

Practical Method for Blind Inversion of Wiener Systems

Kun Zhang and Lai-Wan Chan
 Department of Computer Science and Engineering
 The Chinese University of Hong Kong
 Shatin, Hong Kong
 E-mail: {kzhang, lwchan}@cse.cuhk.edu.hk

Abstract—In this paper, firstly we show that the problem of blind inversion of Wiener systems is a special case of blind separation of post-nonlinear instantaneous mixtures approximately, and derive the learning rule for the former problem using this relationship. Secondly, we review the Gaussianization method for blind inversion of Wiener systems. Based on the fact that the convolutive mixture is close to Gaussian, this method roughly approximates the convolutive mixture by a Gaussian variable and constructs the inverse nonlinearity easily. Thirdly, in order to improve the performance, the Cornish-Fisher expansion is exploited to model the latent convolutive mixture, and then the extended Gaussianization method is developed. We show that the performance of our method is insensitive to the nonlinearity in the Wiener system. Experimental results are presented to illustrate the validity and efficiency of our method.

I. INTRODUCTION

A Wiener system is a linear filter subsystem followed by a memoryless nonlinear distortion (see Figure 1). This system can model many actual systems in nature. It has been used in many areas, such as biology, physiology and phycology. For instance, it is used for describing neural processing [12], investigation of the activity of individual primary neurons in response to prolonged velocity stimuli [13], modelling the dynamic relation between muscle length and tension [9], and estimation of the nonlinear nonstationary model of infant EEG seizures [5].

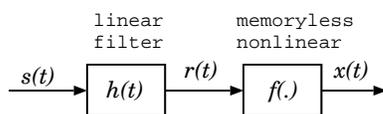


Fig. 1. Wiener system: a linear filter subsystem, $h(t)$, is followed by a memoryless nonlinearity, $f(\cdot)$. For simplicity, we have neglected the additive noise in the figure.

Traditionally, identification of the Wiener system aims to estimate both the characteristic of the nonlinear subsystem and the impulse response of the linear part from the input and output signals. As for the input signal, usually it is assumed to be a Gaussian white random process. The main difficulty in the identification is caused by the fact that the inner system signal, i.e. the convolutive mixture, is not measured. A popular method for identifying this system is the iterative method [10]. In this method, neither the linear subsystem nor the nonlin-

earity needs to be inverted. Therefore the nonlinearity is not restricted to be monotonic.

However, in some situations the input signal of the Wiener system is unknown and then the system can not be identified using traditional methods. In this paper we consider the problem—“blind inversion” of Wiener systems [4]. Under the assumption that the input $s(t)$ is an independent and identically distributed (i.i.d.) process with a non-Gaussian distribution, our goal is to identify the Wiener system, as well as to recover the latent input $s(t)$ from only the observation $x(t)$. In order to recover $s(t)$ from $x(t)$, this system should be invertible. And hence the function f is assumed to be one-to-one. Under these assumptions, the Wiener system can be inverted by the Hammerstein structure, which is a memoryless nonlinear stage followed by a linear filter (see Figure 2).

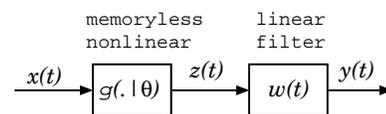


Fig. 2. Hammerstein system: memoryless nonlinearity, $g(\cdot|\theta)$, is followed by a linear filter subsystem, $w(t)$, where θ denotes the parameter vector. For simplicity, we have neglected the additive noise in the figure.

Some researchers have developed algorithms for blind inversion of the Wiener system with different models for the inverse nonlinearity, i.e. the function g in Figure 2 [4], [16], [2]. Without any prior information on the nonlinearity f , usually a multi-layer perceptron (MLP), or polynomial is used to model the inverse nonlinearity. The number of parameters needed to represent the inversion system is quite large. This causes high computation and makes the optimization slow and prone to local optima. And if one chooses a parametric family that is not appropriate, there is a danger of resulting in incorrect conclusions.

In this paper, we exploit the fact that the sum of independent variables tends to be Gaussian, which helps to reduce the freedom of the nonlinearity in the inversion system greatly. This paper is organized as follows. In Section II, we consider the Wiener system as a special case of the post-nonlinear (PNL) mixture model, and develop the learning rule for blind inversion of Wiener systems. In Section III, the Gaussianization method is reviewed, and most importantly, the extended

Gaussianization method is proposed. This method can be considered as a specific application of the general learning rule, with Gaussianization as the first step and the Cornish-Fisher expansion modelling the remaining nonlinearity in the inversion system. Section IV illustrates the performance of the proposed method with experiments. Section V discusses the main results and concludes the paper.

II. FROM BLIND SEPARATION OF PNL MIXTURES TO BLIND INVERSION OF WIENER SYSTEMS

Some researchers have investigated the relationship between blind deconvolution (BD) and blind source separation (BSS) [3], [7], [8]. Similarly, the relationship between blind separation of PNL mixtures and blind inversion of Wiener systems has been discussed in [16]. Now we use this relationship to derive the learning rule for blind inversion of Wiener systems.

A. Blind Separation of PNL Mixtures

Blind separation of PNL mixtures (see Figure 3) is a special case of nonlinear independent component analysis (ICA). By constraining the nonlinearity to be post nonlinear, this model can extract the independent sources up to permutation, scaling and mean indeterminacies under some weak assumptions [15], [17]. The observations $\mathbf{x} = (x_1, x_2, \dots, x_m)^T$ are generated from statistically independent sources $\mathbf{s} = (s_1, s_2, \dots, s_n)^T$ in the form:

$$x_i = f_i(r_i) = f_i\left(\sum_{j=1}^n a_{ij}s_j\right) \quad (1)$$

where s_i are statistically independent sources, f_i are unknown invertible nonlinear functions, a_{ij} denote entries of a regular $m \times n$ mixing matrix A , and r_i are latent linear mixtures.

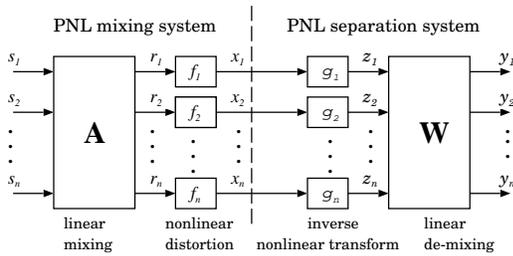


Fig. 3. PNL mixing system and PNL separation system (assuming the number of sources equals that of observations): The mixing system consists of a linear mixing stage with the mixing matrix A and a nonlinear transform stage applying \mathbf{f} to linear mixtures. The separation system does an inversion operation, which is a nonlinear transform stage followed by a linear de-mixing stage.

Assuming the number of sources is equal to that of observations, Taleb and Jutten have given the general learning rule for blind separation of PNL mixtures based on the minimization of mutual information between y_i [15]. Let θ_i denote the parameter vector in the function g_i , and $I(\mathbf{y})$ denote the mutual information between the components of \mathbf{y} . In the

nonlinear stage, θ_i are tuned to minimize $I(\mathbf{y})$:

$$\begin{aligned} \Delta\theta_i &\propto -\frac{\partial I(\mathbf{y})}{\partial W} \\ &= E\left\{\frac{\partial \log |g'_i(x_i|\theta_i)|}{\partial \theta_i}\right\} + E\left\{\left(\sum_{k=1}^n \psi_{y_k}(y_k)w_{ki}\right)\frac{\partial g_i(x_i|\theta_i)}{\partial \theta_i}\right\} \end{aligned} \quad (2)$$

where ψ_{y_i} is the score function of the distribution of y_i :

$$\psi_{y_i}(u) = (\log p_{y_i}(u))' = \frac{p'_{y_i}(u)}{p_{y_i}(u)} \quad (3)$$

Since the separation system of PNL mixtures is cascade structured, in the linear stage W is tuned to minimize $I(\mathbf{y})$ with \mathbf{z} as input:

$$\Delta W \propto -\frac{\partial I(\mathbf{y})}{\partial W} = [W^T]^{-1} + E\{\psi_{\mathbf{y}}(\mathbf{y})\mathbf{z}^T\} \quad (4)$$

where $\psi_{\mathbf{y}} = (\psi_{y_1}, \psi_{y_2}, \dots, \psi_{y_n})^T$.

B. Learning Rule for Blind Inversion of Wiener Systems

In Figure 3, let $\mathbf{s}(t) = (s(t), s(t+1), \dots, s(t+n-1))^T$, $\mathbf{r}(t) = (r(t), r(t+1), \dots, r(t+n-1))^T$, $\mathbf{x}(t) = (x(t), x(t+1), \dots, x(t+n-1))^T$, A be a Toeplitz matrix containing h_i , the coefficients of the filter $h(t)$, as its rows, at different positions for each row:

$$A = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots \\ \dots & h_0 & h_{-1} & h_{-2} & \dots \\ \dots & h_1 & h_0 & h_{-1} & \dots \\ \dots & h_2 & h_1 & h_0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \quad (5)$$

and all of f_i be the same nonlinearity, denoted by f . As $n \rightarrow \infty$, we can neglect the inaccuracy near the top and bottom channels in Figure 3, and the PNL mixing system becomes the Wiener system (Figure 1). In the same way we construct the Toeplitz matrix W with w_i , the coefficients of the filter $w(t)$, as its rows, and constrain g_i to be the nonlinearity g . As $n \rightarrow \infty$, the separation system in Figure 3 becomes the Hammerstein system (Figure 2). In order to eliminate the indeterminacies in the inversion of Wiener systems, we assume the nonlinearity g is strictly monotonically increasing, and y is zero-mean.

Since the separation system of PNL mixtures is cascade structured, the linear stage aims at extracting components y_i from input z_i as independent as possible. Therefore the learning rule in the linear stage is the same as that for ordinary ICA. Similarly, the linear stage in the inversion system of the Wiener system performs the ordinary BD task. In [7], a circulant matrix structure is constructed for the de-mixing matrix W in the BSS problem, and the natural gradient method for BD (see [1]) is derived by exploiting the relationship between BD and BSS. Its batch-mode version is as follows (assuming $w(t)$

has equal length, L , for the causal and noncausal parts):

$$\begin{aligned} \Delta w_j &\propto w_j + E_t\{\psi_y(t)u^*(t-j)\} \\ u(t) &= \sum_{q=-L}^L w_q y(t+q) \\ y(t) &= \sum_{q=-L}^L w_q x(t-q) \end{aligned} \quad (6)$$

where $*$ denotes complex-conjugate.

Due to the specific structure of W , \mathbf{x} and \mathbf{y} as addressed above, and taking into account that all of θ_i are θ , we have

$$\begin{aligned} &E_t\left\{\left(\sum_{k=1}^n \psi_{y_k}(y_k)w_{ki}\right)\frac{\partial g_i(x_i|\theta_i)}{\partial \theta_i}\right\} \\ &= E_t\left\{\left(\sum_{q=-L}^L \psi_y(t+q) \cdot w_q\right)\frac{\partial g(x(t)|\theta)}{\partial \theta}\right\} \\ &= E_t\{\psi_y(t)\sum_{q=-L}^L w_q \frac{\partial g(x(t-q)|\theta)}{\partial \theta}\} \end{aligned}$$

Then Eq. 2 becomes

$$\begin{aligned} \Delta \theta &\propto E_t\left\{\frac{\partial \log |g'(x(t)|\theta)|}{\partial \theta}\right\} \\ &+ E_t\{\psi_y(t)\sum_{q=-L}^L w_q \frac{\partial g(x(t-q)|\theta)}{\partial \theta}\} \end{aligned} \quad (7)$$

which is the learning rule in the nonlinear stage for blind inversion of Wiener systems.

III. GAUSSIANIZATION AND EXTENDED GAUSSIANIZATION

The classical central limit theorem states that the sum of i.i.d variables is asymptotically normally distributed as the number of independent variables increases. In Figure 1, since $s(t)$ is a i.i.d. process and $r(t) = \sum_k h_k s(t-k)$, the variable associated with $r(t)$ is a weighted sum of i.i.d variables. Therefore in general the distribution of r will be closer to the Gaussian distribution than s .^a Hence we assume the distribution of r is close to be Gaussian. Of course there is no straight definition on the ‘‘closeness’’, and it depends on the filter $h(t)$ and the distribution of s . One can construct counterexamples against our assumption. For instance, s is extremely non-Gaussian and one coefficient of $h(t)$ dominates the others. However, in this paper we consider the general cases.

A. Gaussianization

Gaussianization transforms a random variable into a standard Gaussian random variable. In the Wiener system, we simply approximate the distribution of r with the Gaussian distribution, and then $z(t)$, as an estimate of $r(t)$, can be constructed directly by Gaussianization. $y(t)$ can be obtained by applying BD on $z(t)$. This idea has been independently

^aWe drop the time index t of a signal to represent the random variable associated with this signal.

proposed in [14], [18]. In the latter paper, this technique is used to perform blind separation of PNL mixtures.

Let $F_x(x)$ denote the cumulative distribution function (cdf) of the variable x , and $p_x(x)$ denote the probability density function (pdf) of x . $F_x(x) = \int_{-\infty}^x p_x(t)dt$. Assume $F_z(z)$, the cdf of variable z , is continuous and strictly increasing in the variation range of z so that its inverse mapping exists and is unique. Since $z = g(x)$ and g is assumed to be strictly increasing, we have

$$\begin{aligned} F_z(g(x)) &= F_x(x), \quad \text{therefore} \\ g(x) &= F_z^{-1}(F_x(x)) \end{aligned}$$

Denote by G the cdf of the standard Gaussian distribution, $G(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$. As the distribution of z is approximated by the Gaussian distribution, the nonlinear mapping g is constructed by

$$g = G^{-1} \circ F_x \quad (8)$$

When $z(t) = g(x(t))$ is obtained, we can extract $y(t)$ by applying any BD method on $z(t)$: $y(t) = w(t) * z(t)$. For instance, the natural gradient method (Eq. 6) is adopted in our experiments.

In [17], some practical issues for constructing the nonlinearity g were considered.

B. Extended Gaussianization

In general it is rough to approximate the distribution of r by the Gaussian distribution. However, the nonlinearity g constructed using the Gaussianization method compensates the nonlinear distortion caused by f significantly.

Now let the nonlinearity constructed by Gaussianization be denoted by $g^{(1)}$. In order to improve the performance, we introduce a parametric form, $g^{(2)}$, to model the remaining nonlinearity to give more freedom to the distribution of z . Since the convolutive mixture $r(t)$ is assumed to be close to Gaussian, the nonlinearity $g^{(2)}$ is much ‘‘milder’’ than g . Figure 4 shows this structure. In the figure v is a standard Gaussian variable. Taking $v(t)$ as input, the learning rule for the Hammerstein system ($g^{(2)}, w(t)$) can be derived by applying the general learning rule for blind inversion of Wiener systems. This idea has been proposed for blind separation of PNL mixtures in [17].

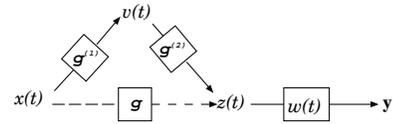


Fig. 4. The system for blind inversion of Wiener systems used in the extended Gaussianization method. We do not construct the nonlinearity g directly, but construct $g^{(1)}$ and $g^{(2)}$ instead. $g = g^{(2)} \circ g^{(1)}$

$g^{(1)}$ can be constructed according to Eq. 8. The Cornish-Fisher expansion helps to provide a parametric model for $g^{(2)}$.

1) *Cornish-Fisher Expansion*: It is possible to obtain useful approximate representation of a distribution with known cumulants in terms of a known distribution. The Edgeworth expansion and the Gram-Charlier expansion arise from choosing the Gaussian distribution as the known distribution.

Quantiles of a distribution are totally determined by the pdf of this distribution, hence one can approximate the quantiles of a distribution with those of the standard Gaussian distribution and known cumulants. Cornish and Fisher propounded a form of expansion in which the terms are polynomial functions of the appropriate standard Gaussian quantile, with functions of known cumulants as coefficients [6]. The four-term Cornish-Fisher (C-F) expansion for α -quantile is

$$x(u_\alpha) \cong u_\alpha + \frac{1}{6}(u_\alpha^2 - 1)\kappa_3 + \frac{1}{24}(u_\alpha^3 - 3u_\alpha)\kappa_4 - \frac{1}{36}(2u_\alpha^3 - 5u_\alpha)\kappa_3^2 \quad (9)$$

where κ_3 and κ_4 are the third-order and fourth-order cumulants (or skewness and kurtosis) of x respectively, u_α is the α -quantile of the standard Gaussian distribution, and $x(u_\alpha)$ is zero-mean and of unit variance.

It was shown in [11] that the C-F approximation is a competitive technique if the target distribution is relatively close to normal. And the most prominent use and motivation of this expansion is to model the distribution of the normalized sum of independent variables.

2) *Learning Rule for Extended Gaussianization*: As the distribution of z is assumed to be close to Gaussian, one can use the C-F expansion to approximate z in terms of v . According to Eq. 9, we have

$$\begin{aligned} z &= g^{(2)}(v|\boldsymbol{\theta}) = g^{(2)}(v|\kappa_3, \kappa_4) \\ &= v + \frac{1}{6}(v^2 - 1)\kappa_3 + \frac{1}{24}(v^3 - 3v)\kappa_4 - \frac{1}{36}(2v^3 - 5v)\kappa_3^2 \end{aligned} \quad (10)$$

Let D be the derivative of $g^{(2)}(v|\boldsymbol{\theta})$ with respect to v :

$$\begin{aligned} D &= g^{(2)'}(v|\boldsymbol{\theta}) \\ &= 1 + \frac{1}{3}\kappa_3 v + \frac{1}{8}(v^2 - 1)\kappa_4 - \frac{1}{36}\kappa_3^2(6v^2 - 5) \end{aligned}$$

According to Eq. 2, one can find the learning rule for κ_3 and κ_4 :

$$\begin{aligned} \Delta\kappa_3 &\propto E_t \left\{ \frac{-6\kappa_3 v^2(t) + 6v(t) + 5\kappa_3}{18D} + \frac{\psi_y(t)}{18} \right. \\ &\quad \left. \sum_{q=-L}^L w_q [-2\kappa_3 v^3(t-q) + 3v^2(t-q) + 5\kappa_3 v(t-q) - 3] \right\}, \\ \Delta\kappa_4 &\propto E_t \left\{ \frac{v^2(t) - 1}{8D} + \frac{\psi_y(t)}{24} \sum_{q=-L}^L w_q [v^3(t-q) - 3v(t-q)] \right\} \end{aligned} \quad (11)$$

This is the learning rule in the nonlinear stage. The learning rule in the linear stage is Eq. 6.

3) *Estimation of the Score Function*: As in blind separation of PNL mixtures, the performance of blind inversion of Wiener systems depends greatly on the estimation accuracy of the score function (Eq. 3). In our experiments the density of output y is modelled as a mixture of m Gaussians with means m_q , variances v_q , and mixing weights π_q ($q = 1, \dots, m$):

$$p_y(y) = \sum_{q=1}^m \pi_q \mathcal{G}(y|m_q, v_q)$$

Then the score function can be obtained analytically:

$$\psi_y(y) = \frac{p_y'(y)}{p_y(y)} = \sum_{q=1}^m p(q|y) \frac{y - m_q}{v_q}$$

where $p(q|y)$ denotes the posterior probability that y is generated by the q -th Gaussian. After each update of $(g^{(2)}, w(t))$, all parameters in the Gaussian mixture model are tuned to trace current output $y(t)$ based on maximum likelihood. The EM algorithm is used to maximize the likelihood. When all of the nonlinear stage, the linear stage and the EM stage converge, the learning process terminates. For details see [17].

4) *Discussion*: In the Gaussianization method, the composite of g and f transforms r to the standard Gaussian variable z . When r is given, whatever the nonlinearity f is, $z = g(f(r))$ does not change. Furthermore, for the extended Gaussianization method, v is not affected by the nonlinearity f . And consequently the performance of the extended Gaussianization method is insensitive to f in the noiseless situation.

The extended Gaussianization method is an application of the general learning rule, with Gaussianization as the first step and the C-F expansion for representing the inverse of remaining nonlinearity. The performance of this method depends on how well the C-F expansion can fit the actual distribution of the convolutive mixture. Hence in order to analyze the performance of our method, we first investigate the property of the C-F expansion.

To ensure that the C-F expansion corresponds to the quantile of a distribution, and that the mapping $x(u_\alpha)$ (Eq. 9) is one-to-one, Eq. 9 should be a strictly monotonic function of u_α . However, it is not always true for all possible values of (κ_3, κ_4) . We can identify the region in the (κ_3, κ_4) -plane guaranteeing the strict monotonicity of the mapping $x(u_\alpha)$. For values of (κ_3, κ_4) in this region, the derivative

$$\frac{dx(u_\alpha)}{du_\alpha} = 1 + \frac{1}{3}\kappa_3 u_\alpha + \frac{1}{8}(u_\alpha^2 - 1)\kappa_4 - \frac{1}{36}\kappa_3^2(6u_\alpha^2 - 5) \quad (12)$$

should be positive definite. Since Eq. 12 is a simple function of u_α with the highest order no more than 2, we can easily find the domain for (κ_3, κ_4) in which $\frac{dx(u_\alpha)}{du_\alpha}$ is always positive for any value of u_α , as shown in Figure 5 (a). Values in this region always result in super-Gaussian distributions of $x(u_\alpha)$. In practice the value of u_α is bounded, and hence the monotonicity constraint is relaxed. For instance, we can implement the monotonicity constraint for u_α between the 0.001 quantile (-3.09) and the 0.999 quantile (3.09) for the standard Gaussian distribution. After some straightforward but

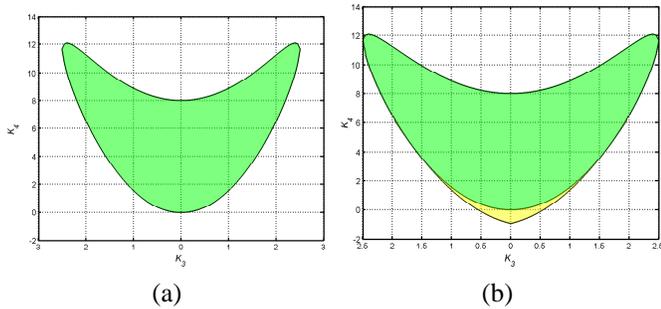


Fig. 5. The domain of (κ_3, κ_4) guaranteeing monotone of the 4-term C-F expansion (Eq. 9). (a) theoretical solution. (b) the domain guaranteeing monotone for $u_\alpha \in [-3.09, 3.09]$.

tedious derivations, the domain for κ_3 and κ_4 is obtained, as sketched in Figure 5 (b). The region in this figure shows us the feasibility of approximating z_i with the C-F expansion when z_i is close to Gaussian, especially when it is super-Gaussian. When (κ_3, κ_4) lies in the lower part of Figure 5 (b), the C-F expansion may become less and less reliable for $\alpha \rightarrow 0$ (or $\alpha \rightarrow 1$).

Due to the limitation of approximation of a distribution by the C-F expansion, the performance of the extended Gaussianization method varies for different characteristics of the distribution of the convolutive mixture. When the convolutive mixture $r(t)$ is close to Gaussian and the combination of its cumulants, (κ_3, κ_4) , lies in the domain shown in Figure 5 (a), the C-F expansion models $r(t)$ well, and hence a good performance is achieved. When $r(t)$ is close to Gaussian, (κ_3, κ_4) lies in the domain indicated by the lower part of Figure 5 (b), and the monotonicity of the C-F expansion is guaranteed for all samples of $v(t)$, $r(t)$ can still be modelled well by the C-F expansion, which also results in a good performance. When $r(t)$ is close to Gaussian, (κ_3, κ_4) lies in the domain indicated by the lower part of Figure 5 (b), and some samples of $v(t)$ are so large as to violate the monotonicity of the C-F expansion, we can drop these samples and the extended Gaussianization method still works. This is because mutual information as an independence measure is robust to outliers. But inevitably the performance decreases compared to the first two cases. When the distribution of $r(t)$ is far from the Gaussian distribution, (for instance, when its distribution is multimodal,) this method may fail.

IV. EXPERIMENTS

In order to investigate the performance of our method when r is not so close to Gaussian, the source signal $s(t)$ in our experiments is i.i.d. and follows the $\Gamma(3, 2)$ distribution. The filter $h(t)$ is a causal filter with coefficients $[1 \ 0.5 \ 0.5 \ 0.2]$. The distribution of the convolutive mixture $r(t)$ is considerably far from the Gaussian distribution, as seen from Figure 6 (b). The nonlinearity is very “hard”: $f(r) = r^3$. The inverse filter $w(t)$ is a causal filter with the length 60. In order to exhibit the performance quantitatively, the classical residual crosstalk $C(y, s) = 10 \log_{10} E[(y - s)^2]$ is used as the performance

index, in which by multiplying by a constant, y and s are made to be of unit variance and with positive correlation.

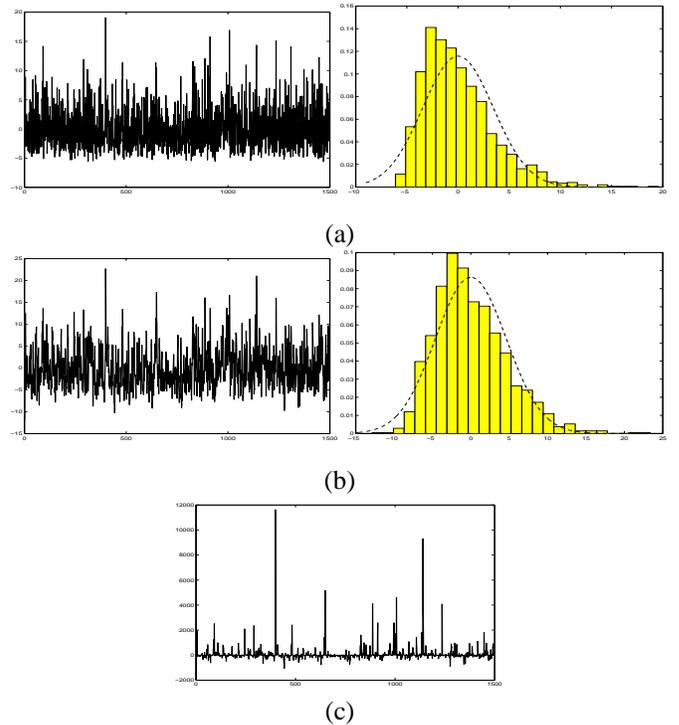


Fig. 6. (a) The source. The figure on the right shows its normalized histogram. For comparison, the Gaussian distribution with the same mean and variance is given by the dashed curve. (b) The latent convolutive mixture. (c) The observation after nonlinear distortion.

Figure 7 shows the experimental result with the Gaussianization method. The residual nonlinear distortion is still considerable due to the departure of the distribution of $r(t)$ from the Gaussian distribution. And accordingly the linear filter $w(t)$ does not invert the filter $h(t)$ well. The recovered signal is fairly different from the input source, which can also be seen from the performance index, which is only -4.8dB .

The experimental result using the extended Gaussianization method is shown in Figure 8. We can see that the nonlinearity f has been compensated by g . The filter $w(t)$ provides a good inversion of $h(t)$. And the input signal is successfully recovered. The performance index is improved to -16.1dB . The result shows the validity and efficiency of the extended Gaussianization method. In the nonlinear stage only two parameters are to be tuned so that this method learns fast.

Figure 9 shows the pdf associated with the C-F expansion and the output pdf modelled by the Gaussian mixture model. We can see the C-F expansion models the latent convolutive mixture well. And the Gaussian mixture model has learned the output density correctly.

V. CONCLUSION

In this paper we derived the general learning rule for blind inversion of Wiener systems. Moreover, we developed the extended Gaussianization method for blind inversion of Wiener systems under the assumption that the weighted sum of i.i.d

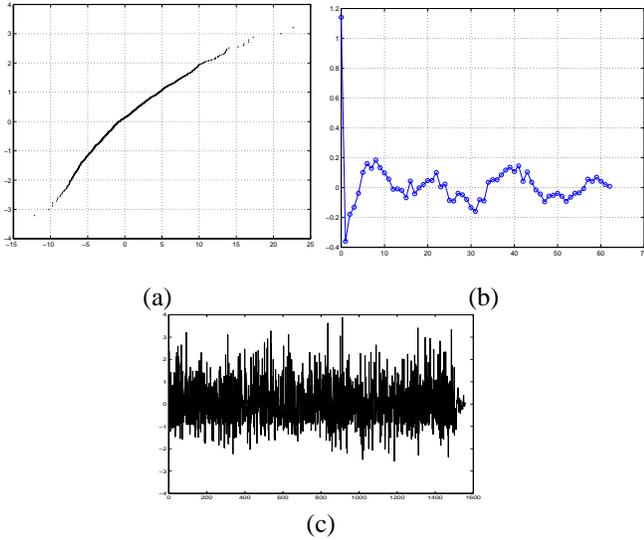


Fig. 7. Blind inversion result obtained by Gaussianization. (a) The residual distortion $g \circ f$ by the Gaussianization method. (b) The convolution of inverse filter w and the original filter h . (c) The recovered signal. Performance index: $C(s, y) = -4.8\text{dB}$.

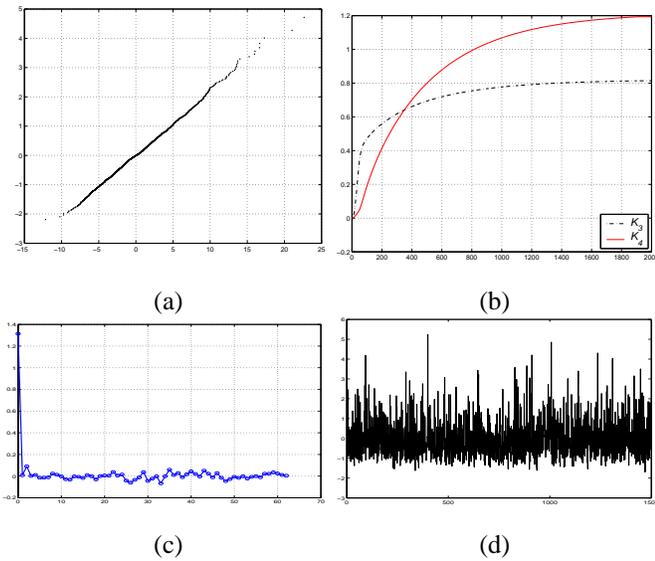


Fig. 8. Blind inversion result obtained by the extended Gaussianization method. (a) The residual distortion. (b) The parameters κ_3 and κ_4 versus iterations. (c) The convolution of w and h at convergence. (d) The recovered signal. Performance index: $C(s, y) = -16.1\text{dB}$.

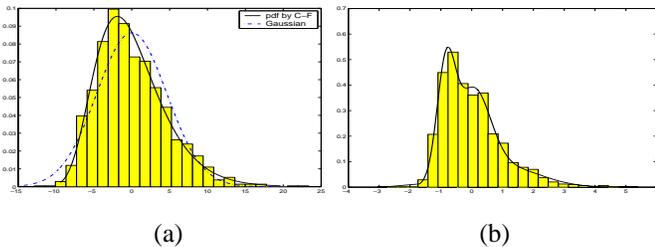


Fig. 9. (a) The pdf associated with the C-F expansion at convergence, together with the distribution histogram of the latent convolutive mixture. (b) The output pdf learned by the Gaussian mixture model, together with its histogram. Here 3 Gaussians are used.

variables are close to Gaussian. This assumption is reasonable in general cases according to the central limit theorem. Our method takes Gaussianization as the first step. After this step the inverse nonlinearity can be modelled by only two unknown parameters with the help of the Cornish-Fisher expansion. We also discussed the applicability of our method in different situations. The performance of this method is insensitive to the nonlinearity in the Wiener system. Experimental results illustrate the limitation of the Gaussianization method and the efficiency of the extended Gaussianization method.

ACKNOWLEDGMENT

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region.

REFERENCES

- [1] S.-I. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang. Multichannel blind deconvolution and equalization using the natural gradient. In *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, pages 101–104, Paris, France, 1997.
- [2] Massoud Babaie-Zadeh, Jordi Solé, and Christian Jutten. Blind inversion of Wiener system using a minimization-projection (MP) approach. In *ICA2003*, pages 681–686, Nara, Japan, April 2003.
- [3] Anthony J. Bell and Terrence J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.
- [4] Jordi Doléi Casals, Anis Taleb, and Christian Jutten. Parametric approach to blind deconvolution of nonlinear channels. In *ESANN'2000*, pages 21–26, April 2000.
- [5] P. Celka and P. Colditz. Nonlinear nonstationary Wiener model of infant EEG seizures. *IEEE Transaction on Biomedical Engineering*, 49(6):556–564, 2002.
- [6] E. A. Cornish and R. A. Fisher. Moments and cumulants in the specification of distributions. *Review of the International Statistical Institute*, 5:307–320, 1937.
- [7] S. C. Douglas and S. Haykin. On the relationship between blind deconvolution and blind source separation. In *Proc. 31st Asilomar Conf. Signals, Systems, and Computers*, pages 1591–1595, 1997.
- [8] Lambert R. H. Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures. *Ph.D. dissertation*, 1996.
- [9] I.W. Hunter. Experimental comparison of Wiener and Hammerstein cascade models of frog muscle fiber mechanics. *Biophys J*, 49:81a, 1986.
- [10] I.W. Hunter and M.J. Korenberg. Identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics*, 55:135–144, 1986.
- [11] Stefan R. Jaschke. The Cornish-Fisher expansion in the context of delta-gamma-normal approximations. *Journal of Risk*, 4(4):33–52, 2002.
- [12] V.Z. Marmarelis and K.I. Naka. White-noise analysis of a neuron chain: an application of the Wiener theory. *Science*, 175:1276–1278, 1972.
- [13] B.N. Segal and J.S. Outerbridge. Vestibular (semicircular canal) primary neurons in bullfrog: nonlinearity of individual and population response to rotation. *J Neurophys*, 47:545–562, 1982.
- [14] Jordi Sol-Casals, Massoud Babaie-Zadeh, Christian Jutten, and Dinh-Tuan Pham. Improving algorithm speed in pml mixture separation and wiener system inversion. In *Proc. ICA2003*, pages 639–644, 2003.
- [15] A. Taleb and C. Jutten. Source separation in post-nonlinear mixtures. *IEEE Trans. on Signal Processing*, 47(10):2807–2820, 1999a.
- [16] Anis Taleb, Jordi Solé, and Christian Jutten. Quasi-nonparametric blind inversion of Wiener systems. *IEEE Transaction on Signal Processing*, 49(5):917–924, 2001.
- [17] Kun Zhang and Lai-Wan Chan. Extended gaussianization method for blind separation of post-nonlinear mixtures. Paper in submission, 2003.
- [18] Andreas Ziehe, Motoaki Kawanabe, Stefan Harmeling, and Klaus-Robert Mueller. Blind separation of post-nonlinear mixtures using Gaussianizing transformations and temporal decorrelation. In *Proc. ICA2003*, pages 269–274, 2003.