MODELING AND COMPENSATING INPUT DISTORTION IN MIMO SYSTEMS VIA BLIND DECONVOLUTION

Jani Even [*] **Kenji Sugimoto** [*,1]

[*] *Graduate School of Information Science*
Nara Institute of Science and Technology
Keihanna Science City, Nara, 630-0192, JAPAN

Abstract: This paper proposes a method for modeling and compensating nonlinear input distortion in MIMO dynamical systems. This is conducted in three steps. First, the distorted signals are recovered from the observed output by using blind signal deconvolution based on statistical independence of the input signals. The linear part of the system is also identified in this first step. Secondly, the recovered signal, together with the input signal, gives a model of distortion as a nonlinear function. Finally the distortion is compensated by applying the inverse of the estimated function.

Keywords: Nonlinear distortion, Modeling, Independent component analysis

1. INTRODUCTION

In practical situations, a system to be controlled is more or less input-distorted, i.e., its input signals undergo some nonlinear distortion before they excite a linear dynamical system. This model is called Hammerstein system (Fig.1). Such distortion often arises due to saturation, dead-zone, or other imperfections of actuators even if the nominal system is linear. It is hence of great interest to model and compensate the distortion for achieving good control performance. In this paper, we study a method for identifying both the linear part and the nonlinear distortion part simultaneously for multi-input multi-output (MIMO) systems.

Traditional identification methods fail to provide a model in this case because they have no access to the distorted signals, which then pass through the (unknown) linear part. Consequently some complicated identification methods have been proposed. Most of the existing methods express the nonlinearity as a combination of nonlinear basis-functions and use the input output relationship to perform the identification, see (Pawlak, 1991; McKelvey and Hanner, 2003; Greblicki, 1989, among others).

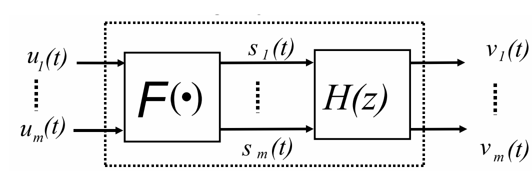

Fig. 1. Hammerstein system

This paper proposes yet another approach based on "blind signal deconvolution". Under the assumption that the distortion is memoryless (static) and component wise (no interaction among channels), the distorted signal is retrieved by deconvoluting the output signal. Then we provide a simple way to compensate the nonlinearity.

[1] Partially supported by the Japan Society for Promotion of Science, under the grant #18560436

In general, the goal of blind signal deconvolution is to retrieve the source signals applied to a linear dynamical system when only its output signals are observed. When retrieving the source, we also obtain the system parameter up to certain indeterminacies; see Section 2 for a detail. The research conducted on blind signal deconvolution during the last decades gave birth to many efficient methods; see review paper (Cardoso, 1998) and book (Cichocki and Amari, 2002) for details. Since the methods are mainly based on the independence of signals, these techniques are also referred to as ICA (Independent Component Analysis). ICA techniques seem promising in control engineering, too, although few papers have appeared so far (Sugimoto and Nitta, 2005; Sugimoto *et al.*, 2005; Even and Sugimoto, 2006).

Based on ICA, this paper gives an efficient method for identification of Hammerstein system. In fact, the simulation results below are quite satisfactory. In particular, the method is successful in ameliorating the performance of a feedforward controller. It is hence expected that the proposed method may help us to broaden the range of operation in various application.

The remainder of the paper is organized as follows. In Section 2, as preliminary, we briefly present blind signal deconvolution. In 3, the problem is stated and the proposed method is detailed. In 4, a numerical simulation illustrates the effectiveness of the proposed method.

Throughout the paper the signals are in discrete-time with time $t = 0, 1, 2, \cdots$. Matrices and vectors are written in boldface. $\mathcal{I}_m$ and $\mathcal{O}_{p\times m}$ respectively denote $m \times m$ identity matrix and $p \times m$ zero matrix.

2. PRELIMINARIES: BLIND DECONVOLUTION

The blind deconvolution problem is formulated as follows. $\mathbf{v}(t) = [v_1(t), \ldots, v_m(t)]^T$ are the output signals of a dynamical system $\mathbf{H}(z)$, called "mixer," excited by the "source" signals $\mathbf{s}(t) = [s_1(t), \ldots, s_m(t)]^T$. Both the mixer and the sources are assumed to be unknown. The goal is to recover the source signals as closely as possible.

A dynamical system $\mathbf{W}(z)$, called "demixer," is applied to the observed signals, see Fig. 2. Then the demixer is adapted so that its outputs $\mathbf{y}(t) = [y_1(t), \ldots, y_m(t)]^T$ estimate the inputs $\mathbf{s}(t)$. A common hypothesis used in blind signal deconvolution is that the signals $s_i(t)$ are

i) statistically independent,

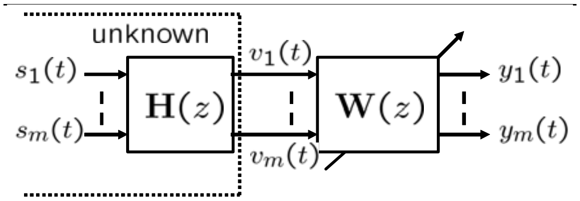

Fig. 2. Blind signal deconvolution

ii) identically independently distributed (i.i.d.),
iii) non Gaussian.

With these hypotheses, the goal can be achieved by adapting the demixer $\mathbf{W}(z)$ until the components of $\mathbf{y}$ are mutually statistically independent; for a detail, see (Cardoso, 1998) and references herein. Usually, the statistical independence is measured by a cost function based on their higher order statistics. The mutual information of $\mathbf{y}(t)$

$$\mathrm{MI}(\mathbf{y}) = -H(\mathbf{y}) + \sum_{i=1}^{m} H(y_i) \qquad (1)$$

is a commonly used cost function (see Appendix A).

Even under the above conditions the blind identification of the transfer matrix has indeterminacies: a permutation, a delay or a scaling of any components cannot be detected. This is illustrated by:

$$\mathbf{W}(z)\mathbf{H}(z) = \mathbf{P}\ \Lambda(z), \qquad (2)$$

where $\mathbf{P}$ is some permutation matrix and

$$\Lambda(z) = \begin{bmatrix} \lambda_1 z^{-\tau_1} & & \\ & \ddots & \\ & & \lambda_m z^{-\tau_m} \end{bmatrix}$$

for some real number λ_i and some integer τ_i. These indeterminacies, however, can be resolved with our identification setting, as will be shown in Section 3.4.

3. MAIN RESULTS

3.1 Problem formulation

Now consider a Hammerstein system composed of a nonlinear part and a linear part as described in Fig. 3. There are m memoryless nonlinear functions $f_i(\cdot)$ that distort the signals $\mathbf{u}(t) = [u_1(t), \ldots, u_m(t)]^T$. Let us assume that $f_i(\cdot)$ is smooth with $f_i(0) = 0$. The linear part is a dynamical system modeled by an $m \times m$ stable, minimal phase and biproper rational transfer matrix $\mathbf{H}(z)$. The signal $\mathbf{u}(t)$ is assumed to satisfy the assumptions i)-iii) in Section 2.

Our final goal is to compensate the nonlinear functions $f_i(\cdot)$. First the unknown signal $\mathbf{s}(t) =$

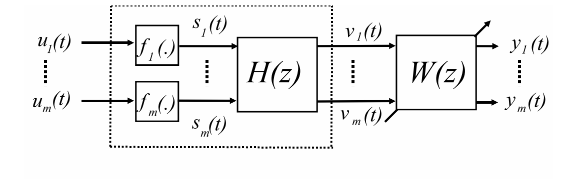

Fig. 3. Recovery of source $s(t)$ by blind signal deconvolution

$[s_1(t), \ldots, s_m(t)]^T$, where $s_i(t) = f_i(u_i(t))$, is estimated. This is done by means of a blind deconvolution technique outlined in Section 2. Then using the signal $\mathbf{u}(t)$ and the estimates of $\mathbf{s}(t)$, we compute approximates of the nonlinear functions $f_i(\cdot)$. Finally, we compensate their effect by applying the reverse functions. These are done in batch processes and are detailed below.

3.2 Estimation of linear part

The assumptions i)-iii) of Section 2 also hold for the input signals $\mathbf{s}(t) = [s_1(t), \ldots, s_m(t)]^T$ of $\mathbf{H}(z)$ because the nonlinear part of the system is non mixing and memoryless. We further assume that $f_i(u_i(t))$ has a non-Gaussian distribution without much loss of generality. These assumptions allow us to recover $\mathbf{s}(t)$ by means of an approximate blind deconvolution method, using the following "finite impulse response (FIR)" demixer $\mathbf{W}(z)$.

The demixer is represented by the state space model

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}_W \ \mathbf{x}(t) + \mathbf{B}_W \ \mathbf{v}(t) \\ \mathbf{y}(t) &= \mathbf{C}_W \ \mathbf{x}(t) + \mathbf{D}_W \ \mathbf{v}(t) \end{aligned} \tag{3}$$

with $\quad \mathbf{A}_W = \begin{bmatrix} \mathcal{O}_{m \times m(l+1)} \\ \mathcal{I}_{ml} \ \ \mathcal{O}_{ml \times m} \end{bmatrix} \mathbf{B}_W = \begin{bmatrix} \mathcal{I}_m \\ \mathcal{O}_{ml \times m} \end{bmatrix}$

$$\begin{bmatrix} \mathbf{D}_W & \mathbf{C}_W \end{bmatrix} = \begin{bmatrix} \mathbf{W}_0 & \mathbf{W}_1 & \ldots & \mathbf{W}_l \end{bmatrix}$$

$\mathbf{W}(z)$ is an FIR filter of sufficiently large length $l+1$, the matrices $\mathbf{A}_W$ and $\mathbf{B}_W$ are fixed whereas $\mathbf{C}_W$ and $\mathbf{D}_W$ are adapted with a batch algorithm based on the on-line method proposed in (Cichocki and Amari, 2002), which exploits the relative gradient of the mutual information MI$[\mathbf{y}(t)]$, as follows.

Let $\mathbf{C}_W(k)$ and $\mathbf{D}_W(k)$ denote the matrices at iteration k and $\mu(k)$ be the positive adaptation step used at iteration k. The adaptation law for $\mathbf{C}_W(k)$ and $\mathbf{D}_W(k)$ are

$$\begin{aligned} \mathbf{C}_W(k+1) &= \mathbf{C}_W(k) - \mu(k)\Delta\mathbf{C}_W(k) \\ \mathbf{D}_W(k+1) &= \mathbf{D}_W(k) - \mu(k)\Delta\mathbf{D}_W(k), \end{aligned}$$

with

$$\begin{aligned} \Delta\mathbf{C}_W(k) = & \left[\mathcal{I}_m - < \psi(\mathbf{y}(t))\mathbf{y}^T(t) >_t\right]\mathbf{C}_W(k) \\ & - < \psi(\mathbf{y}(t))\mathbf{x}^T(t) >_t \\ \Delta\mathbf{D}_W(k) = & \left[\mathcal{I}_m - < \psi(\mathbf{y}(t))\mathbf{y}^T(t) >_t\right]\mathbf{D}_W(k) \end{aligned}$$

where $< . >_t$ denotes the time average on the data block. $\psi(.) = [\psi_1(.) \ldots \psi_m(.)]^T$ is a vector containing the score functions associated with the input signal (see next paragraph). In iteration k, the whole block of the signal $\mathbf{y}(t)$ and the state $\mathbf{x}(t)$ has to be computed using Eq. (3) with $\mathbf{C}_W(k)$ and $\mathbf{D}_W(k)$ (hence this is a batch algorithm). At initialization $[\mathbf{D}_W(0)| \ \mathbf{C}_W(0)]$ is set to $[\mathcal{I}_m| \ \mathcal{O}_{m \times ml}]$. An important property of the above adaptation law is that once $\mathbf{W}(z)$ is initialized to a biproper filter, it remains biproper during adaptation (Cichocki and Amari, 2002). After convergence of the algorithm, $\mathbf{W}(z)$ is an approximate inverse of $\mathbf{H}(z)$.

The components of the signal $\mathbf{u}(t)$ are not observed and thus neither their probability distribution functions $P_{u_i}(u_i)$ nor their score functions $\psi_{u_i}(u_i) = -\partial \ln[P_{u_i}(u_i)]/\partial u_i$ are known. A simple approach is to use some prior knowledge in order to determine an approximation of the required score functions (Cardoso, 1998). Conversely, as in this paper, the score functions can be estimated during the adaptation so that no prior knowledge on the form of the nonlinear functions $f_i(\cdot)$ is necessary. These approximate score functions are used to compute the update matrices $\Delta\mathbf{D}_W(k)$ and $\Delta\mathbf{C}_W(k)$. During adaptation, $\mathbf{y}(t)$ converges to $\mathbf{u}(t)$ and thus $\widehat{\psi}_k(.)$ converges to the desired score function; see (Taleb and Jutten, 1999) for a detail on adaptive score function estimation.

3.3 Approximation of the nonlinear functions

After the blind deconvolution method has converged, the output $\mathbf{y}(t)$ of the demixer is an estimation of the input $\mathbf{s}(t)$ of the linear part of the system. Let us first assume that the permutation indeterminacy was solved (see next section for resolving permutation indeterminacy). Then, for each of the nonlinear functions $f_i(\cdot)$, a set of input $\mathbf{u}(t)$ and an estimation of the corresponding set of output $\mathbf{s}(t)$ are available. Thus an approximation of the inverse of the nonlinear functions $f_i^{-1}(\cdot)$ can be obtained by using a function approximation method based on neural network, splines or polynomials among others.

In this paper, a generalized regression neural network provides the approximated inverse functions $\widehat{f_i^{-1}}(\cdot)$ (Wasserman, 1993).

Since the input signal $\mathbf{s}(t)$ is observed, it is possible to detect and correct the permutation $\mathbf{P}$ of the components of $\mathbf{y}(t)$, i.e. $\mathbf{y}(t) \approx \mathbf{P}\ \mathbf{u}(t)$, by computing the covariance

$$\Gamma_{sy} = \mathcal{E}\left\{\mathbf{s}(t)\mathbf{y}(t)^T\right\} \approx \mathcal{E}\left\{\mathbf{s}(t)\mathbf{u}(t)^T\right\}\ \mathbf{P}^T \approx \begin{bmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_m \end{bmatrix} \mathbf{P}^T$$

since $\mathcal{E}\left\{s_i(t) f_j(s_j(t))\right\} = \gamma_i\ \delta_{i,j}$ with $\delta_{i,j}$ the Kronecker symbol (the γ_i are unknown but supposed non null). By taking $\mathbf{P} = (\text{sign}(\Gamma_{sy}))^T$, any inversion of a the components is corrected.

In the proposed identification setting, both $\mathbf{H}(z)$ and $\mathbf{W}(z)$ are biproper thus there is no delay indeterminacy. But in the non biproper case, the delay indeterminacy is resolved along with the permutation indeterminacy by computing the covariances of $s_i(t)$ and $y_j(t+\tau)$ for different values of τ.

The scale indeterminacy of the blind deconvolution method is compensated by the approximation of the inverse nonlinear functions. Suppose $y_i(t) \approx \lambda_i u_i(t)$, then the estimated inverse is $\widehat{f_i^{-1}}(x) = f_i^{-1}(x/\lambda_i)$

4. NUMERICAL SIMULATION

Consider the transfer matrix

$$\mathbf{H}(z) = \begin{bmatrix} \dfrac{z}{z-0.7} & \dfrac{-0.7(z-0.3)}{z+0.4} \\ \dfrac{0.2(z-0.5)}{z-0.6} & \dfrac{z-0.5}{z+0.3} \end{bmatrix},$$

as the linear part, whose impulse response is given in Fig. 4 and the nonlinear functions

$$f_1(u_1) = \tanh(2u_1),$$
$$f_2(u_2) = \tanh(3u_2).$$

For all figures representing impulse or frequency responses, the subplot at the i^{th} row and j^{th} column corresponds to the transfer from the j^{th} component of the input to the i^{th} component of the output. Namely, (i,j)-subplot represents (i,j) entry in the transfer matrix.

The input signal components are uniformly distributed and 5000 samples have been used for the identification. The observed signals are corrupted by small additive Gaussian noises. The SNR are 33dB for $v_1(t)$ and 31dB for $v_2(t)$.

The joint density of the components is plotted in Fig.5-a (This plot is obtained by plotting 500 points of the signals). A severe effect of the nonlinear functions is visible on the joint density of $\mathbf{s}(t)$ in Fig.5-b. Then the linear part of the system transform the joint density in the one plotted in Fig.5-c.

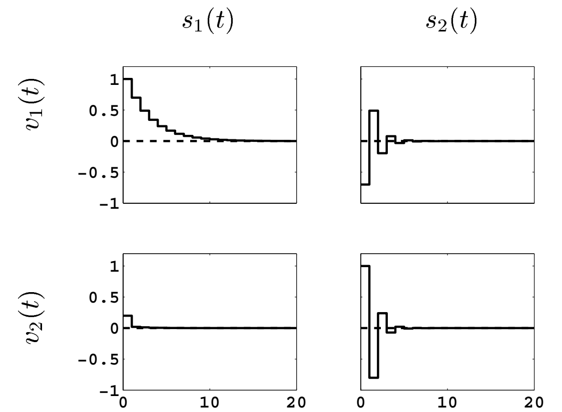

Fig. 4. Impulse response of each entry in $\mathbf{H}(z)$.

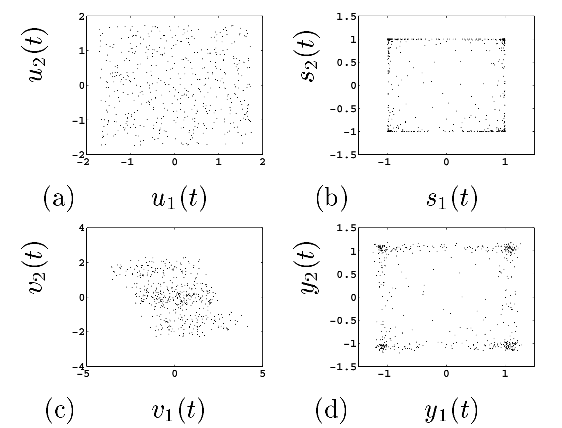

Fig. 5. Evolution of joint density: original source (a), after distortion (b), after filtering (c) and recovered distorted source (d).

4.1 Linear part

The length of the FIR demixer $\mathbf{W}(z)$ is 15 (degree $l = 14$). Figure 6 shows that, after convergence, $\mathbf{W}(z)\mathbf{H}(z)$ is a diagonal constant matrix. This means that the blind deconvolution is successfully achieved. Thus the joint density of the components of $\mathbf{y}(t)$ (Fig.5-d) is similar, except the dispersion introduced by the additive noise and a scaling of the components, to the one of the components of $\mathbf{s}(t)$ (Fig.5-b). Namely, $(y_1(t), y_2(t)) \approx (\alpha_1 s_1(t), \alpha_2 s_2(t))$ + estimation noise (where α_1 and α_2 are the scaling factors, here close to one).

4.2 Compensation of nonlinear functions

The inverses of the functions f_1, f_2 are approximated with a generalized regression neural net-

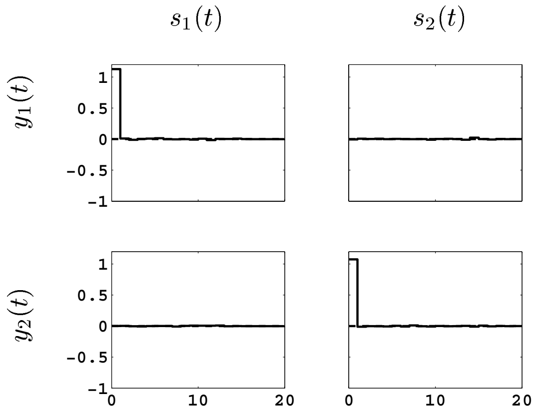

Fig. 6. Impulse response of each entry in the cascade $\mathbf{W}(z)\mathbf{H}(z)$.

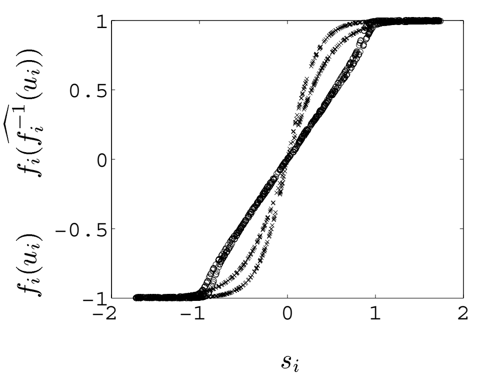

Fig. 7. Compensation of nonlinear functions.

work (the spread of the radial basis functions is 0.07). In Fig. 7, the input of the linear part are plotted with and without compensation. The hyperbolic distortions are linearized by the compensation.

4.3 *Example: Feedforward control*

In this part, the proposed method is used to improve the performance of a feedforward controller for the previous system. If there is no nonlinearity, the ideal feedforward controller is the inverse of the system. First, as a reference, the estimate of the inverse system is obtained by using an identification method that does not consider the nonlinear part. Here, the *N4sid* method from the identification toolbox is used. Then the proposed method is also applied. In fact the feedforward controller is obtained by the blind deconvolution method during the estimation of the nonlinear part. The inverse of the system obtained by N4sid and the proposed method are both very close to the true inverse as seen in figs.8 and 9.

However, the tracking performances are not the same because of the non-linear part. In figs.10 and 11, the tracking performances are significantly

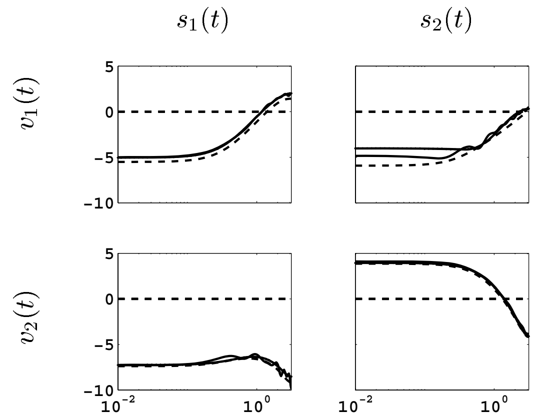

Fig. 8. Gain of the true inverse of the system (dashed line), the inverse obtained with N4sid (dashed dotted line) and the inverse obtained with the proposed method (solid line).

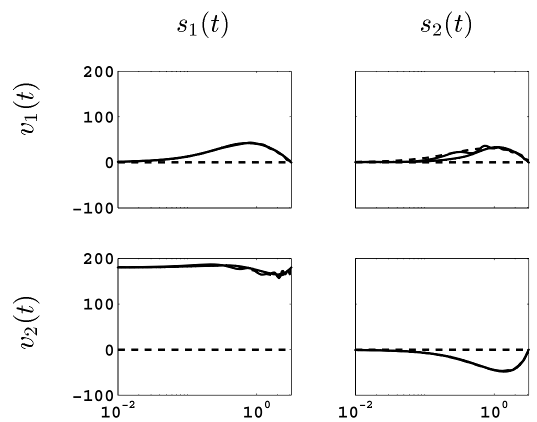

Fig. 9. Phase of the true inverse of the system (dashed line), the inverse obtained with N4sid (dashed dotted line) and the inverse obtained with the proposed method (solid line).

better with the proposed method. Figs.12 and 13 show the corresponding quadratic error between commands and outputs with the two methods. These results show that the proposed method gives a significant improvement of the feedforward controller performance by modeling and compensating the nonlinearities.

5. CONCLUSION

This paper proposes a simple method for the identification of linear systems with component wise non-linearities at the inputs. By using independent component analysis, this method recovers the outputs of the nonlinear blindly from the outputs of the linear part. Thus this methods gives an effective way to separate the identification of linear and non linear parts. Hence it is possible to apply

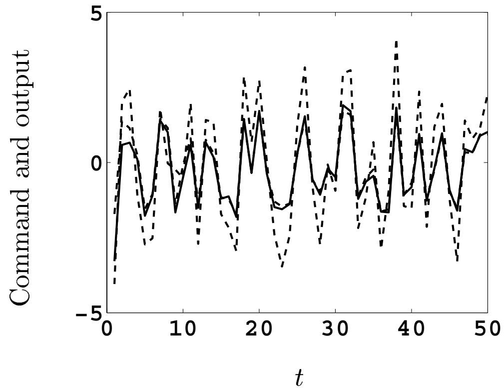

Fig. 10. Command input (solid line), output by proposed method (solid line) and output by N4sid (dashed line) of the first channel, respectively.

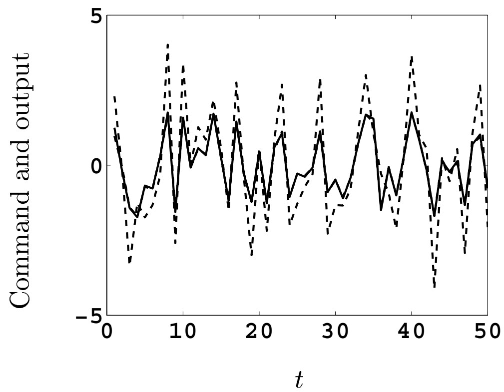

Fig. 11. Command input (solid line), output by proposed method (solid line) and output by N4sid (dashed line) of the second channel, respectively.

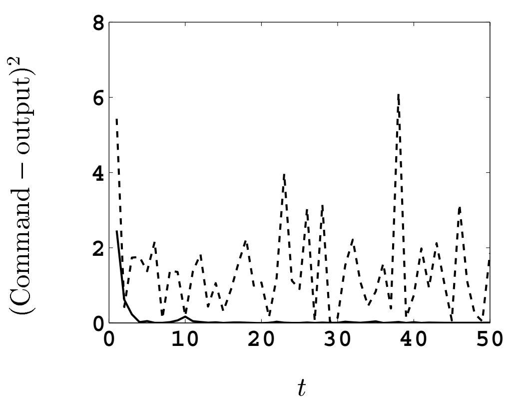

Fig. 12. Quadratic error with N4sid (dashed line) and with the proposed method (solid line) for the first channel.

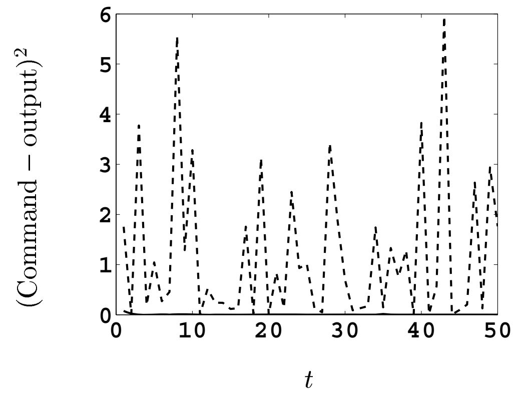

Fig. 13. Quadratic error with N4sid (dashed line) and with the proposed method (solid line) for the second channel.

another identification method for the linear part if the FIR filter model of the proposed method is not well suited for a given application. In this case, the proposed method can be seen as a preprocessing that compensates the non-linearities in order to apply a method designed for linear systems.

Appendix A. MUTUAL INFORMATION

Signal components are called statistically independent, if their joint probability density function $P_{\mathbf{y}}(\mathbf{y})$ can be factorized in the product of the marginal probability density functions P_{y_i}:

$$P_{\mathbf{y}}(\mathbf{y}) = \prod_{i=1}^{m} P_{y_i}(y_i)$$

In order to measure how far the components are from statistical independence, we can use the Kullback-Leibler divergence between the joint probability density function and the product of marginal probability density functions:

$$\mathcal{K}\left(P_{\mathbf{y}}(\mathbf{y}) | \prod_{i=1}^{m} P_{y_i}(y_i)\right) = \int P_{\mathbf{y}}(\mathbf{y}) \ln \left| \frac{P_{\mathbf{y}}(\mathbf{y})}{\prod_{i=1}^{m} P_{y_i}(y_i)} \right| d\mathbf{y},$$

which vanishes if and only if the output components are independent. This quantity measures the mutual information between the components of the output signals.

It can be expressed with the entropy of $\mathbf{y}$ and the marginal entropies:

$$\mathrm{MI}(\mathbf{y}) = \mathcal{K}\left(P_{\mathbf{y}}(\mathbf{y}) | \prod_{i=1}^{m} P_{y_i}(y_i)\right)$$
$$= -H(\mathbf{y}) + \sum_{i=1}^{m} H(y_i),$$

where $H(\mathbf{x}) = -\int P_{\mathbf{x}}(\mathbf{x}) \ln P_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}$.

Assuming ergodic statistical properties of the output signals, the cost function is:

$$\mathrm{MI}(\mathbf{y}(k)) = -H(\mathbf{y}(k)) + \sum_{i=1}^{m} H(y_i(k)).$$

The true marginal distributions of the sources are unknown so that the marginal entropies $H(y_i(k))$ must be estimated using a distribution Q_{y_i}. Q_{y_i} may be determined using some prior knowledge or estimated from the output components.

REFERENCES

Cardoso, J. F. (1998). Blind signal separation: statistical principles. *Proceeding of the IEEE* **10**(9), 2009–2025.

Cichocki, A. and S. Amari (2002). *Adaptive Blind Signal and Image Processing : Learning Algorithms and Applications*. John Wiley & Sons.

Even, J. and K. Sugimoto (2006). Blind estimation of row relative degree via constrained mutual information minimization. *6th International Conference on Independent Component Analysis and Blind Source Separation ICA 06, Charleston, USA* pp. 352–359.

Greblicki, W. (1989). Non-parametric orthogonal series identification of hammerstein systems. *International journal of systems science* **20**(12), 2355–2367.

McKelvey, T. and C. Hanner (2003). On identification of hammerstein systems using excitation with a finite number of levels. *Proceeding of the 13th International Symposium on System Identification, Rotterdam, The Netherland* pp. 57–60.

Pawlak, M. (1991). On the series expansion approach to the identification of hammerstein systems. *IEEE Trans. Automatic Control* **36**, 736–767.

Sugimoto, K., A. Suzuki, M. Nitta and N. Adachi (2005). Polynomial matrix approach to independent component analysis: (part ii) application. *IFAC World Congress, Prague, Czech Republic.*

Sugimoto, K. and M. Nitta (2005). Polynomial matrix approach to independent component analysis: (part i) basics. *IFAC World Congress, Prague, Czech Republic.*

Taleb, A. and C. Jutten (1999). Source separation in post-nonlinear mixtures. *IEEE Trans. on SP* **47**(10), 2807–2820.

Wasserman, P.D. (1993). *Advanced Methods in Neural Computing*. Van Nostrand Reinhold, New York.