Quasi-Nonparametric Blind Inversion of

Wiener Systems

Anisse Taleb (1,2), Jordi Solé (3), Christian Jutten (1,2)

(1) INPG-LIS, 46 Ave Flix Viallet, 38031 Grenbole Cedex, France

(2) GDR-PRC ISIS

(3) UNIVERSITAT DE VIC, c/ Sagrada família, 7 - 08500 Vic, Catalunya, Spain.

EDICS: 2SPTM 2NLIN 1MCHT 1NLHO

Correspondance author, email: taleb@lis-viallet.inpg.fr

Abstract

An efficient procedure for the blind inversion of a nonlinear Wiener system is proposed. We proved that the problem can be expressed as a problem of blind source separation in nonlinear mixtures, for which a solution has been recently proposed. Based on a quasi-nonparametric relative gradient descent, the proposed algorithm can perform efficiently even in the presence of hard distortions.

I. Introduction

When linear models fail, nonlinear models, because of their better approximation capabilities, appear to be powerfull tools for modeling practical situations. Examples of actual nonlinear systems include digital satellite and microwave channels which are composed of a linear filter followed by a memoryless nonlinear travelling wave tube amplifier [6], magnetic recording channel, etc. Identification of such systems is, therefore, of great theoretical and practical interest.

Many researches have been done for the identification and/or the inversion of such systems. Traditional nonlinear system identification methods have been based on higher-order input/output cross-correlations [2], or on the application of the Bussgang and Prices theorems [3], [10] for nonlinear systems with Gaussian inputs.

These methods assume the availability of the nonlinear system input. However, in a real world situation, one often does not have access to the system input: blind identification of the nonlinearity becomes then a necessary and a powerful tool. Although higher order statistics of the output signal has been used in the detection of nonlinearities [11], [12] or in the identification of cascade of linear system - memoryless polynomial nonlinearity - linear system [13], blind identification of nonlinear systems has remained an intractable problem, except for a very restricted class of inputs, mainly Gaussian.

It is well known that, unlike the case of linear systems, prior knowledge of the model is necessary for nonlinear system identification [14]. This paper is concerned by a particular class of nonlinear systems. These are composed by a linear time invariant subsystem (filter h) followed by a memoryless nonlinear distortion (function f) (Figure 1). This class of nonlinear systems, also known as Wiener systems, is not only another nice and mathematically attracting model, but also a model found in various areas, such as biology: study of the visual system [5], relation between the muscle length and tension [8], industry: description of a distillation plant [1], sociology and psychology. See also [9] and the references therein. Despite its interest, at our knowledge, no blind procedure exists for the inversion of such systems.

The paper organizes as follows. In section II, the model, notations and assumptions are detailed. The cost function, based on mutual information, is proposed in section III. Derivation of the cost function optimization algorithm is exposed in section IV. In V, practical issues for implementing the algorithm are adressed. Finally, section VI illustrates the good behavior of the proposed algorithm in a hard situation, and its efficiency by means of Monte-Carlo simulations.

II. MODEL AND ASSUMPTIONS

We suppose that the input of the system, $\mathcal{S} = \{s(t)\}$ is an unknown non-Gaussian independent and identically distributed (iid) process, and that both subsystems h, f are unknown and invertible. We are concerned by the restitution of $s(t)$ by only observing the system's output. This implies the blind design of an inverse structure (Figure 2), this structure will be composed of the same subsystems than the Wiener system, *i.e.* a linear filter and a memoryless nonlinearity, but in the reverse order, *i.e.* a Hammerstein system. The nonlinear part g is concerned by the compensation of the distortion f without access to its input, while the linear part w is

a deconvolution filter.

The following notation will be adopted through the paper. For each process $\mathcal{Z} = \{z(t)\}$, $\boldsymbol{z}$ denotes a vector of infinite dimension, whose t-th entry is $z(t)$. Following this notation, the unknown input-output transfert can be written as:

$$\boldsymbol{e} = \boldsymbol{f}(\boldsymbol{Hs}) \tag{1}$$

where:

$$\boldsymbol{H} = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & h(t+1) & h(t) & h(t-1) & \cdots \\ \cdots & h(t+2) & h(t+1) & h(t) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{2}$$

denotes a square Toeplitz matrix of infinite dimension and represents the action of the filter h on $s(t)$. This matrix is nonsingular provided that the filter h is invertible, *i.e.* h^{-1} exists and satisfies $h^{-1} * h = h * h^{-1} = \delta_0$, where δ_0 is the Dirac impulse at $t = 0$.

One can recognise in equation (1) the postnonlinear (pnl) model [15]. However, this model has been studied only in the finite dimentional case, in which it has been shown that, under mild conditions, the system is separable provided that the input $\boldsymbol{s}$ has independent components, and that the matrix $\boldsymbol{H}$ has at least two nonzero entries per row or per column.

Here, we conjecture that this will remain true for the infinite dimension case. The first separability condition is fullfiled since $\boldsymbol{s}$ has independent components due to the iid assumption. Moreover, due to the particular structure of the matrix $\boldsymbol{H}$, the second condition of separability will always hold except if h is proportional to a pure delay.

The output of the inversion structure can be written in the same way than (1):

$$y = Wx \qquad (3)$$

with $x(t) = g(e(t))$. Following [15], the inverse system g, w is estimated by minimizing the output mutual information. In fact, this criterion is always positive and vanishes iff y has independent components, *i.e.* $y(t)$ is an iid process.

III. COST FUNCTION

The mutual information of a random vector of dimension n is defined by:

$$I(z) = \sum_{i=1}^{n} H(z_i) - H(z_1, z_2, \ldots, z_n) \qquad (4)$$

Since we are interested in the mutual information of infinite dimension random vectors, a natural question is "how does this quantity grow with n?". This comes by using the notion of *entropy rates* of stochastic processes [4]. The entropy rate of the stochastic process $\mathcal{Z} = \{z(t)\}$ is defined as:

$$H(\mathcal{Z}) = \lim_{T \to \infty} \frac{1}{2T + 1} H(z(-T), \ldots, z(T)) \qquad (5)$$

if the limit exists. Theorem 4.2.1 of [4] states that this limit exists for a stationary stochastic process. We shall then define the mutual information rate of a stationary stochastic process by:

$$I(\mathcal{Z}) = \lim_{T \to \infty} \frac{1}{2T + 1} \left\{ \sum_{t=-T}^{T} H(z(t)) - H(z(-T), \ldots, z(T)) \right\} = H(z(\tau)) - H(\mathcal{Z}) \qquad (6)$$

Here τ is arbitrary due to the stationarity assumption. We shall notice that $I(\mathcal{Z})$ is always positive and vanishes iff $\mathcal{Z}$ is iid.

Now, since $\mathcal{S}$ is stationary, and since h and w are time-invariant filters, then $\mathcal{Y}$ is also stationary, and $I(\mathcal{Y})$ is defined by:

$$I(\mathcal{Y}) = H(y(\tau)) - H(\mathcal{Y}) \tag{7}$$

We need the following lemma, proof of which is given in Appendix A:

LEMMA 1: Let $\mathcal{X}$ be a stationary stochastic process, and let h be an invertible linear time invariant filter, then:

$$H(h * \mathcal{X}) = H(\mathcal{X}) + \frac{1}{2\pi} \int_0^{2\pi} \log|\sum_{t=-\infty}^{+\infty} h(t)e^{-jt\theta}|d\theta \tag{8}$$

□

Since $y(t) = h * x(t)$, one can write, applying lemma 1:

$$H(\mathcal{Y}) = H(\mathcal{X}) + \frac{1}{2\pi} \int_0^{2\pi} \log|\sum_{t=-\infty}^{+\infty} w(t)e^{-jt\theta}|d\theta \tag{9}$$

Moreover, since $x(t) = g(e(t))$ and by using the stationarity of $\mathcal{E} = e(t)$, one can write:

$$H(\mathcal{X}) = \lim_{T \to \infty} \frac{1}{2T+1} \left\{ H(e(-T), \dots, e(T)) + \sum_{t=-T}^{T} E\left[\log g'(e(t))\right] \right\}$$

$$= H(\mathcal{E}) + E\left[\log g'(e(\tau))\right] \tag{10}$$

Combining (9) and (10) in (7) leads finally to:

$$I(\mathcal{Y}) = H(y(\tau)) - \frac{1}{2\pi} \int_0^{2\pi} \log|\sum_{t=-\infty}^{+\infty} w(t)e^{-jt\theta}|d\theta - E\left[\log g'(e(\tau))\right] - H(\mathcal{E}) \tag{11}$$

IV. THEORETICAL DERIVATION OF THE INVERSION ALGORITHM

To derive the optimization algorithm we need the derivatives of $I(\mathcal{Y})$ (11) with respect to the linear part w and with respect to the nonlinear function g.

A. Linear subsystem

For the linear subsystem w, this is quite easy since the filter w is well parameterized by its coefficients. For the coefficient $w(t)$, corresponding to the t-th lag, we have:

$$\frac{\partial H(y(\tau))}{\partial w(t)} = -E[\frac{\partial y(\tau)}{\partial w(t)}\psi_{y(\tau)}(y(\tau))] = -E[x(\tau-t)\psi_{y(\tau)}(y(\tau))] \tag{12}$$

where $\psi_{y(\tau)}(u) = (\log p_{y(\tau)})'(u)$. Since, by stationarity, $\psi_{y(\tau)}$ is independent of τ, in the following it will be denoted simply ψ_y. The second term of interest is (see appendix B):

$$\frac{\partial}{\partial w(t)}\left\{\frac{1}{2\pi}\int_0^{2\pi}\log|\sum_{t'=-\infty}^{+\infty}w(t')e^{-jt'\theta}|d\theta\right\} = \frac{1}{2\pi}\int_0^{2\pi}\frac{e^{-jt\theta}}{\sum_{t=-\infty}^{+\infty}w(t)e^{-jt\theta}}d\theta \tag{13}$$

One recognises the $\{-t\}$-th coefficient of the inverse of the filter w, which we denote $\overline{w}(-t)$. The derivatives of the other terms with respect to the w coefficients are null, which leads, by combining (12) and (13), to:

$$\frac{\partial I(\mathcal{Y})}{\partial w(t)} = -E[x(\tau-t)\psi_y(y(\tau))] - \overline{w}(-t) \tag{14}$$

Equation (14) is the gradient of $I(\mathcal{Y})$ with respect to $w(t)$. Considering a small relative variation of w, expressed as the convolution of w by a "small" filter ϵ:

$$w \to w + \epsilon * w \tag{15}$$

the first order variation of $I(\mathcal{Y})$ writes as :

$$\Delta I(\mathcal{Y}) = -\left\{(\gamma_{y,\psi_y(y)} + \delta) * \epsilon\right\}(0) \tag{16}$$

where $\gamma_{y,\psi_y(y)}(t) = E[y(\tau-t)\psi_y(y(\tau))]$ and δ is the identity filter. One immediately notices that taking:

$$\epsilon = \mu\left\{\gamma_{y,\psi_y(y)} + \delta\right\}, \tag{17}$$

where μ is a small positive[1] real constant, insures a continuous decrease of $I(\mathcal{Y})$. It then provides the following gradient descent algorithm:

$$w \leftarrow w + \mu \left\{ \gamma_{y,\psi_y(y)} + \delta \right\} * w \tag{18}$$

which may be seen as a high-order correlation function of $y(t)$. It is interesting to notice that if $y(t)$ is Gaussian $\mathcal{N}(0,1)$ then $\psi_y(u) = -u$, and $\gamma_{y,\psi_y(y)}(t) = -E[y(\tau - t)y(\tau)]$, which is nothing but the correlation function, up to the sign, of $y(t)$. In this case, the algorithm reduces to a spectrum whitening.

B. Nonlinear subsystem

For this subsystem, we choose a nonparametric approach, implying no parametric-type restriction concerning its functional form. In consequence, and since the family of all possible characteristics is so wide, the only possible parametrisation of g is itself. This may seem to be confusing, but it will appear clearly in subsection V-D. Consider a small relative deviation of g, expressed as the composition of g by a "small" function ϵ:

$$g \rightarrow g + \epsilon \circ g \tag{19}$$

Then, the variations of the two terms $H(y(\tau))$ and $E\left[\log g'(e(\tau))\right]$ in (11) are considere. We have:

$$\Delta E[\log g'(e(\tau))] = E[\log(1 + \epsilon' \circ g(e(\tau)))] = E[\epsilon'(x(\tau))] \tag{20}$$

and,

$$\Delta H(y(\tau)) = -E[\psi_y(y(\tau))\{w * \epsilon(x)\}(\tau)] \tag{21}$$

---

[1]Small enough to insure the validity of the first order variation approximation.

It leads to a variation:

$$\Delta I(\mathcal{Y}) = -E[\psi_y(y(\tau))\{w * \epsilon(x)\}(\tau)] - E[\epsilon'(x(\tau))] \tag{22}$$

of $I(\mathcal{Y})$. Now let us write:

$$\epsilon'(x) = \int_{\mathbb{R}} \epsilon(v)\delta'(x - v)dv \tag{23}$$

$$\epsilon(x) = \int_{\mathbb{R}} \epsilon(v)\delta(x - v)dv \tag{24}$$

then:

$$\Delta I(\mathcal{Y}) = -\int_{\mathbb{R}} \underbrace{E[\psi_y(y(\tau))\{w * \delta(x - v)\}(\tau) + \delta'(x(\tau) - v)]}_{J(v)} \epsilon(v)dv \tag{25}$$

To make a gradient descent, we may take:

$$\epsilon(v) = \mu Q * J(v) \tag{26}$$

where μ is a small positive real constant, and Q is any function such that:

$$\int_{\mathbb{R}} J(v)Q * J(v)dv \geq 0, \tag{27}$$

to insure that $I(\mathcal{Y})$ decreases. Using the Parseval equality, the last condition becomes

$$\int_{\mathbb{R}} |\mathcal{J}(\nu)|^2 \Re\{\mathcal{Q}(\nu)\}d\nu \geq 0 \tag{28}$$

where $\mathcal{J}$ and $\mathcal{Q}$ are the Fourier transforms of J and Q respectively and $\Re$ denotes the real part.

It is thus sufficient to take $\Re\{\mathcal{Q}(\nu)\} > 0$ for insuring the *descent* condition. The algorithm writes then as:

$$g \leftarrow g + \mu \{Q * J\} \circ g \tag{29}$$

V. Practical issues

It is clear that (18) and (29) are unusable in practice. This section is concerned by adapting these algorithms to an actual situation. We consider then a finite discrete sample $\mathcal{E} = \{e(1), e(2), \ldots, e(T)\}$. We assume that we already have computed the output of the inversion system, *i.e.* $\mathcal{X} = \{x(1), x(2), \ldots, x(T)\}$ and $\mathcal{Y} = \{y(1), y(2), \ldots, y(T)\}$. The first question of interest is the estimation of the quantities involved in equations (18) and (29).

A. Estimation of ψ_y

Two approaches can be used to estimate ψ_y. The first approach consists in estimating the pdf of y, and by $\psi_y = (\log p_y)'$ one gets an estimate of ψ_y. A possible choice of an estimator of p_y consists in using nonparametric approaches, in particular methods based on kernels [7]. Kernel density estimators are easy to implement and have a very flexible form. However, they suffer from the difficulty of the choice of the kernel bandwidth B. In fact, this choice strongly depends on the criterion used to evaluate the performance of the estimation.

Formally, using normalized kernels[2], one estimates p_y by:

$$\hat{p}_y(u) = \frac{1}{BT} \sum_{t=1}^{T} K\left(\frac{u - y(t)}{B}\right) \tag{30}$$

and ψ_y by:

$$\hat{\psi}_y(u) = \frac{\sum_{t=1}^{T} K'\left(\frac{u - y(t)}{B}\right)}{B \sum_{t=1}^{T} K\left(\frac{u - y(t)}{B}\right)} \tag{31}$$

In our experiments we used the Gaussian kernel (K has a Gaussian shape), however many other kernel shapes can be good candidates (see [7] for more details). A *"quick and dirty"* method

---

[2] $\int_{\mathbb{R}} K(u) du = 1.$

for the choice of the bandwith consists in using the rule of thumb $h = 1.06\,\hat{\sigma}_y\,T^{-1/5}$ which is based on the minimum asymptotic mean integrated error criterion. Probably better choices could be used, but experimentally we noticed that the proposed estimator works fine.

A second approach for the estimation of ψ_y is based on the direct mean square minimisation of:

$$\xi(\hat{\psi}_y) = E[(\hat{\psi}_y(y) - \psi_y(y))^2] \tag{32}$$

which can be done without knowing the true function ψ_y. In fact, using:

$$E[\psi_y(y)\hat{\psi}_y(y)] = -E[\hat{\psi}'_y(y)]$$

equation (32) reduces to:

$$\xi(\hat{\psi}_y) = E[\hat{\psi}_y(y)^2 + 2\hat{\psi}'_y(y)] + E[\psi_y(y)^2] \tag{33}$$

in which the true function ψ_y appear as a constant term. $\hat{\psi}_y$ being nonlinear, it could then be modeled, for example, by a neural network [16] and trained using the backpropagation algorithm. This approach has not been used in this paper where we prefer a nonparametric approach.

B. Estimation of $\gamma_{y,\psi_y(y)}$

According to the previous subsection, it is assumed that one has an estimator of ψ_y. We are now interested in estimating:

$$\gamma_{y,\psi_y(y)}(t) = E[y(\tau - t)\psi_y(y(\tau))] \tag{34}$$

Assuming $y(t)$ is ergodic, one can get the estimate:

$$\hat{\gamma}_{y,\psi_y(y)}(t) = \frac{1}{T}\sum_{\tau=1}^{T} y(\tau - t)\hat{\psi}_y(y(\tau)) \tag{35}$$

where, by convention, $y(\tau) = 0$ for $\tau < 1$ and $\tau > T$. This estimate corresponds to the use of a rectangular window. Other estimates can be obtained by choosing different windows. Since, in theory, $\gamma_{y,\psi_y(y)}(0) = -1$, $\hat{\gamma}_{y,\psi_y(y)}(0)$ will be set to -1 without computing. Equation (35) is a convolution, it implementation is done by resort to the FFT which reduces dramatically the computation load.

C. Filter estimation

In pratical situations, the filter w, updated according to:

$$w \leftarrow w + \mu\{\hat{\gamma}_{y,\psi_y(y)} + \delta\} * w, \tag{36}$$

has a finite length (FIR). The result of the convolution of w with $\hat{\gamma}_{y,\psi_y(y)} + \delta$ should be truncated to fit the size of w. A smooth truncation, e.g. using of a Hamming window, is preferable to avoid overshooting.

D. Nonlinear subsystem estimation

We recall that no parametrisation of g is used. Then, one would ask the intriguing question "How would I compute the output of the nonlinear subsystem without having g?". In fact, applying the equation (29) to the t-th element of the sample $\mathcal{E}$, and using $x(t) = g(e(t))$, one gets:

$$x(t) \leftarrow x(t) + \mu\{Q * J\}(x(t)) \tag{37}$$

As (37) converges, for each input sample $e(t)$, we know that $x(t) = g(e(t))$. A simple nonlinear approximation using the samples $[x(t), e(t)]$, $t = 1, \dots, T$ could lead to an approximation of teh function g. However, this is not necessary and (37) has the advantage to be nonparametric, and efficient for any g.

E. Estimation of $Q * J$

This function is necessary to adapt the output of the nonlinear subsystem, and can be estimated by:

$$Q * J(v) = \frac{1}{T} \sum_{t=1}^{T} -Q'(v - x(t)) + \psi_y(y(t))\{w * Q(v - x)\}(t) \tag{38}$$

A possible choice of Q is:

$$Q(u) = \begin{cases} -u & \text{if } u \geq 0 \\ \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

which is very simple from a computational point of view. The real part of its Fourier transform is $\Re\mathcal{Q}(\omega) = 1/\omega^2$ and satisfies (28).

In order to reduce the computational load when T is large, equation (38) can be simplified. Instead of computing $Q * J$ at each sample $x(t)$, it is computed only at $N \ll T$ equally-spaced points in the range $[x_{min}, x_{max}]$ of the input signal $x(t)$. The values of $Q * J$ at the sample $x(t)$ are then obtained by linear interpolation. Equation (37) is in fact quite robust to sampling: this can be explained by the low-pass filtering effect of Q on J.

F. Indeterminacies

The output of the nonlinear subsystem $x(t), t = 1, \ldots, T$ should be centered and normalized. In fact, the inverse of the nonlinear distortion can be restored only up to a scale and a shift factor. For the linear subsystem, the output $y(t), t = 1, \ldots, T$ should also be normalized due

to scale indeterminacy on w. For example, the mean and variance of $x(t)$ are estimated by:

$$\hat{m}_x = \frac{1}{T} \sum_{t=1}^{T} x(t) \tag{40}$$

$$\hat{\sigma}_x = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (x(t) - \hat{m}_x)^2} \tag{41}$$

and $x(t)$ is normalized by:

$$x(t) \leftarrow \frac{x(t) - \hat{m}_x}{\hat{\sigma}_x} \tag{42}$$

One can use the same normalisation scheme for $y(t)$.

VI. Experimental results

A. Hard experiment

To test the previous algorithm, we simulate a hard situation. The iid input sequence $s(t)$, shown in figure 5, is generated by applying a cubic distortion to an iid Gaussian sequence. The filter h is FIR, with the coefficients:

$$h = [0.826, -0.165, 0.851, 0.163, 0.810]$$

Its frequency response is shown in figure 3. Figure 4 shows that h has two zeros outside the unit circle, which indicates that h is non-minimum phase.

The nonlinear distortion is a hard saturation $f(u) = \tanh(10u)$. The observed sequence is shown in figure 5.

The algorithm was provided with a sample of size $T = 1000$. The size of the impulse response of w was set to 51 with equal size for the causal and anti-causal parts. Estimation results, shown in figures 5,6 and 7, prove the good behavior of the proposed algorithm. The phase of

w (Figure 6) is composed of a linear part which corresponds to an arbitrary uncontrolled but constant delay, and of a nonlinear part which cancels the h phase.

B. Performance analysis

The performance of the proposed algorithm is evaluated by means of Monte-Carlo simulations. Since we are interested in restoring the original sequence, a good performance index consists in evaluating the error between the original signal $s(t)$ and the restored signal $y(t)$. The performance index writes then as:

$$\mathcal{P}_T = E\left[\frac{1}{T}\sum_{t=1}^{T}(y(t) - s(t))^2\right] \qquad (43)$$

in which, it is supposed that $y(t)$ has been properly delayed and rescaled.

A uniformally distributed signal in the range $[-\sqrt{3}, \sqrt{3}]$ is chosen for the simulations. Filter h is a simple low-pass filter with coefficients $[1, 0.5]$, and the nonlinear distortion is still $f(u) = \tanh(10u)$. Figure 8 shows the results of the simulations, where each point is the average of 50 experiments. As one can expect, the performance index decreases when the sample size T increases.

C. Real data

The previous procedure has been tested on real data. Figure 9 shows a seismic refraction profile recorded using a vessel and an Ocean Bottom Seismometer (OBS) over the Beaulieu plateau in France.

Each wave correspond to the signal recorded by the unique sensor, located on the sea bottom (OBS), in response to the shots which are produced at regular time intervals by the vessel which moves at constant speed.

In this experiment, the proposed procedure provides an estimate of the nonlinearity g which is far from a linear characteristic. The output of the nonlinear processing is shown in Figure 10.

Figure 11 compares a purely linear deconvolution and the proposed nonlinear method assuming the existence of a nonlinear memoryless distortion. We can notice that the proposed procedure provides an enhanced version of the purely linear deconvolution revealing hidden structures not emphasized by the linear deconvolution. However, these results are currently not well understood since seismic experts consider usually purely linear models. The only thing that can be said is that the linear model does not hold since our procedure revealed the existence of a nonlinearity.

VII. Final remarks and conclusion

In this paper a blind procedure for the inversion of a nonlinear Wiener system was proposed. Contrary to other blind identification procedures, the system input is not assumed to be Gaussian. Moreover, the nonlinear subsystem is unknown or not directly identifiable. The inversion procedure is based on a quasi-nonparametric relative gradient descent of the mutual information rate of the inversion system output. It is inspired from recent results on blind source separation in nonlinear mixtures.

One may notice that some quantities involved in the algorithm can be efficiently estimated by resort to FFT and interpolation, which reduces dramatically the computational cost. The estimation of g is done implicitely: only the values of $x(t) = g(e(t)), t = 1, \ldots, T$ are estimated. One can further use any regression algorithm based on these data to estimate g, e.g. neural networks, splines, etc.

The proposed procedure performs well, even in hard situations (hard nonlinearity and non minimum phase filter). Extension to complex channels is quite easy and is currently under investigation.

This approach could be promising in many application areas, as suggest the preliminary results on seismic data. These results show that the method emphasizes novel information that cannot be obtained by purely linear methods.

Appendix

I. Proof of lemma 1

We first show that, for two arbitrary stochastic processes $\mathcal{X}$ and $\mathcal{Y}$, we have:

$$I(\mathcal{X}, h * \mathcal{Y}) = I(\mathcal{X}, \mathcal{Y}) \tag{44}$$

In fact:

$$I(\mathcal{X}, h * \mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}/h * \mathcal{Y}) \tag{45}$$

and since h is supposed invertible, then the uncertaincy on $\mathcal{X}$ due to the knowledge of $h * \mathcal{Y}$ is equal to that of $\mathcal{X}$ due to the knowledge of $\mathcal{Y}$:

$$H(\mathcal{X}/h * \mathcal{Y}) = H(\mathcal{X}/\mathcal{Y}) \tag{46}$$

and leads obviously to (44).

Now, let $\mathcal{Z}$ be an arbitrary stationary stochastic process independent form $\mathcal{X}$. We have:

$$I(\mathcal{X} + \mathcal{Z}, \mathcal{Z}) = H(\mathcal{X} + \mathcal{Z}) + H(\mathcal{Z}) - H(\mathcal{X} + \mathcal{Z}, \mathcal{Z}) \tag{47}$$

However, using the independence of $\mathcal{Z}$ and $\mathcal{X}$, we have:

$$H(\mathcal{X} + \mathcal{Z}, \mathcal{Z}) = H(\mathcal{Z}) + H(\mathcal{X}) \tag{48}$$

Combining (47) and (48) we get:

$$I(\mathcal{X} + \mathcal{Z}, \mathcal{Z}) = H(\mathcal{X} + \mathcal{Z}) - H(\mathcal{X}) \tag{49}$$

And similarily for $h * \mathcal{Z}$ and $h * \mathcal{X}$, we have:

$$I(h * \mathcal{X} + h * \mathcal{Z}, h * \mathcal{Z}) = H(h * \mathcal{X} + h * \mathcal{Z}) - H(h * \mathcal{X}) \tag{50}$$

Using (44), the last equation becomes

$$I(h * \mathcal{X} + h * \mathcal{Z}, h * \mathcal{Z}) = I(\mathcal{X} + \mathcal{Z}, \mathcal{Z}) \tag{51}$$

which leads finaly to:

$$H(h * \mathcal{X} + h * \mathcal{Z}) - H(\mathcal{X} + \mathcal{Z}) = H(h * \mathcal{X}) - H(\mathcal{X}) \tag{52}$$

for any $\mathcal{Z}$. This expression shows that the difference between the output and input entropy is independent from the input distribution. We can then write:

$$H(h * \mathcal{X}) - H(\mathcal{X}) = H(h * \mathcal{G}) - H(\mathcal{G}) \tag{53}$$

where $\mathcal{G}$ is a white stationary Gaussian process. However, for a Gaussian process, we have from the Szegö-Kolmogorov-Krein theorem [17]:

$$H(h * \mathcal{G}) - H(\mathcal{G}) = \frac{1}{4\pi} \int_0^{2\pi} \log S(e^{j\theta}) d\theta \tag{54}$$

where $S(e^{j\theta})$ is the spectrum of $h * \mathcal{G}$. But since $\mathcal{G}$ is white, we have:

$$S(e^{j\theta}) = |\sum_{t=-\infty}^{+\infty} h(t)e^{-jt\theta}|^2 \tag{55}$$

from which, we deduce immediately that:

$$H(h * \mathcal{X}) - H(\mathcal{X}) = \frac{1}{2\pi} \int_0^{2\pi} \log|\sum_{t=-\infty}^{+\infty} h(t)e^{-jt\theta}|d\theta \tag{56}$$

II. Evaluation of (13)

Let us denote:

$$W(e^{j\theta}) = \sum_{t'=-\infty}^{+\infty} w(t')e^{-jt'\theta} \tag{57}$$

then,

$$\frac{1}{2\pi} \int_0^{2\pi} \log|\sum_{t'=-\infty}^{+\infty} w(t')e^{-jt'\theta}|d\theta = \frac{1}{4\pi} \int_0^{2\pi} \log W(e^{j\theta})W(e^{-j\theta})d\theta \tag{58}$$

and,

$$\frac{\partial}{\partial w(t)} \left\{ \frac{1}{2\pi} \int_0^{2\pi} \log|\sum_{t'=-\infty}^{+\infty} w(t')e^{-jt'\theta}|d\theta \right\} = \frac{1}{4\pi} \int_0^{2\pi} \frac{e^{-jt\theta}W(e^{-j\theta}) + e^{jt\theta}W(e^{j\theta})}{|W(e^{j\theta})|^2}d\theta$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \Re\left\{ \frac{e^{-jt\theta}}{W(e^{j\theta})} \right\} d\theta = \Re\frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-jt\theta}}{W(e^{j\theta})}d\theta \tag{59}$$

However, since the filter w is real, the imaginary part of the intergral is equal to zero leading

to the result (13).

References

[1]   R. Bars, I. Bèzi, B. Pilipàr, and B. Ojhelyi. Nonlinear and long range control of a distillation pilot plant. In *Identification and Syst. Parameter Estimation; Preprints 9th IFAC/IFORS Symp.*, pages 848–853, Budapest (Hungary), July 1990.

[2]   S. A. Billings and S. Y. Fakhouri. Identification of a class of nonlinear systems using correlation analysis. *Proc. IEEE*, 66:691–697, July 1978.

[3] E. D. Boer. Cross-correlation function of a bandpass nonlinear network. *Proc. IEEE*, 64:1443–1444, September 1976.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

[5] A. C. den Brinker. A comparison of results from parameter estimations of impulse responses of the transient visual system. *Biol. Cybern.*, 61:139–151, 1989.

[6] K. Feher. *Digital Communications–Satellite/Earth Station Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[7] W. Härdle. *Smoothing Techniques, with implementation in S*. Springer-Verlag, 1990.

[8] I. W. Hunter. Frog muscle fiber dynamic stiffness determined using nonlinear system identification techniques. *Biophys. J.*, 49:81a, 1985.

[9] I. W. Hunter and M. J. Korenberg. The identification of nonlinear biological systems: Wiener and Hamerstein cascade models. *Biol. Cybern.*, 55:135–144, 1985.

[10] G. Jacovitti, A. Neri, and R. Cusani. Methods for estimating the autocorrelation function of complex stationary processes. *IEEE Trans. ASSP.*, 35:1126–1138, August 1987.

[11] C. L. Nikias and Petropulu A. P. *Higher-Order Spectra Analysis - A Nonlinear Signal Processing Framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[12] C. L. Nikias and M. R. Raghuveer. Bispectrum estimation: A digital signal processing framework. *Proc. IEEE*, 75:869–890, July 1987.

[13] S. Prakriya and D. Hatzinakos. Blind identification of LTI-ZMNL-LTI nonlinear channel models. *Biol. Cybern.*, 55:135–144, 1985.

[14] M. Schetzen. Nonlinear system modeling based on the Wiener theory. *Proc. IEEE*, 69:1557–1573, December 1981.

[15] A. Taleb and C. Jutten. Source separation in postnonlinear mixtures. *IEEE trans. S.P.*- To appear.

[16] A. Taleb and C. Jutten. Entropy optimization, application to blind source separation. In *ICANN 97*, pages 529–534, Lausanne (Switzerland), October 1997.

[17] P. Whittle. *Some recent contributions to the theory of stationary processes*. Almquist and Wiksell, 1954. Appendix 2 in H. Wold, A study in the analysis of stationary time series.
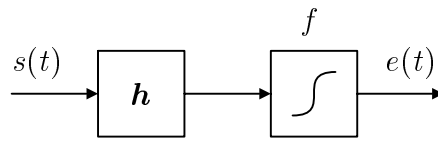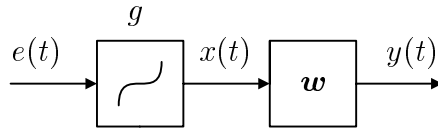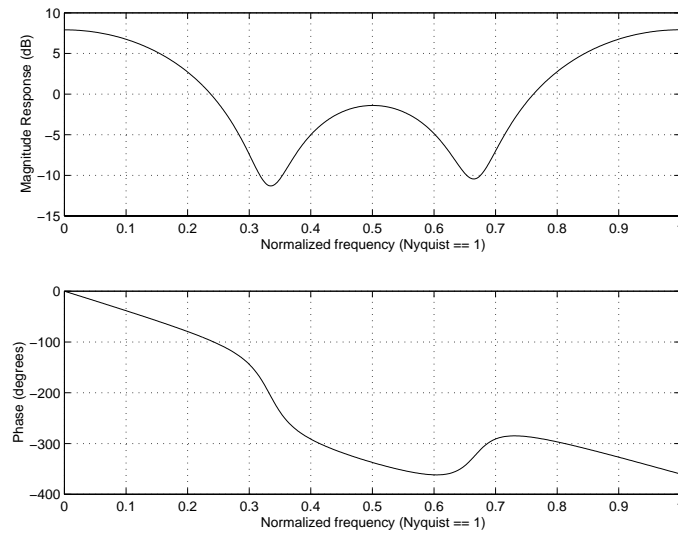
Fig. 1. Wiener system.

Fig. 2. Wiener system inversion structure - Hammerstein system.
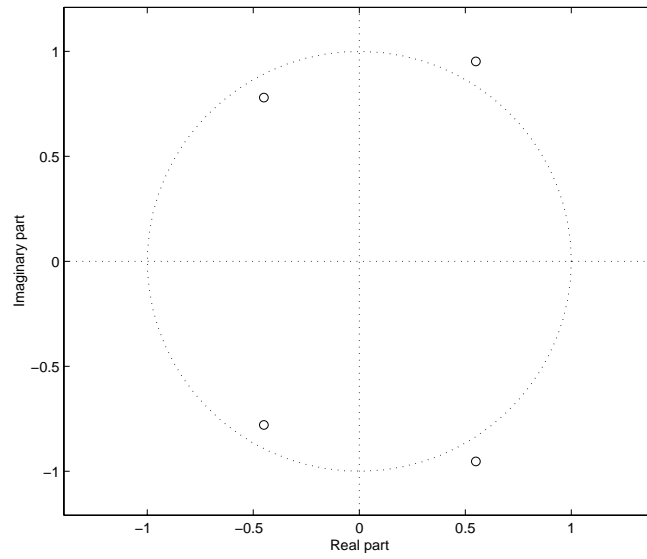

Fig. 3. h frequency domain response.
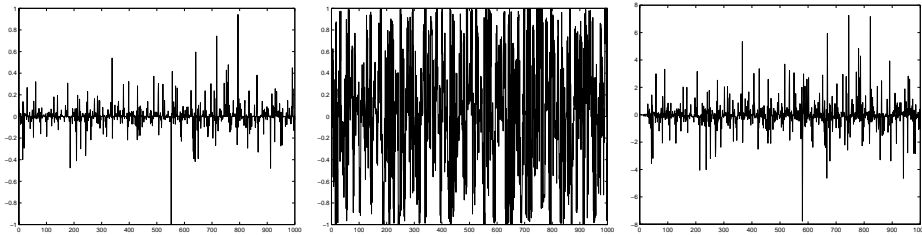
Fig. 4. Zeros of h.

Fig. 5. From left to right: Original input sequence $s(t)$, Observed sequence $e(t)$, Restored sequence $y(t)$.
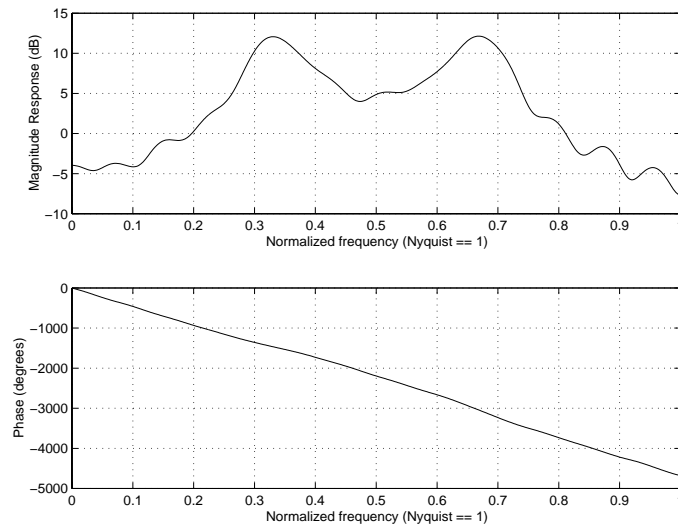

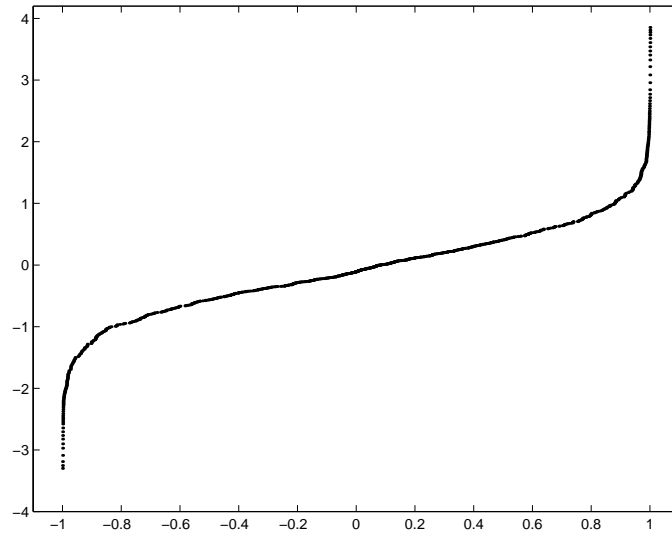Fig. 6. Estimated inverse of h: w frequency domain response.

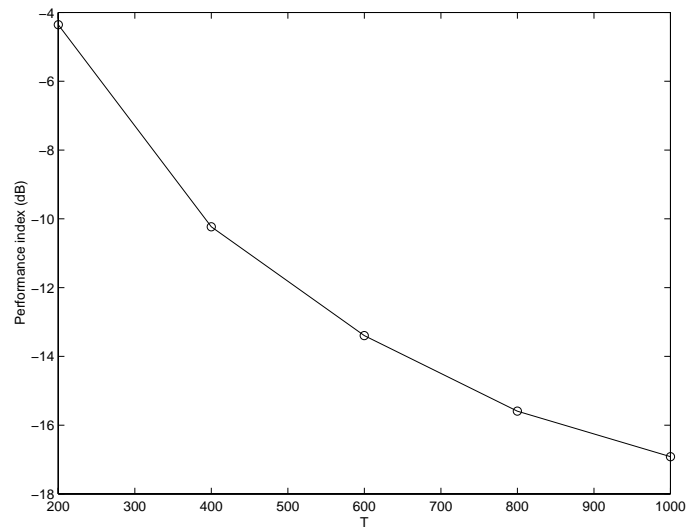Fig. 7. Estimated inverse of the nonlinear characteristic f: $x(t)$ vs. $e(t)$
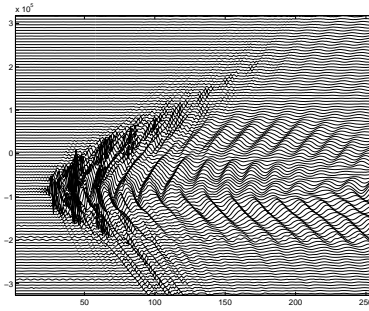

Fig. 8. Performance index versus sample size T.

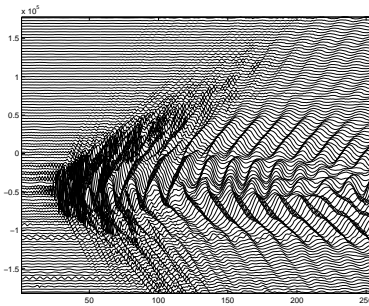Fig. 9.   Original seismic data

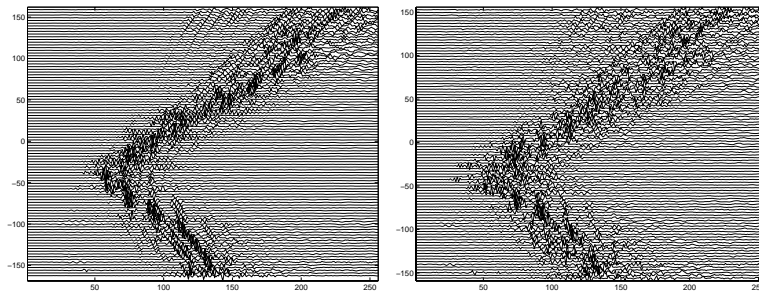Fig. 10.   Output of the nonlinear processing stage

Fig. 11.   Linear deconvolution vs. Nonlinear Wiener system deconvolution