Semi-blind Technique for Non-stationary Noise Cancellation in System Identification

Kenji Sugimoto and Jani Even

Graduate School of Information Science
Nara Institute of Science and Technology
8916-5, Takayama-cho, Ikoma, Nara, 630-0192, Japan

E-mail: {kenji, even}@is.naist.jp

Abstract

This paper proposes a method for signal recovery from noisy observation in system identification. From an ICA (Independent Component Analysis) point of view, this method estimates the noise model by resorting the problem to what is called semi-blind deconvolution, and then cancels the noise effect, thereby enabling us to identify the system more accurately. An adaptive (on-line) semi-blind method is described with focus upon the cost function. Numerical example for a spring-mass-damper system illustrates the result.

1 Introduction

In system identification, we need to measure the input and output signals with sufficient accuracy. Since the output is often contaminated by noise, it is desirable if we can obtain clean signals by canceling the noise effect. In this paper we consider the problem of recovering the output signal from noisy observation by means of a semi-blind and adaptive deconvolution method.

In general, the blind deconvolution problem is that of recovering input signals of a multivariable dynamical system only from its output signals. Many efficient methods have been proposed to this end and applied in various fields such as communication or biomedical signal processing (see book [1] and references therein). Since most of the methods use statistic independence, we generically call this methodology Independent Component Analysis (ICA). Otherwise we call it blind deconvolution when dynamical systems are treated [2, 3, 4, 5].

Recently it has turned out that this methodology is also useful in control engineering because it enables us to identify a system without complete knowledge of input signals, if they are statistically independent along with some other assumptions. This is because the signal recovery (deconvolution) amounts to constructing an inverse system, which then gives information of the original system. Possible applications are, among others, disturbance suppression, change detection, modeling and canceling nonlinear distortion [6, 7, 8]. Extensions to strictly proper and nonlinear cases are treated in [9, 10].

In system identification for control, we manipulate the

plant ourselves and hence (command) input signals are usually available, although there are some unavailable inputs like additive noise. If we make use of the available components of input, it will be more realistic than regarding the input as totally unknown. In view of this, the authors proposed a semi-blind deconvolution method [11], where we identify the system with known components of input and observed outputs. The authors then gave an adaptive (on-line) version of this method, with application to noise cancellation [12]. However, its adaptation process has not been fully described yet.

The objective of this paper is as follows. First, we give a complete description of the adaptive semi-blind deconvolution, which makes full use of projection of the gradient for search [11, 12], and we clarify the corresponding cost function in optimization. This process amounts to constructing a noise model by reducing the signals to basic components named “source” signals. They may not have a physical meaning but do form a virtual noise generator together with the noise model driven by them. Thus we can analyze statistical properties of noise.

Secondly, we recover the noise by means of the deconvoluted source signal and the noise model. Then we cancel the noise effect by subtracting the recovered noise signal from the observed signal. This enables us to identify the system more accurately. We illustrate the method with a typical example in control: a spring-mass-damper system.

The paper is organized as follows. As preliminary, Sect. 2 introduces blind deconvolution and the method we will modify in later sections. In Sect. 3, the main part of the paper, the semi-blind deconvolution is proposed using an extended system representation. We then explain adaptive noise cancellation. Finally in Sect. 4, numerical simulations are given to validate effectiveness of our method.

In what follows, $\mathcal{I}_n$ and $\mathcal{O}_{p \times q}$ respectively denote the identity and zero matrices whose sizes are in the suffices. A state space representation for a transfer matrix may be written in the compact form

$$H(z) = \begin{bmatrix} A_H & B_H \\ C_H & D_H \end{bmatrix} = C_H(z\mathcal{I}_n - A_H)^{-1}B_H + D_H,$$

where n is the state dimension. Throughout the paper t denotes discrete-time. We sometimes confuse time and frequency domain signals with a slight abuse.

2 Preliminary: Blind Deconvolution

Fig. 1 illustrates the blind deconvolution scheme. The unknown signal vector $\mathbf{s}(t) \in \mathbb{R}^m$, called “source”, is filtered by the unknown transfer matrix $H(z)$, called “mixer”, and the signal vector $\mathbf{v}(t) \in \mathbb{R}^m$ is observed. $H(z)$ is assumed to be rational, stable, minimal phase, and biproper; i.e., $H(\infty)$ is nonsingular.

The goal of blind deconvolution is to recover the source $\mathbf{s}(t)$; i.e., to find the transfer matrix $W(z)$, called “demixer”, which generates $\mathbf{y}(t) \in \mathbb{R}^m$ that gives a certain estimate of the source; i.e., $\mathbf{y}(t) \approx \mathbf{s}(t)$ in a sense.

Fig. 1: Blind deconvolution scheme.

Most of the existing methods rely on the following theorem (the original result can be traced back to [14] and was adapted by some authors [3, 4]).

Theorem 1. Consider the relation $\mathbf{y} = W(z)H(z)\mathbf{s}$. Assume that the components of signal $\mathbf{s}(t)$ are:

- i) statistically mutually independent and independently identically distributed (i.i.d.), that is: $s_i(t)$ and $s_j(t')$ are statistically independent for $i \neq j$ or $t \neq t'$,
- ii) and have non Gaussian distributions (whose variances are assumed to be equal to unity for simplicity).

Then, the components of $\mathbf{y}(t)$ are statistically independent and have non Gaussian distributions if and only if

$$W(z)H(z) = P \Gamma(z) \quad (1)$$

for some permutation matrix P and

$$\Gamma(z) = \text{diag} \left(\alpha_1 z^{-\beta_1}, \dots, \alpha_m z^{-\beta_m} \right). \quad \square$$

P and $\Gamma(z)$ are referred to as “indeterminacy.” The above theorem states that the blind deconvolution problem can be solved, up to the indeterminacy of P and $\Gamma(z)$, by finding estimates $\mathbf{y}(t)$ that are statistically independent. Thus most of the blind deconvolution methods adapt the demixer $W(z)$ so that the components of $\mathbf{y}$ are “as independent as possible”. This is done by minimizing a cost function measuring the independence of the estimates. A popular choice is the mutual information

$$\begin{aligned} \text{MI}(\mathcal{Y}) &= \text{KL} \left(P(\mathcal{Y}) \middle| \prod_{i=1}^m \prod_{t=1}^L P(y_i(t)) \right) \\ &= \frac{1}{L} \int P(\mathcal{Y}) \log \left(\frac{P(\mathcal{Y})}{\prod_{i=1}^m \prod_{t=1}^L P(y_i(t))} \right) d\mathcal{Y}, \end{aligned} \quad (2)$$

where we define $\mathcal{Y} = (\mathbf{y}(1), \dots, \mathbf{y}(L))$.

Let us now describe the blind deconvolution method by Zhang and Cichocki [15]. This is based on the adaptation of a finite impulse response (FIR) demixer $W(z)$ of length $l + 1$ taps, represented by the state space model

$$\begin{aligned} \mathbf{x}(t+1) &= A_W \mathbf{x}(t) + B_W \mathbf{u}(t) \\ \mathbf{y}(t) &= C_W \mathbf{x}(t) + D_W \mathbf{u}(t), \end{aligned}$$

$$A_W = \begin{bmatrix} \mathcal{O}_{m \times ml} & \mathcal{O}_{m \times m} \\ \mathcal{I}_{ml} & \mathcal{O}_{ml \times m} \end{bmatrix}, B_W = \begin{bmatrix} \mathcal{I}_m \\ \mathcal{O}_{ml \times m} \end{bmatrix}. \quad (3)$$

$W(z)$ being an FIR filter, the matrices A_W and B_W are fixed as above whereas C_W and D_W will be adapted to minimize $\text{MI}(\mathbf{y}(t))$. Zhang and Cichocki [15] have shown that with this demixer, the on-line mutual information minimization is equivalent to minimizing the cost function

$$\mathcal{J}(\mathbf{y}(t), D_W, C_W) = -\log |\det(D_W)| - \sum_{i=1}^m \log P_{y_i}(y_i) \quad (4)$$

derived from equation (2). Let $C_W^{(t)}$ and $D_W^{(t)}$ denote the matrices at time t and $\mu(t)$ be the positive adaptation step used at time t . Adaptation rules are obtained by using the natural gradient of $\mathcal{J}(\mathbf{y}(t), D_W, C_W)$ (see [15] for derivation of the adaptation rules). The matrices are updated by

$$\begin{aligned} C_W^{(t+1)} &= C_W^{(t)} - \mu(t) \Delta C_W^{(t)} \\ D_W^{(t+1)} &= D_W^{(t)} - \mu(t) \Delta D_W^{(t)}, \end{aligned} \quad (5)$$

with

$$\begin{aligned} \Delta C_W^{(t)} &= [\Lambda(t) - \psi(\mathbf{y}(t))\mathbf{y}^T(t)] C_W^{(t)} - \psi(\mathbf{y}(t))\mathbf{x}^T(t) \\ \Delta D_W^{(t)} &= [\Lambda(t) - \psi(\mathbf{y}(t))\mathbf{y}^T(t)] D_W^{(t)}, \end{aligned}$$

where

$$\Lambda(t) = \text{diag} \left(\psi_1(y_1(t))y_1(t), \dots, \psi_m(y_m(t))y_m(t) \right)$$

and $\psi(\cdot) = (\psi_1(\cdot) \dots \psi_m(\cdot))^T$ is a vector consisting of approximations of the score functions $\psi_i(s) = -\partial \ln[P_{s_i}(s)]/\partial s$, associated with the (unknown) source signals. For example, a good approximation for sub-Gaussian distribution is $\psi_i(y) = y^3$ (see [1] for a discussion on approximation of the score function). The matrix $\Lambda(t)$ is used to enforce a non-holonomic constraint so that the terms $\psi_i(y_i(t))y_i(t)$ are not constrained to specific values.

We take $D_W^{(0)} = \mathcal{I}_m$, $C_W^{(0)} = \mathcal{O}_{m \times ml}$ at initialization. Consequently the demixer remains biproper during adaptation because of the update rule equation (5) [15].

3 Main Results

3.1 Problem Formulation

Let us now consider the model depicted in Fig. 2. $H_c(z)$ denotes the plant to be identified. The command input signal $\mathbf{c}(t) \in \mathbb{R}^q$ is assumed to be known, but the observed signals $\mathbf{v}(t) \in \mathbb{R}^p$ is contaminated by the additive noise $\mathbf{n}(t) \in \mathbb{R}^p$. The square transfer matrix $H_d(z)$ and the source $\mathbf{d}(t) \in \mathbb{R}^p$

Fig. 2: Additive noise model

form a generative model of the additive noise. Our goal is to build a filter that eliminates the additive noise from the output signal, in order to recover ‘clean’ observation.

We assume that both $H_c(z)$ and $H_d(z)$ are proper and stable. Furthermore, $H_d(z)$ is minimal phase and biproper (If not biproper, then it suffices to modify the source $\mathbf{d}(t)$, since our goal is to identify $\mathbf{n}(t)$ and not $\mathbf{d}(t)$; see Sect. 3.4). $H_c(z)$ can be strictly proper and non-minimal phase, and is not restricted to be square.

Note that $\mathbf{c}(t)$ is known contrary to the conventional blind deconvolution. We assume that the components of $\mathbf{d}(t)$ are statistically independent and *i.i.d.* with non Gaussian distribution, and that they are also independent of the components of $\mathbf{c}(t)$. We regard $\mathbf{d}(t)$ as a set of basic factors that may not have any physical meaning but drive the model $H_d(z)$ to generate the noise $\mathbf{n}(t)$.

Conventionally the noise was assumed to be Gaussian mainly because of simplicity, but the above generative model gives a richer description. For example, in a later section we will treat a distribution similar to Gaussian but with a finite support, which is rather realistic (see the histograms in Fig. 7 in numerical simulation).

We will derive a learning rule by modifying that of Zhang and Cichocki [15], where they assumed the source signals are *i.i.d.*, in order to show that these signals are recovered by their method. However, we will show that our modified learning is still valid without *i.i.d.* assumption on $\mathbf{c}(t)$.

3.2 Extended Mixer

Fig. 3: Virtual observations and extended mixer.

By regarding the known $\mathbf{c}(t)$ as “virtual” observations, we

obtain the extended mixer (see Fig. 3):

$$\begin{bmatrix} \mathbf{v}(t) \\ \mathbf{c}(t) \end{bmatrix} = H(z) \begin{bmatrix} \mathbf{d}(t) \\ \mathbf{c}(t) \end{bmatrix}, \quad H(z) = \begin{bmatrix} H_d(z) & H_c(z) \\ \mathcal{O}_{q \times p} & \mathcal{I}_q \end{bmatrix} \quad (6)$$

By assumptions on H_d and H_c , the system (6) is also stable, biproper, and minimal phase. Let us take the demixer

$$\begin{bmatrix} \mathbf{y}_p(t) \\ \mathbf{y}_q(t) \end{bmatrix} = W(z) \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{c}(t) \end{bmatrix}, \quad W(z) = \begin{bmatrix} W_d(z) & W_c(z) \\ \mathcal{O}_{q \times p} & \mathcal{I}_q \end{bmatrix}, \quad (7)$$

with blocks compatible with equation (6). It is clear that $\mathbf{y}_q(t) \equiv \mathbf{c}(t)$. We have to adapt $W_d(z)$ and $W_c(z)$ to obtain $\mathbf{y}_p(t) \approx \mathbf{d}(t)$.

Note: A similar approach was proposed in [13] for using known input signals, but this was in an instantaneous blind signal separation problem, i.e. when the mixer and the demixer are constant matrices.

3.3 Adaptation with Fixed Blocks

We realize $W_d(z)$ and $W_c(z)$ as FIR filters of length $l + 1$ taps as follows:

$$W_d(z) = \left[\begin{array}{c|c} A_{W_d} & B_{W_d} \\ \hline C_{W_d}^{(t)} & D_{W_d}^{(t)} \end{array} \right], \quad W_c(z) = \left[\begin{array}{c|c} A_{W_c} & B_{W_c} \\ \hline C_{W_c}^{(t)} & D_{W_c}^{(t)} \end{array} \right] \quad (8)$$

where matrices

$$A_{W_d} = \begin{bmatrix} \mathcal{O}_{p \times pl} & \mathcal{O}_{p \times p} \\ \mathcal{I}_{pl} & \mathcal{O}_{pl \times p} \end{bmatrix}, \quad B_{W_d} = \begin{bmatrix} \mathcal{I}_p \\ \mathcal{O}_{pl \times p} \end{bmatrix},$$

$$A_{W_c} = \begin{bmatrix} \mathcal{O}_{q \times ql} & \mathcal{O}_{q \times q} \\ \mathcal{I}_{ql} & \mathcal{O}_{ql \times q} \end{bmatrix}, \quad B_{W_c} = \begin{bmatrix} \mathcal{I}_q \\ \mathcal{O}_{ql \times q} \end{bmatrix}$$

are fixed, while $C_{W_d}^{(t)}$, $D_{W_d}^{(t)}$, $C_{W_c}^{(t)}$, $D_{W_c}^{(t)}$ are adapted. Then the demixer $W(z)$ is obtained by combining the above two systems:

$$\begin{bmatrix} \mathbf{x}_d(t+1) \\ \mathbf{x}_c(t+1) \end{bmatrix} = A_W \begin{bmatrix} \mathbf{x}_d(t) \\ \mathbf{x}_c(t) \end{bmatrix} + B_W \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{c}(t) \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{y}_p(t) \\ \mathbf{y}_q(t) \end{bmatrix} = \bar{C}^{(t)} \begin{bmatrix} \mathbf{x}_d(t) \\ \mathbf{x}_c(t) \end{bmatrix} + \bar{D}^{(t)} \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{c}(t) \end{bmatrix}, \quad (9)$$

$$A_W = \begin{bmatrix} A_{W_d} & \mathcal{O}_{p(l+1) \times q(l+1)} \\ \mathcal{O}_{q(l+1) \times p(l+1)} & A_{W_c} \end{bmatrix},$$

$$B_W = \begin{bmatrix} B_{W_d} & \mathcal{O}_{p(l+1) \times q} \\ \mathcal{O}_{q(l+1) \times p} & B_{W_c} \end{bmatrix}$$

$$\bar{C}^{(t)} = \begin{bmatrix} C_{W_d}^{(t)} & C_{W_c}^{(t)} \\ \mathcal{O}_{q \times pl} & \mathcal{O}_{q \times ql} \end{bmatrix}, \quad \bar{D}^{(t)} = \begin{bmatrix} D_{W_d}^{(t)} & D_{W_c}^{(t)} \\ \mathcal{O}_{q \times p} & \mathcal{I}_q \end{bmatrix}.$$

In order to obtain the update rules for $\bar{C}^{(t)}$ and $\bar{D}^{(t)}$, we want to use equation (5) given for the blind case. However, because of semi-blind nature, $\bar{C}^{(t)}$ and $\bar{D}^{(t)}$ have fixed blocks as above, so that we need to update only the upper blocks and keep the lower blocks unchanged (i.e., the zero matrices and the identity matrix).

To this end, let us project the original update rule in Sect. 2 onto the constrained optimization search plane by means of the projection matrix

$$P_d = \begin{bmatrix} \mathcal{I}_p & \mathcal{O}_{p \times q} \end{bmatrix}. \quad (10)$$

Namely, instead of equation (5), we have

$$\begin{aligned}\Delta \bar{C}^{(t)} &= P_d \left\{ [\Lambda(t) - \psi(y(t))y^T(t)] \bar{C}^{(t)} - \psi(y(t))x^T(t) \right\}, \\ \Delta \bar{D}^{(t)} &= P_d [\Lambda(t) - \psi(y(t))y^T(t)] \bar{D}^{(t)}.\end{aligned}$$

Then we can compute

$$\begin{aligned}\Delta C_{W_d}^{(t)} &= [\Lambda_p(t) - \psi(y_p(t))y_p^T(t)] C_{W_d}^{(t)} - \psi(y_p(t))x_d^T(t) \\ \Delta C_{W_c}^{(t)} &= [\Lambda_p(t) - \psi(y_p(t))y_p^T(t)] C_{W_c}^{(t)} - \psi(y_p(t))x_c^T(t) \\ \Delta D_{W_d}^{(t)} &= [\Lambda_p(t) - \psi(y_p(t))y_p^T(t)] D_{W_d}^{(t)} \\ \Delta D_{W_c}^{(t)} &= [\Lambda_p(t) - \psi(y_p(t))y_p^T(t)] D_{W_c}^{(t)} - \psi(y_p(t))y_q^T(t).\end{aligned}$$

where

$$\Lambda_p(t) = \text{diag}(\psi_1(y_1(t))y_1(t), \dots, \psi_p(y_p(t))y_p(t)).$$

The update rules of our semi-blind deconvolution method are thus given by

$$\begin{aligned}C_{W_d}^{(t+1)} &= C_{W_d}^{(t)} - \mu(t)\Delta C_{W_d}^{(t)}, \quad C_{W_c}^{(t+1)} = C_{W_c}^{(t)} - \mu(t)\Delta C_{W_c}^{(t)}, \\ D_{W_d}^{(t+1)} &= D_{W_d}^{(t)} - \mu(t)\Delta D_{W_d}^{(t)}, \quad D_{W_c}^{(t+1)} = D_{W_c}^{(t)} - \mu(t)\Delta D_{W_c}^{(t)}.\end{aligned}$$

Similarly as in Sect. 2, $W_d^{(t)}(z)$ remains biproper during adaptation when it is initialized to identity.

3.4 Noise Cancellation by Demixer

After convergence, the above learning rule gives

$$y_p(t) \approx d(t), \text{ and } W_d(z) \approx H_d(z)^{-1}$$

up to indeterminacy mentioned in Sect. 2. Then we have

$$\begin{aligned}\widehat{H}_d(z) &= W_d(z)^{-1} \\ &= \left[\begin{array}{c|c} A_{W_d} - B_{W_d}D_{W_d}^{(t)-1}C_{W_d}^{(t)} & B_{W_d}D_{W_d}^{(t)-1} \\ \hline -D_{W_d}^{(t)-1}C_{W_d}^{(t)} & D_{W_d}^{(t)-1} \end{array} \right].\end{aligned}$$

The inversion of $D_{W_d}^{(t)}$ is possible because this matrix remains nonsingular during adaptation. Furthermore, $W_d(z)$ should be minimal phase in order to assure stability of the estimated $\widehat{H}_d(z)$. This is the case for a sufficiently large number l of taps FIR approximation. But the simulations have shown that even with few taps $l \approx 12$, $W_d(z)$ is minimal phase and the stable inversion is possible.

Now we proceed to give our noise cancellation scheme as shown in Fig. 4. We recover the ‘clean’ observations by subtracting the estimated noise from the observed signal as

$$z(t) = v(t) - \hat{n}(t), \quad \hat{n}(t) = \widehat{H}_d(z)y_p(t). \quad (11)$$

Here, it is worth to note that, unlike y_p , $\hat{n}$ is *not* subject to the indeterminacy mentioned in Sect. 2.

We can thus regard the obtained z as the output of the plant to be identified: $z(t) \approx H_c(z)c(t)$. By making use of this signal, we can perform system identification with more accuracy.

This method also allows us to recover the observation even if the noise dynamics $H_d(z)$ is slowly varying with time (relatively to the adaptation rate of the semi-blind method).

Fig. 4: Adaptive noise cancellation

3.5 Discussion on the Cost Function

We have derived the update rule for the semi-blind deconvolution by projecting the original adaptation rule [15]. Let us consider the meaning of this projection from the viewpoint of the cost function in comparison with (4).

By (10) and (9) we have $\det \bar{D}^{(t)} = \det D_{W_d}^{(t)}$ and the projected cost function is written as

$$\begin{aligned}\mathcal{J}(y(t), D_{W_d}, C_{W_d}, D_{W_c}, C_{W_c}) &= -\log |\det D_{W_d}| \\ &- \sum_{i=1}^p \log P_{y_i}(y_i) - \sum_{i=p+1}^{p+q} \log P_{y_i}(y_i),\end{aligned}$$

where $y_p = (y_1, \dots, y_p)$ and $y_q = (y_{p+1}, \dots, y_{p+q})$. But the last term does not depend on the proposed demixer, thus

$$\begin{aligned}\mathcal{J}(y_p(t), D_{W_d}, C_{W_d}, D_{W_c}, C_{W_c}) &= \\ &-\log |\det D_{W_d}| - \sum_{i=1}^p \log P_{y_i}(y_i).\end{aligned} \quad (12)$$

Now define

$$\begin{aligned}\mathcal{Y}_p &= (y_p(1), \dots, y_p(L)) \\ \mathcal{C} &= (c(1), \dots, c(L)).\end{aligned}$$

Then the goal is to factorize the joint density in

$$P(\mathcal{Y}_p, \mathcal{C}) = \prod_{i=1}^p \prod_{t=1}^L P(y_i(t))P(\mathcal{C}) \quad (13)$$

in order to obtain output $y(t)$ that has statistically independent *i.i.d.* components that are also statistically independent of the command $c(t)$ (Note that the joint density of the command $P(\mathcal{C})$ is *not* factorized. This is why we need no *i.i.d.* assumption on the command signals). Consequently

$$\text{MI}(\mathcal{Y}_p, \mathcal{C}) = \text{KL} \left(P(\mathcal{Y}_p, \mathcal{C}) \middle| \prod_{i=1}^p \prod_{t=1}^L P(y_i(t))P(\mathcal{C}) \right)$$

Because of the structure of the demixer we have

$$\begin{bmatrix} y_p(1) \\ c(1) \\ \vdots \\ y_p(L) \\ c(L) \end{bmatrix} = \mathcal{H} \begin{bmatrix} v(1) \\ c(1) \\ \vdots \\ v(L) \\ c(L) \end{bmatrix}, \quad \mathcal{H} = \begin{bmatrix} \bar{D} & & 0 \\ & \bar{D} & \\ * & & \ddots \end{bmatrix}.$$

Fig. 5: Configuration of mass-spring-damper system.

Since $\det \mathcal{H} = (\det \bar{D})^L = (\det D_{W_d})^L$ we obtain

$$P(\mathcal{Y}_p, \mathcal{C}) = \frac{P(\mathcal{V}, \mathcal{C})}{|\det D_{W_d}|^L},$$

where $\mathcal{V} = (\mathbf{v}(1), \dots, \mathbf{v}(L))$. Keeping only the terms depending of the demixer the cost function simplifies to

$$\begin{aligned} \mathcal{J}(\mathcal{Y}_p, D_{W_d}, C_{w_d}, D_{W_c}, C_{w_c}) = \\ -L \log |\det D_{W_d}| - \sum_{i=1}^p \sum_{t=1}^L \log P(y_i(t)). \end{aligned}$$

As in [15], we obtain the cost function (12) for the online algorithm, thus the projected adaptation rule minimizes the desired cost function.

4 Numerical Simulation

In this section, we apply the proposed method to a mass-spring-damper system, in which observation is corrupted by a non stationary additive noise. In particular, we show that the proposed method is able to filter out the additive noise.

The mass-spring-damper system is depicted in Fig. 5 and the value of the parameters are given in Table 1.

Table 1: Parameters of the mass-spring-damper system

M_1	0.007 kg	K_1	3 N/m	D_1	0.9 Ns/m
M_2	0.015 kg	K_2	1.5 N/m	D_2	0.6 Ns/m

Simulation is performed with the discretized system (obtained by the bilinear transform to the Laplace representation of the continuous system). The sampling frequency is 50 Hz and the observation time is 300s (corresponding to 15000 samples).

As command signals $\mathbf{c}(t)$ we have applied pseudo random numeric sequences with uniform distribution (Fig. 6) to the

Fig. 6: command input signals for identification (first second).

Fig. 7: Histograms of $n_1(t)$ and $n_2(t)$.

masses M_1 and M_2 while the positions of these masses are observed. But the observed signals $\mathbf{v}(t)$ are corrupted with additive noise

$$\begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} = H_c(z) \begin{bmatrix} c_1(t) \\ c_2(t) \end{bmatrix} + K \begin{bmatrix} n_1(t) \\ n_2(t) \end{bmatrix},$$

where $H_c(z)$ is the transfer matrix of this system and K is a diagonal matrix used to set a given SNR.

The additive noises are generated by

$$\begin{bmatrix} n_1(t) \\ n_2(t) \end{bmatrix} = H_d^{(t)}(z) \begin{bmatrix} d_1(t) \\ d_2(t) \end{bmatrix},$$

where the source signals $\mathbf{d}(t) = [d_1(t), d_2(t)]^T$ are i.i.d., have uniform distributions and are statistically independent. The noise dynamics is modeled by a transfer function varying continuously from

$$H_d^{t=0}(z) = \begin{bmatrix} \frac{z}{z-0.3} & \frac{0.3z(z+0.3)}{(z-0.4)(z-0.1)} \\ \frac{-0.4z^2}{(z+0.6)(z+0.5)} & \frac{z^2}{(z-0.6)(z+0.4)} \end{bmatrix},$$

for $t = 0$ to

$$H_d^{t=L}(z) = \begin{bmatrix} \frac{z}{z+0.2} & \frac{-0.2z(z-0.5)}{(z-0.8)(z+0.3)} \\ \frac{0.7z^2}{(z+0.3)(z-0.7)} & \frac{z^2}{(z-0.3)(z+0.2)} \end{bmatrix}$$

for $t = L$ (to obtain the continuously varying filters, for simplicity, we use a FIR filter whose impulse response is a linear combination of the truncated impulse responses of $H_d^{t=0}(z)$ and $H_d^{t=L}(z)$). This additive noise is non stationary and have correlated and colored components. Moreover, this generative model gives additive noises that have distributions close to the Gaussian but with finite supports (see histograms in Fig. 7). The average SNR on the 300s of observation was set to 6dB.

The parameters of the proposed method are $l = 31$ and $\mu(t) = 0.01$ (fixed step size). The approximated score functions are $\psi_i(s) = s^3$. With these parameters, the proposed method gives an estimate of the noises $\hat{\mathbf{n}}(t)$ that is close to the true ones after convergence, see Fig. 8.

To measure the quality of the recovered observation, we define the clean observation $\mathbf{v}_c(t) = H_c(z)\mathbf{c}(t)$. This is the output signal of the system part $H_c(z)$ when there is no disturbance. Then the measure of performance is the mean squared error (MSE) between the clean observation and the modified observations $\mathbf{z}(t)$ obtained by using the observation recovery scheme (for each time sample the means are obtained by low pass filtering of the instantaneous squared error and the value is the average of the two components).

Fig. 8: True (dashed) and estimated (solid) noises $n_1(t)$ and $n_2(t)$ (last 2 seconds).

Fig. 9: Mean squared error for $v_c(t) - v(t)$ (top curve) and $v_c(t) - z(t)$ (bottom curve).

The square marker curve in Fig. 9 shows the evolution of the MSE between the noisy observations $v(t)$ and the clean observations $v_c(t)$ which remains constant. The evolution of the MSE for the proposed method is the curve with circle in Fig. 9. We can see that the MSE is converging to small values and remains small whereas the disturbance dynamic $H_d^{(t)}(z)$ is varying. This indicates a good tracking performance of the proposed method.

Using the output of the proposed method and the command it is then possible to perform the identification of the system part $H_c(z)$ without the influence of the non stationary noise dynamics.

5 Conclusions

In this paper, we have proposed a realistic model for the additive noise corrupting the output of an unknown linear system whose input, the command, is known. Then we have shown that, in the presence of non stationary colored noise, the output can be estimated by means of a modified semi-blind adaptive blind deconvolution method.

References

[1] A. Cichocki and S. Amari: *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, John Wiley & Sons, 2002.
[2] B. Ans, J. Héroult, and C. Jutten: Adaptive neural architectures: Detection of primitives, *COGNITIVA'85*, pp. 593-597, 1985

[3] O. Shalvi and E. Weinstein: New criteria for blind deconvolution of non-minimum phase systems, *IEEE Trans. Inf. Theory*, **36**-2, pp. 312-321, 1990
[4] P. Comon: Independent Component Analysis, A New Concept?, *Signal Processing*, **36**, pp. 287-314, 1994.
[5] S. Amari, A. Cichocki, and H. H. Yang: *A new learning algorithm for blind signal separation*, MIT Press, Advances in neural information processing systems 8, 1996
[6] M. Nitta and K. Sugimoto: Disturbance rejection system based on Independent Component Analysis (in Japanese), *Transactions of SICE*, **42**, pp. 1313-1319, 2006
[7] M. Nitta, A. Suzuki, K. Sugimoto, and N. Adachi: Blind Identification of Mechanical Systems via Independent Component Analysis and Detections of Structure Change (in Japanese), *Transactions ISCIE*, **19**, pp. 177-184, 2006
[8] J. Even and K. Sugimoto: Modeling and Compensating Input Distortion in MIMO Systems via Blind Deconvolution, *IFAC Workshop on Adaptation and Learning in Control and Signal Processing*, 2007
[9] J. Even and K. Sugimoto: Blind Identification for Systems Non-Invertible at Infinity, *IEICE Transactions of Fundamentals*, E90-A, 1133-1143, 2007
[10] S. Fukunaga and K. Fujimoto: Nonlinear Independent Component Analysis Using State-space Models (in Japanese), *Transactions of ISCIE*, **20**-10, pp. 404-412, 2007
[11] J. Even and K. Sugimoto: An ICA Approach to Semi-blind Identification of Strictly Proper Systems Based on Interactor Polynomial Matrix, *International Journal of Robust and Nonlinear Control, Special issue on polynomial methods*, **17**-8, pp. 752-768, 2007.
[12] J. Even and K. Sugimoto: ICA Based Adaptive Disturbance Cancellation for MIMO Systems with Unknown Dynamics, *SICE-ICASE Int. Joint Conf*, Busan, Korea, 2006.
[13] M. Joho, H. Mathis, and G. S. Moschitz: Combined Blind/nonblind Source Separation Based on the Natural Gradient, *IEEE Signal Processing Letters*, **8**-8, pp. 236-238, 2001.
[14] G. Darmon: Analyse générale des liaisons stochastiques, *Rev. Inst. Internat. Stat*, **21**, pp. 2-8, 1953
[15] L. Zhang and A. Cichocki: Blind deconvolution of Dynamical System : A state Space Approach (Invited paper), *Journal of Signal Processing*, **4**-2, pp. 111-130, 2000