# Blind Inversion of Wiener Systems

Anisse Taleb[1], Jordi Solé[2], and Christian Jutten[1,3]

[1] LIS-INPG, 46 Avenue Félix Viallet - 38031 Grenoble Cedex, France
[2] Universitat de Vic, c/ Sagrada família, 7 - 08500 Vic, Catalunya, Spain.
[3] ISTG-UJF, B.P 53 - 38041 Grenoble Cedex 9, France

**Abstract.** A system in which a linear dynamic part is followed by a non-linear memoryless distortion, a Wiener system, is blindly inverted. This kind of systems can be modelised as a postnonlinear mixture, and using some results about these mixtures, an efficient algorithm is proposed. Results in a hard situation are presented, and illustrate the efficiency of this algorithm.

## 1 Introduction and assumptions

In many areas of signal processing, nonlinear systems are present. Many research has been done in the identification and/or the inversion of such systems. These assume that the input of the distortion is available. One can get an estimate of the nonlinearity, or its inverse, and then the compensation of the distortion is straightforward.

However, in a real world situation, one often does not have access to the input. In this case, blind identification of the nonlinearity becomes the only way to solve such a problem. It is well known that, unlike the case of linear systems, prior knowledge of the model is necessary for nonlinear system identification [11].

Traditional nonlinear system identification methods have relied on higher-order cross-correlations of the input and the output [1]. The Bussgang and Prices theorems [2], [8] have been applied to identification of nonlinear models with real and complex Gaussian inputs. Though higher order statistics of the output signal have been used in the detection of nonlinearities [9], [10], blind identification of nonlinear systems has remained an intractable problem, except for the very restricted class of Gaussian inputs.
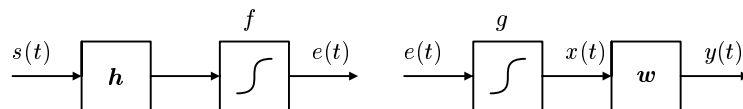


**Fig. 1.** Nonlinear Wiener system and its inversion structure

This paper is concerned by a particular class of nonlinear systems. These are composed by a linear subsystem (filter $h$) and a memoryless nonlinear function $f$

(Figure 1 Left). 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 [4], relation between the muscle length and tension [6], industry: description of a distillation plant, sociology and psychology, see also [7] and the references therein. Despite its interest, at our knowledge, no blind procedure exists for the identification of such systems.

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 output of the system. This implies that we will blindly design an inverse structure $g, w$ (Figure 1 Right). 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 linear deconvolution filter.

## 2    Design of the cost function

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 input-output transfert can be written as:

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

where:

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

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.

One can recognise in equation (1) the postnonlinear (pnl) model [12]. However, this model has been studied only in the finite dimentional case, in which it has been shown that, under mild conditions, the system was separable provided that the input $\boldsymbol{s}$ has independent components, and that matrix $\boldsymbol{H}$ has at least two nonzero entries per row or per column.

We conjecture that this will remain true in the infinite dimensional case. Here the first separability condition is fullfiled since $\boldsymbol{s}$ has independent components due to the iid assumption. Moreover, due to the particular structure of 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):

$$y = Wx \tag{3}$$

with $x(t) = g(e(t))$. Following [12], to invert such a system, the inverse system $g, w$ is estimated by minimizing the output mutual information.

Mutual information of a random vector of dimension $n$ is defined by:

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

Since we are interested in the mutual information of infinite dimension random vectors, a natural question to ask is "how does this quantity grow with n?". This comes by using the notion of *entropy rates* of stochastic processes [3]. The entropy rate stochastic process $\mathcal{Z} = \{z(t)\}$ is defined as:

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

when the limit exists. Theorem 4.2.1 of [3] states that this limit exists for a stationary stochastic process. We shall then define mutual information rate of a stationary stochastic process by:

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

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

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

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

From (3), one can write:

$$\begin{pmatrix} y(T) \\ \vdots \\ y(-T) \end{pmatrix} = \underbrace{\begin{pmatrix} w(0) & \cdots & w(-T) \\ \vdots & \ddots & \vdots \\ w(T) & \cdots & w(0) \end{pmatrix} \begin{pmatrix} x(T) + v(T) \\ \vdots \\ x(-T) + v(-T) \end{pmatrix}}_{\boldsymbol{W}_T(\boldsymbol{x}_T + \boldsymbol{v}_T)} \tag{8}$$

where $\boldsymbol{v}_T$ is a random vector which contains the remaining terms corresponding to the convolution truncation.

$$H(y(-T), \ldots, y(T)) = H(\boldsymbol{W}_T(\boldsymbol{x}_T + \boldsymbol{v}_T)) = \log|\det \boldsymbol{W}_T| + H(\boldsymbol{x}_T + \boldsymbol{v}_T) \tag{9}$$

The entropy rate of $\mathcal{Y}$ can then be expressed as:

$$H(\mathcal{Y}) = \lim_{T \to \infty} \frac{1}{2T+1} H(y(-T), \ldots, y(T)) = \lim_{T \to \infty} \frac{1}{2T+1} \log|\det \boldsymbol{W}_T| + H(\mathcal{X}) \tag{10}$$

since as $T \to \infty$, $\boldsymbol{x}(t) + \boldsymbol{v}(t) \xrightarrow{\text{m.s.}} \boldsymbol{x}(t)$. The first term of this last equation is:

$$\lim_{T \to \infty} \frac{1}{2T+1} \log|\det \boldsymbol{W}(t)| = \frac{1}{2\pi} \int_0^{2\pi} \log|\sum_{t=-\infty}^{+\infty} w(t)e^{-jt\theta}|d\theta \qquad (11)$$

In fact, as $T \to \infty$, the eigenvalues of $\boldsymbol{W}$ tend to the Fourier coefficents of $w$. Finally, by the stationarity of $\mathcal{E}$, one can write:

$$H(\mathcal{X}) = \lim_{T \to \infty} \frac{1}{2T+1} \left\{ H(e(-T), \dots, e(T)) + \sum_{t=-T}^{T} E\left[\log g'(e(t))\right] \right\}$$
$$= H(\mathcal{E}) + E\left[\log g'(e(\tau))\right] \qquad (12)$$

Combining (11) and (12) in (7) leads to:

$$I(\mathcal{Y}) = H(y(\tau)) - \frac{1}{2\pi} \int_0^{2\pi} \log|\sum_{t=-\infty}^{+\infty} w(t)e^{-jt\theta}|d\theta - E\left[\log g'(e(\tau))\right] - H(\mathcal{E})$$
$$(13)$$

# 3   Theoritical derivation of the inversion algorithm

To derive the optimization algorithm we need the derivatives of $I(\mathcal{Y})$ (13) with respect to the linear part $w$ and with repect to the nonlinear function $g$.

## 3.1   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:

$$\frac{\partial H(y(\tau))}{\partial w(t)} = -E[\frac{\partial y(\tau)}{\partial w(t)} \psi_{y(\tau)}(y(\tau))] = -E[x(\tau - t)\psi_{y(\tau)}(y(\tau))] \qquad (14)$$

where $\psi_{y(\tau)}(u) = (\log p_{y(\tau)})'(u)$. Since, by stationarity, it is independent from $\tau$ it will be denoted simply $\psi_y$. The second term of interest is:

$$\frac{\partial}{\partial w(t)} \frac{1}{2\pi} \int_0^{2\pi} \log|\sum_{t=-\infty}^{+\infty} w(t)e^{-jt\theta}|d\theta = \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-jt\theta}}{\sum_{t=-\infty}^{+\infty} w(t)e^{-jt\theta}} \qquad (15)$$

One recognises the $\{-t\}$-th coefficient of the inverse of the filter $w$, which we denote $\overline{w}(-t)$. The derivative of other terms with respect to $w$ coefficients are null, which leads by combining (14) and (15):

$$\frac{\partial I(\mathcal{Y})}{\partial w(t)} = -E[x(\tau - t)\psi_y(y(\tau))] - \overline{w}(-t) \qquad (16)$$

Equation (16) is the gradient of $I(\mathcal{Y})$ with respect to $w(t)$. Consider a small relative variation of $w$, expressed in terms of a convolution by a small filter $\epsilon$:

$$w \rightarrow w + \epsilon * w \qquad (17)$$

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

$$\Delta I(\mathcal{Y}) = - \left\{ \left( \gamma_{y,\psi_y(y)} + \delta \right) * \epsilon \right\}(0) \qquad (18)$$

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

$$\epsilon = \mu \left\{ \gamma_{y,\psi_y(y)} + \delta \right\}, \qquad (19)$$

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

$$w \leftarrow w + \mu \left\{ \gamma_{y,\psi_y(y)} + \delta \right\} * w \qquad (20)$$

### 3.2   Nonlinear subsystem

For this subsystem, we use a nonparametric approach. We make 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 the consequences are simple. The same technique as the linear subsystem is used here. In fact, consider a small relative deviation of $g$, expressed in terms of composition by a *"small"* function:

$$g \rightarrow g + \epsilon \circ g \qquad (21)$$

In this case, we have:

$$\Delta E[\log g'(e(\tau))] = E[\log(1 + \epsilon' \circ g(e(\tau)))] = E[\epsilon'(x(\tau))] \qquad (22)$$

and,

$$\Delta H(y(\tau)) = -E[\psi_y(y(\tau))\{w * \epsilon(x)\}(\tau)] \qquad (23)$$

which gives the variation of $I(\mathcal{Y})$:

$$\Delta I(\mathcal{Y}) = -E[\psi_y(y(\tau))\{w * \epsilon(x)\}(\tau)] - E[\epsilon'(x(\tau))] \qquad (24)$$

Now let us write:

$$\epsilon'(x) = \int_{\mathbb{R}} \epsilon(v)\delta'(x - v)dv \qquad (25)$$

$$\epsilon(x) = \int_{\mathbb{R}} \epsilon(v)\delta(x - v)dv \qquad (26)$$

---

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

then:

$$\Delta I(\mathcal{Y}) = - \int_{\mathbb{R}} \underbrace{E[\psi_y(y(\tau))\{w * \delta(x - v)\}(\tau) + \delta'(x(\tau) - v)]}_{J(v)} \epsilon(v)dv \qquad (27)$$

To make a gradient descent, we may take:

$$\epsilon(v) = \mu Q * J(v) \qquad (28)$$

where $Q$ is any function such that:

$$\int_{\mathbb{R}} J(v)Q * J(v)dv \geq 0 \qquad (29)$$

Using Parseval equality, this condition becomes

$$\int_{\mathbb{R}} |\mathcal{J}(\nu)|^2 \Re\{\mathcal{Q}(\nu)\}d\nu \geq 0 \qquad (30)$$

It suffice to take $\Re\{\mathcal{Q}(\nu)\} > 0$ to insure this condition. Based on the gradient descent, the algorithm writes then as:

$$g \leftarrow g + \mu \{Q * J\} \circ g \qquad (31)$$

## 4  Practical issues

It is clear that (20) and (31) are unusable in practice. This section is concerned by adapting these algorithms to an actual situation. We consider then a finite discrete sample $\mathcal{E} = \{e(1), e(2), \dots, e(T)\}$. The first question of interest is the estimation of the quantities involved in equations (20) and (31). We assume that we already have computed the output of the inversion system, *i.e.* $\mathcal{X} = \{x(1), x(2), \dots, x(T)\}$ and $\mathcal{Y} = \{y(1), y(2), \dots, y(T)\}$.

**Estimation of $\psi_y$:** Since we are concerned by nonparametric estimation, we will use a kernel density estimator [5]. This estimator is easy to implement and has a very flexible form, but suffers from the difficulty of the choice of the kernel bandwidths. Formally, we estimate $p_y$ by:

$$\hat{p}_y(u) = \frac{1}{hT} \sum_{t=1}^{T} K\left(\frac{u - y(t)}{h}\right) \qquad (32)$$

from which we get an estimate of $\psi_y$ by $\hat{\psi}_y(u) = \dfrac{\hat{p}'_y(u)}{\hat{p}_y(u)}$. Many kernel shapes can be good candidates, for our experiments we used the Gaussian kernel. A "*quick and dirty*" method for the choice of the bandwith consists in using the rule of thumb $h = 1.06\,\hat{\sigma}\,T^{-1/5}$. Better estimators may be found, and used, but experimentally we noticed that the proposed estimator works fine.

**Estimation of $\gamma_{y,\psi_y(y)}$:** Provided we have an estimator of $\psi_y$, we can compute $\hat{\psi}_y(y(t)), t = 1, \ldots, T$. Then:

$$\gamma_{y,\psi_y(y)}(t) = E[y(\tau - t)\psi_y(y(\tau))] \tag{33}$$

is estimated by:

$$\hat{\gamma}_{y,\psi_y(y)}(t) = \frac{1}{T}\sum_{\tau=1}^{T} y(\tau - t)\hat{\psi}_y(y(\tau)) \tag{34}$$

assuming ergodicity. Since $\gamma_{y,\psi_y(y)}(0) = -1$, $\hat{\gamma}_{y,\psi_y(y)}(0)$ may be set to $-1$ without computing it.

**Estimation of $Q * J$:** This function is necessary to adapt the output of the nonlinear subsystem, and can be estimated by:

$$Q * J(v) = \frac{1}{T}\sum_{t=1}^{T} -Q'(v - x(t)) + \psi_y(y(t))\{w * Q(v - x)\}(t) \tag{35}$$

**Nonlinear subsystem parametrisation and estimation:** No parametrisation of $g$ is used. One would ask the intriguing question *"How would I compute the output of the nonlinear subsystem without $g$?"*. In fact applying the equation (31) to the $t$-th element of the sample $\mathcal{E}$, and using $x(t) = g(e(t))$, one gets:

$$x(t) \leftarrow x(t) + \mu\{Q * J\}(x(t)) \tag{36}$$

This equation will then compute the output of $g$ without having a particular form of this function. A possible choice of $Q$ is:

$$Q(u) = \begin{cases} -u & \text{if } u \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{37}$$

which is very simple from a computational point of view.

**Filter parametrisation and estimation:** In pratical situations, the filter $w$ is of finite length (FIR). We also suppose that $w$ has equal length in its causal and anti-causal parts. Result of the convolution of $w$ with $\hat{\gamma}_{y,\psi_y(y)}$ should be truncated to fit the size of $w$. A smooth truncation, e.g. use of a Hamming window, is preferable to avoid overshooting.

**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 linear function. For the linear subsystem, the output $y(t), t = 1, \ldots, T$ should also be normalized.

# 5    Experimental results

To test the previous algorithm, we simulate a hard situation. The iid input sequence $s(t)$, shown in figure 3, 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 2. The nonlinear distortion is a hard
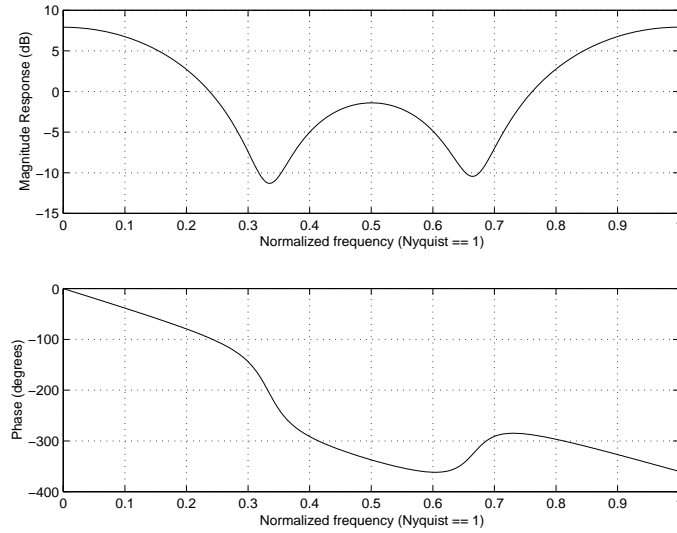


**Fig. 2.** $h$ frequency domain response.

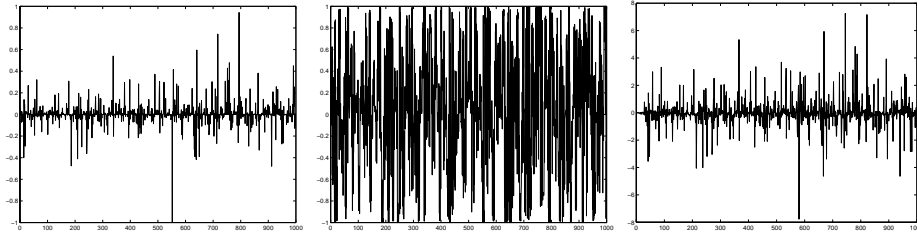saturation $f(u) = \tanh(10u)$. The observed sequence is shown in figure 3. The



**Fig. 3.** From left to right: Original input sequence $s(t)$, Observed sequence $e(t)$, Restored sequence $y(t)$.

algorithm was provided with a sample of size $T = 1000$. The size of the impulse

response of $w$ was set to 51. Estimation results, shown in figures 3,4,5, prove the good behavior of the proposed algorithm. The phase of filter $w$, Figure 4, is composed of a linear part which corresponds to an arbitrary uncontrolled but constant delay, and of a nonlinear part which compensates the $h$ phase.
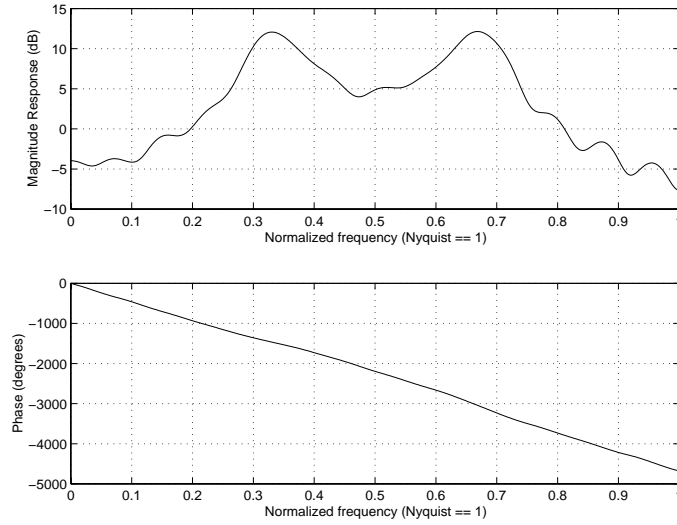


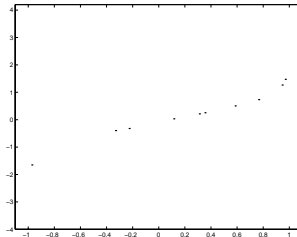**Fig. 4.** Estimated inverse of $h$: $w$ frequency domain response.



**Fig. 5.** Estimated inverse of the nonlinear characteristic $f$: $x(t)$ vs. $e(t)$

## 6    Final remarks and conclusion

In this paper a blind procedure for the inversion of a nonlinear Wiener system was proposed. This procedure is based on a relative gradient descent of the mutual information rate of the inversion system output.

One may notice that some quantities involved in the algorithm can be efficiently estimated by resort to the FFT 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 this data to estimate $g$, e.g. neural networks, splines, ect. The relation between the choice of $Q$ and the performances of the algorithm are not well understood and is currently under investigation.

The proposed procedure shows good performance on simulated data, and is now applied to real data. Extension to multichannel Wiener systems is currently under investigation.

# References

1. S. A. Billings and S. Y. Fakhouri. Identification of a class of nonlinear systems using correlation analysis. *Proc. IEEE*, 66:691–697, July 1978.
2. E. D. Boer. Cross-correlation function of a bandpass nonlinear network. *Proc. IEEE*, 64:1443–1444, September 1976.
3. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.
4. 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.
5. W. Härdle. *Smoothing Techniques, with implementation in S*. Springer-Verlag, 1990.
6. I. W. Hunter. Frog muscle fiber dynamic stiffness determined using nonlinear system identification techniques. *Biophys. J.*, 49:81a, 1985.
7. I. W. Hunter and M. J. Korenberg. The identification of nonlinear biological systems: Wiener and Hamerstein cascade models. *Biol. Cybern.*, 55:135–144, 1985.
8. 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.
9. C. L. Nikias and Petropulu A. P. *Higher-Order Spectra Analysis - A Nonlinear Signal Processing Framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
10. C. L. Nikias and M. R. Raghuveer. Bispectrum estimation: A digital signal processing framework. *Proc. IEEE*, 75:869–890, July 1987.
11. M. Schetzen. Nonlinear system modeling based on the Wiener theory. *Proc. IEEE*, 69:1557–1573, December 1981.
12. A. Taleb and C. Jutten. Source separation in postnonlinear mixtures. Jan 1998. Submitted to *IEEE trans. S.P.*- under revision.