depth Δ [2], [3]. Therefore, the estimate $\hat{s}(t)$ of s(t) can be obtained by the filtered output y(t).

III. FUNCTIONAL LINK NETWORKS

The basic scheme of the FLN is shown in Fig. 2. Taking a perceptron with N_0 input nodes and N_1 enhancement nodes as an example, the output function calculated by this network is

$$y(t) = \delta^{(1)}(t)^T \mathbf{w}^{(1)}(t) = \hat{s}(t)$$
 (3)

where $\mathbf{w}^{(1)}(t) = [w_{11}^{(1)}(t), w_{12}^{(1)}(t), \dots, w_{1N_1}^{(1)}(t)]^T$ denotes the vector of weights between enhancement nodes and the output of the FLN. $\boldsymbol{\delta}^{(1)}(t) = [\delta_1^{(1)}(t), \delta_2^{(1)}(t), \dots, \delta_{N_1}^{(1)}(t)]^T$ is the vector of enhancement nodes and

$$\delta_k^{(1)}(t) = \operatorname{sigm}\left(\sum_{l=1}^{N_0} w_{kl}^{(0)} \delta_l^{(0)}(t)\right)$$
(4)

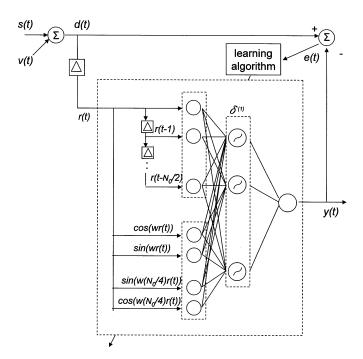


Fig. 2. Scheme of functional link networks.

where the sigmoidal function is given by

$$sigm(net) = \frac{1}{1 + e^{-\beta net}}.$$
 (5)

Here, $w_{kl}^{(0)}$ is the weight between the lth input node and the kth enhancement node and is randomly given. $\delta_l^{(0)}(t)$ denotes the lth input node. In this study, the input nodes consist of the sequence of reference signals and its orthonormal basis functions. The orthonormal basis functions of reference signals are used to enhance the representation of input space.

The mean-square output error y(t) - d(t) is commonly defined as the criterion of learning. Therefore, $\mathbf{w}^{(1)}(t)$ can be adapted by using gradient method [3]

$$\mathbf{w}^{(1)}(t+1) = \mathbf{w}^{(1)}(t) - \frac{\mu}{\left(a + \boldsymbol{\delta}^{(1)}(t)^T \boldsymbol{\delta}^{(1)}(t)\right)} \boldsymbol{\delta}^{(1)}(t) \left(y(t) - d(t)\right) \quad (6)$$

where μ is the learning rate, and a is a small constant. Since d(t) contains noises v(t), the system output noise e(t) is also correlated to noise v(t), as follows:

$$e(t) = d(t) - y(t) = s(t) + v(t) - \hat{s}(t). \tag{7}$$

From (6), we can find that (7) is affected directly by uncorrelated noises.

IV. FUNCTIONAL LINK NETWORKS WITH HIGHER ORDER STATISTICS

A. Learning Algorithm With Higher Order Statistics

For a set of n real variables $\{x_i(t)\}$, $i=1,2,\ldots,n$, the nth-order cross cumulant of $\{x_i(t)\}$ is defined by

$$C_{x_1 x_2 \dots x_n}(\tau_1, \tau_2, \dots, \tau_{n-1}) \stackrel{\Delta}{=} \operatorname{Cum}[x_1(t), x_2(t+\tau_1), x_3(t+\tau_2), \dots, x_n(t+\tau_{n-1})].$$
 (8)

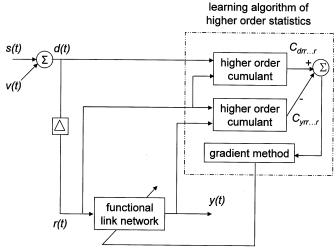


Fig. 3. Scheme of functional link networks with higher order statistics.

Here, $Cum[\bullet]$ in (8) denotes the cumulant operator. Other properties of cumulants are described in [14].

Under the assumption that there exists the nth-order cumulants of the desired and reference signals and they are not identically zero, they are given by

$$C_{drr...r}(\tau_1, \tau_2, \dots, \tau_{n-1}) = \operatorname{Cum} [d(t), r(t+\tau_1), r(t+\tau_2), \dots, r(t+\tau_{n-1})].$$
 (9)

Since s(t) and v(t) are independent, and the *n*th-order cumulants of Gaussian noises are identically zero, (9) can be expressed by

$$C_{drr...r}(\tau_{1}, \tau_{2}, ..., \tau_{n-1})$$

$$= Cum [s(t), s(t + \tau_{1} - \Delta), ..., s(t + \tau_{n-1} - \Delta)]$$

$$= C_{ss...s}(\tau_{1} - \Delta, \tau_{2} - \Delta, ..., \tau_{n-1} - \Delta).$$
(10)

Similarly, we have

$$C_{yrr...r}(\tau_1, \tau_2, \dots, \tau_{n-1}) = C_{\hat{s}s...s}(\tau_1 - \Delta, \tau_2 - \Delta, \dots, \tau_{n-1} - \Delta).$$
 (11)

By the properties of (10) and (11), the basic idea behind the proposed method, shown in Fig. 3, is that the higher order cross cumulants of the desired signal and filtered output are used as the learning criterion for training weights to reduce the influence of additive noises.

$$\xi = \sum_{\tau_{1}} \sum_{\tau_{2}} \dots \sum_{\tau_{n-1}} \frac{1}{2} \left[C_{yrr...r}(\tau_{1}, \tau_{2}, \dots, \tau_{n-1}) - C_{drr...r}(\tau_{1}, \tau_{2}, \dots, \tau_{n-1}) \right]^{2}$$

$$= \sum_{(\tau_{1}, \tau_{2}, \dots, \tau_{n-1}) \in \Gamma} \dots$$

$$\sum_{(\tau_{1}, \tau_{2}, \dots, \tau_{n-1}) \in \Gamma} \dots$$

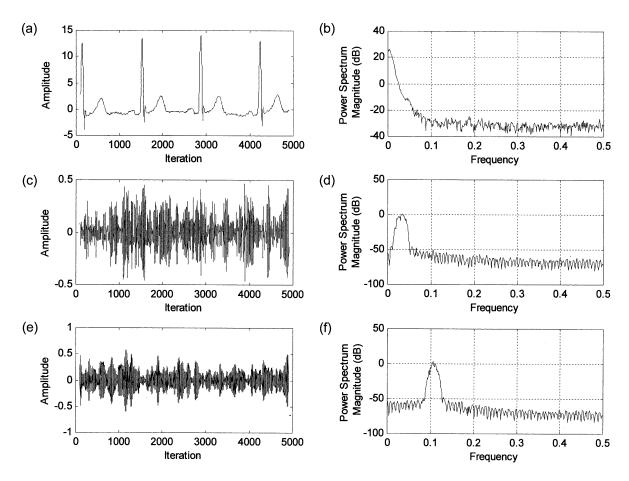


Fig. 4. (a) The waveform of SOI for simulation, and (b) its power spectrum; (c) the waveform and (d) power spectrum of additive Gaussian noises distributed in [0.05, 0.075]; (e) the waveform (f) and power spectrum of additive noises distributed in [0.2, 0.225].

Equation (12) can be rewritten in a matrix form

$$\xi = \frac{1}{2} \left[\mathbf{C}_{\delta^{(1)}\mathbf{r}\mathbf{r}...\mathbf{r}} \mathbf{w}^{(1)} - \mathbf{C}_{\mathbf{d}\mathbf{r}\mathbf{r}...\mathbf{r}} \right]^{2}. \tag{13}$$

Here, $\mathbf{C_{\delta^1}}_{\mathbf{rr...r}}$, and $\mathbf{C_{drr...r}}$ are, respectively, an $M_{\Gamma} \times N_1$ matrix and an $M_{\Gamma} \times 1$ column vector, and M_{Γ} denotes the number of points in the set Γ .

In order to minimize ξ , the gradient descent method is used. The gradient of ξ is given by

$$\nabla_{\mathbf{w}^{(1)}}(t) \equiv \frac{\partial \xi}{\partial \mathbf{w}^{(1)}(t)}$$

$$= 2 \left[\mathbf{C}_{\boldsymbol{\delta}^{(1)}\mathbf{rr}...\mathbf{r}}^{T} \mathbf{C}_{\boldsymbol{\delta}^{(1)}\mathbf{rr}...\mathbf{r}} \mathbf{w}^{(1)} - \mathbf{C}_{\boldsymbol{\delta}^{(1)}\mathbf{rr}...\mathbf{r}}^{T} \mathbf{C}_{\mathbf{drr}...\mathbf{r}} \right]. (14)$$

Consequently, the adaptation formula in this algorithm becomes

$$\mathbf{w}^{(1)}(t+1) = \mathbf{w}^{(1)}(t) - \frac{\mu}{tr\left(\mathbf{C}_{\boldsymbol{\delta}^{(1)}\mathbf{rr}...\mathbf{r}}^{T}\mathbf{C}_{\boldsymbol{\delta}^{(1)}\mathbf{rr}...\mathbf{r}}\right)} \nabla_{\mathbf{w}^{(1)}}(t). \tag{15}$$

From (12), it is straightforward that the influence of v(t) on (15) is reduced.

B. Implementation of Learning Algorithm With Third Order Statistics

In practice, the theoretical higher order cumulants need to be substituted by their estimations. The estimate of third-order cumulants can be recursively computed by

$$\hat{C}_{x_1x_2x_3}(t;\tau_1,\tau_2)
= \langle x_1(t)x_2(t+\tau_1)x_3(t+\tau_2) \rangle
- \langle x_1(t) \rangle \langle x_2(t+\tau_1)x_3(t+\tau_2) \rangle
- \langle x_2(t+\tau_1) \rangle \langle x_1(t)x_3(t+\tau_2) \rangle
- \langle x_3(t+\tau_2) \rangle \langle x_1(t)x_2(t+\tau_1) \rangle
+ 2 \langle x_1(t) \rangle \langle x_2(t+\tau_1) \rangle \langle x_3(t+\tau_2) \rangle.$$
(16)

For zero-mean signals, it can be given by

$$\hat{C}_{x_1 x_2 x_3}(t; \tau_1, \tau_2) = \langle x_1(t) x_2(t + \tau_1) x_3(t + \tau_2) \rangle. \tag{17}$$

Here, the operation $\langle \bullet \rangle$ is given by

$$\langle f(t) \rangle = \lambda \langle f(t-1) \rangle + (1-\lambda)f(t)$$
 (18)

where λ is a forgetting factor.

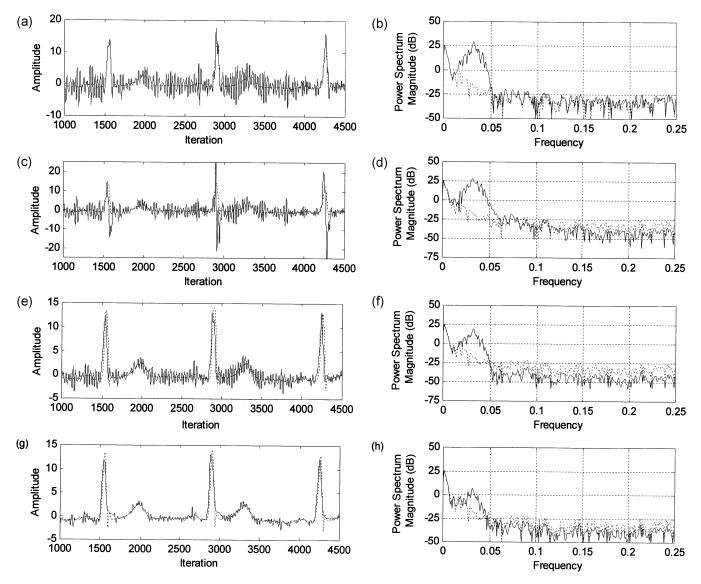


Fig. 5. (a) Simulated trials with the SNR of -5 dB; (b) its power spectrum; the filtered outputs of (c) ALE, (e) FLN, and (g) FLN-TOS, and their power spectra of (d) ALE, (f) FLN, and (h) FLN-TOS; solid line denotes filtered outputs, whereas dashed line denotes SOI.

Therefore, for third-order statistics, the learning criterion can be expressed as

$$\xi = \sum_{\tau_1 = -m_1}^{m_1} \sum_{\tau_2 = 1}^{m_2} \frac{1}{2} \left[\sum_{j=1}^{N_1} w_{j1}^{(1)}(t) \hat{C}_{\delta_j^{(1)}rr}(t; \tau_1, \tau_2) - \hat{C}_{drr}(t; \tau_1, \tau_2) \right]^2.$$
(19)

V. RESULTS AND DISCUSSION

In some simulations, comparisons are conducted for the performance of adaptive line enhancement (ALE) with normalized least mean-square algorithm [3], the FLN, and the FLN with third-order statistics (FLN-TOS) for nonlinear signal enhancement. These comparisons are mainly presented in terms of the MSE between the SOI and the system output obtained by each above method. In this study, we assume that the nonlinear SOI is the electrocardiogram (ECG) pattern in MIT/BIH database shown in Fig. 4(a), and its power

spectrum is shown in Fig. 4(b). The additive Gaussian noises and their power spectrum are respectively shown in Fig. 4(c) and (e) and Fig. 4(d) and (f).

Fig. 5 is the result for stationary Gaussian noises. The additive noise in Fig. 4(c) is used to generate simulated trials, shown in Fig. 5(a) with a signal-to-noise ratio (SNR) of -5 dB. Fig. 5(b) shows that the power spectrum of additive noises heavily overlap that of the SOI. The FLN $(N_0=32,\,N_1=32,\,\mu=0.05,\,\beta=0.02),$ the FLN-TOS $(N_0=32,\,N_1=32,\,\mu=0.05,\,\beta=0.02,\,m_1=30,\,m_2=5,\,\lambda=0.9),$ and 32-taps ALE with the learning rate $\mu=0.05$ are used in this simulation. The result shows that the filtered output of the ALE presents heavy distortion, and both the FLN and the FLN-TOS can effectively enhance the SOI in this case. The FLN provides about 5-dB reduction for Gaussian noises, whereas the FLN-TOS can provide about 20-dB reduction.

Fig. 6 is the result for nonstationary Gaussian noises. The variation of SNR and noise types of additive nonstationary Gaussian noises is shown in Fig. 6(a). Here, *NOISE1* and *NOISE2* denote Gaussian noises in Fig. 4(c) and (e), respectively. The parameters of the FLN, FLN-TOS, and ALE are set the same as above. Result shows that the

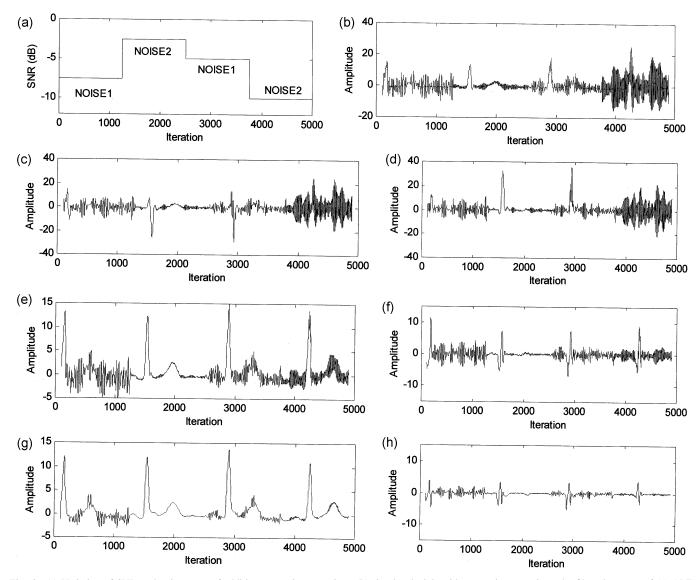


Fig. 6. (a) Variation of SNR and noise types of additive nonstationary noises; (b) simulated trials with nonstationary noises; the filtered outputs of (c) ALE, (e) FLN, and (g) FLN-TOS, and their error curves of (d) ALE, (f) FLN, and (h) FLN-TOS.

efficiency of the FLN for signal enhancement is sensitive to the variation of the SNR of additive noises. For the FLN-TOS, the influence of variation of additive noises on the performance for signal enhancement is negligible.

From above simulations, the FLN-TOS provides a good performance for signal enhancement under stationary and nonstationary Gaussian noises. To investigate the stability of performance of the FLN-TOS, the simulations of the FLN-TOS under different learning rates and noise levels are carried out. Fig. 7 is the comparison of performance of each method with different learning rates. As we expected, the FLN-TOS provides the best performance and its performance is more insensitive to the choice of learning rates. The performances of both the ALN and the FLN are easily influenced by selecting learning rates. In particular, when the learning rate is large, the influence of additional noises on the performance increases. Although smaller learning rates for the FLN can provide better performance, they also cause slow convergence for learning. This may cause the distortion of estimate of the SOI. Therefore, the performance of the FLN becomes poor when the learning rate is less than 0.05.

Fig. 8 is the comparison of performance of each method under different noise levels. The SNR of simulated trials are set from -2.5~dB

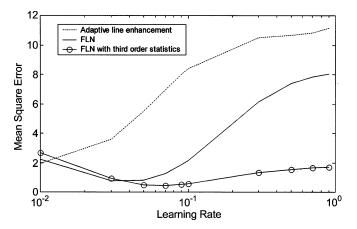


Fig. 7. Comparison of performances of each method with different learning rates.

to -15 dB. The result shows that the performance of ALE rapidly becomes very poor when the SNR becomes poor. The FLN-TOS provides

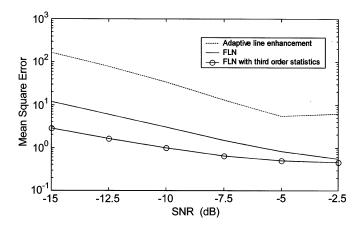


Fig. 8. Comparison of performances of each method under different noise levels.

better performance under different noise levels. Although its performance is also influenced by the variation of the SNR, it is more insensitive than that of the FLN and ALE.

VI. CONCLUSION

The learning algorithm of the FLN with higher order statistics was introduced in this study. By employing higher order statistics techniques to suppress additive Gaussian noises before adapting weights, it can provide a cleaner desired signal for learning. Simulated results show that the FLN-TOS can effectively enhance nonlinear under stationary or nonstationary Gaussian noises. The performance of the FLN-TOS for signal enhancement is less sensitive to its learning rates. An FLN can effectively track nonlinear signals, but its efficiency for suppressing noises is limited. The performance of the FLN-TOS over the FLN and the ALE has demonstrated under different noise levels. Therefore, the FLN-TOS is an effective approach for signal enhancement under additive Gaussian noisy environment.

REFERENCES

- B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, Jr., and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692–1716, 1975.
- [2] J. R. Zeidler, "Performance analysis of LMS adaptive prediction filters," *Proc. IEEE*, vol. 78, pp. 1781–1806, 1990.
- [3] S. Haykin, Adaptive Filter Theory, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [4] B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: perceptron, madaline, and backpropagation," *Proc. IEEE*, vol. 78, no. 9, pp. 1415–1441.
- [5] F. Rosenblatt, Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanism. Washington, DC: Spartan Books, 1961.
- [6] R. P. Lippmann and P. Beckman, "Adaptive neural net preprocessing for signal detection in non-Gaussian noise," Adv. Neural Inf. Proces. Syst., vol. 1, pp. 124–132, 1989.
- [7] P. P. Gandhi and V. Ramamurti, "Neural networks for signal detection in non-Gaussian noise," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2846–2851, Nov. 1997.
- [8] Y. H. Pao, Adaptive Pattern Recognition and Neural Networks. Reading, MA: Addison-Wesley, 1989.
- [9] N. E. Cotter, "The Stone–Weierstrass theorem and its application to neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 4, pp. 290–295, 1990
- [10] B. Igelnik and Y.-H. Pao, "Stochastic choice of basis functions in adaptive function approximation and the functional-link net," *IEEE Trans. Neural Netw.*, vol. 6, no. 6, pp. 1320–1329, 1995.

- [11] B. M. Sadler, G. B. Giannakis, and K.-S. Li, "Estimation and detection in non-Gaussian noise using higher order statistics," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2729–2741, Oct. 1994.
- [12] H. M. Ibrahim, R. R. Gharieb, and M. M. Hassan, "A higher order statistics-based adaptive algorithm for line enhancement," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 527–532, Feb. 1999.
- [13] D. C. Shin and C. L. Nikias, "Adaptive interference canceler for narrowband and wideband interferences using higher order statistics," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2715–2728, Oct. 1994
- [14] C. L. Nikias and A. P. Petropulu, Higher Order Spectral Analysis: A Nonlinear Signal Processing Framework. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [15] M. Brown and C. Harris, Neurofuzzy Adaptive Modelling and Control. Englewood Cliffs, NJ: Prentice-Hall, 1994.

Comments on "CuBICA: Independent Component Analysis by Simultaneous third- and Fourth-Order Cumulant Diagonalization"

Eric Moreau, Member, IEEE

Abstract—Source separation based on joint diagonalization of cumulant matrices or tensors has led now to classical algorithms. In a previous paper by Blaschke and Wiskott, an algorithm based on the above principle is proposed. It combines third- and fourth-order cumulants and it is claimed to possess interesting properties. This comment has two objectives. We point out imprecisions of statement and we give additional references of closely related works, some of them presenting identical issues.

Index Terms—Blind source separation, contrast function, higher order statistics, independent component analysis, matrices joint diagonalization.

I. INTRODUCTION

Among different approaches of the source separation problem, the algebraical ones based on some diagonalization procedures have lead to algorithms that became very well known because of their relative simplicity and efficiency. Initially, the separation procedure is decomposed into two stages. The first one realizes a so-called whitening of the observation signals by diagonalizing for example their covariance matrix (correlation matrix at zero lag). The main goal of this first stage is to constraint the matrix which one seeks to be a unitary one. The goal of the second stage is then to restore a final unitary matrix to estimate the source signals. This final stage has led to numerous interesting solutions, see, e.g., [3], [5], [2], [7], [6], [10], which are based on Jacobi-like procedures for diagonalization of some cumulant tensors.

In practice the above diagonalization procedures are considered through the optimization of a quadratic criterion. More importantly, the above (joint) diagonalization criteria are shown to be equivalent to some contrast functions. In the field of source separation, contrast functions are important in the sense that they constitute the basis of identifiability conditions for effective separation and yield natural separation criteria, see, e.g., [5].

Manuscript received October 21, 2005; revised January 17, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chong-Yug Chi.

The author is with the Telecommunications Department of ISITV, Université du Sud Toulon Var, 83162 La Valette du Var Cedex, France (e-mail: moreau@univ-tln.fr).

Digital Object Identifier 10.1109/TSP.2006.881175