Weighted principal component analysis for Wiener system identification – Regularization and non-Gaussian excitations

Qinghua Zhang, Vincent Laurain, Jiandong Wang

Weighted Principal Component Analysis for Wiener System Identification – Regularization and non-Gaussian Excitations

Qinghua Zhang * **Vincent Laurain** ** **Jiandong Wang** ***

* *INRIA Rennes Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France (email: qinghua.zhang@inria.fr)*
** *CRAN, Université de Lorraine, CNRS, 2, rue Jean Lamour 54506 Vandoeuvre-lès-Nancy Cedex, Franc (email: vincent.laurain@univ-lorraine.fr)*
*** *College of Engineering, Peking University, Beijing, China 100871 (email: jiandong@pku.edu.cn)*

Abstract: Finite impulse response (FIR) Wiener systems driven by Gaussian inputs can be efficiently identified by a well-known correlation-based method, except those involving even static nonlinearities. To overcome this deficiency, another method based on weighted principal component analysis (wPCA) has been recently proposed. Like the correlation-based method, the wPCA is designed to estimate the linear dynamic subsystem of a Wiener system without assuming any parametric form of the nonlinearity. To enlarge the applicability of this method, it is shown in this paper that high order FIR approximation of IIR Wiener systems can be efficiently estimated by controlling the variance of parameter estimates with regularization techniques. The case of non-Gaussian inputs is also studied by means of importance sampling.

Keywords: Wiener system identification, block-oriented nonlinear system, principal component analysis.

1. INTRODUCTION

For Wiener systems with fully parametrized linear and nonlinear sub-models, their identification can generally be formulated as a non convex optimization problem. Alternatively, some Wiener system identification methods are based on some assumed particularity of the system, notably the monotonicity of the static nonlinearity. By appropriately parametrizing the inverse function of the monotonic nonlinearity, the entire Wiener system can be formulated in the form of a linearly parametrized model (Zhu, 1999; Ni et al., 2012). A different approach has been proposed in (Zhang et al., 2006) with the particularity of focusing on linear subsystem identification without assuming any explicit parametrization of the monotonic nonlinearity. Some similar methods assuming other particular properties of the static nonlinearity, more or less related to the monotonicity, have also been developed in (Bai and Reyland Jr, 2008). If the monotonicity of the static nonlinearity is not assumed, a possible solution is to rely on non parametric identification methods (Greblicki, 1992; Hu and Chen, 2005; Pawlak et al., 2007).

When the input signal is randomly generated following a Gaussian distribution, it is well known that the identification of the linear subsystem can be easily separated from the unknown nonlinearity (Billings and Fakhouri, 1982; Greblicki, 1992). This approach is of particular interest when little prior knowledge about the nonlinearity is available. In this case, the cross-correlation between the input and the output of the Wiener system provides a consistent estimation of the finite impulse response (FIR) coefficients, up to a common unknown scalar factor. The related Wiener system identification method will be referred to as the *correlation-based method* in this paper. Under the particularly assumption of Gaussian input, this method is equivalent to the best linear approximation (BLA) in the sense of least squares (see the Appendix at the end of this paper), which is usually applied within a much larger framework (Wong et al., 2012; Enqvist and Ljung, 2005).

Another method based on Gaussian-distributed inputs has been proposed in (Zhang and Laurain, 2014), namely the weighted principle component analysis (wPCA) method, mainly aiming at addressing a deficiency of the correlation-based method, when the static nonlinearity in a Wiener system is an *even function*. As explained in the Appendix at the end of this paper, the correlation-based method is consistent for the estimation of the FIR of the linear subsystem, *except in the case of Wiener systems involving even nonlinearities*. To avoid this singular case, a simple modification of the correlation-based method is also mentioned in the appendix, under some assumptions about experimental conditions.

The present paper enhances the results of (Zhang and Laurain, 2014) in two directions. In order to estimate high order FIR models, typically approximating stable infinite impulse responses (IIR), *regularization* techniques are incorporated into the wPCA method in this paper.

By reducing the variance of parameter estimates, the regularized wPCA method is more suitable for high order model estimations. The other enhancement presented in this paper is for the purpose of relaxing the assumption of Gaussian input distribution, based on techniques of *importance sampling*.

2. PROBLEM DESCRIPTION

The single-input single-output (SISO) Wiener system considered in this paper is in the form of

$$z(t) = \sum_{s=0}^{n-1} h(s)u(t-s) \tag{1a}$$

$$y(t) = f(z(t)) + e(t) \tag{1b}$$

where $t = 1, 2, 3, \ldots$ represents the discrete time instants, the input $u(t) \in \mathbb{R}$ is independently and identically distributed (i.i.d.) with a symmetric probability density function around $u = 0$, $y(t) \in \mathbb{R}$ is the output, $z(t) \in \mathbb{R}$ an internal variable, $e(t) \in \mathbb{R}$ a stationary noise independent of the input, $h(s) \in \mathbb{R}$ denotes the finite impulse response (FIR) coefficients, and $f(\cdot)$ is an unknown nonlinear function. Note that no whiteness nor any particular distribution of the noise $e(t)$ is necessary. The purpose of system identification in this framework is to estimate the FIR sequence $h(0), h(1), \ldots, h(n-1)$ from the input-output sequences $u(t), y(t)$ for $t = 1, 2, 3, \ldots$, *without assuming any parametric form* of the nonlinear function $f(\cdot)$.

Define

$$\varphi(t) \triangleq \begin{bmatrix} u(t) \\ u(t-1) \\ \vdots \\ u(t-n+1) \end{bmatrix}, \quad \theta \triangleq \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(n-1) \end{bmatrix}, \tag{2}$$

then the covariance matrix

$$\mathrm{cov}[\varphi(t)] = \sigma_u^2 I_n \tag{3}$$

where σ_u^2 is the variance of $u(t)$ and I_n the $n \times n$ identity matrix. The Wiener system model (1) is then rewritten as

$$y(t) = f(\varphi^T(t)\theta)) + e(t). \tag{4}$$

To avoid the scale indetermination between θ and $f(\cdot)$, it is assumed that θ has a normalized Euclidean norm, i.e., $\|\theta\| = 1$. The sign of θ remains undetermined.

The considered Wiener system identification problem as formulated above amounts to estimating the vector θ from $\varphi(t)$ and $y(t)$.

In this considered framework, the linear subsystem is modeled by a FIR sequence. As the proposed identification method is numerically efficient for long FIR sequence estimation, most sufficiently stable infinite impulse responses (IIR) can be well approximated by the proposed method.

When the input $u(t)$ is Gaussian distributed, the well known correlation-based method, is based on the estimation of $\mathrm{E}[\varphi(t)y(t)]$. As explained in the Appendix at the end of this paper, this method is consistent, *except in the case of even nonlinearity*. The wPCA-based method studied in this paper remains consistent in this case.

The ability of separately estimating the linear subsystem is particularly important when the static nonlinearity is discontinuous, since joint estimation of both parts may suffer from the discontinuities. Discontinuous nonlinearities with known parametrizations have been studied in (Voros, 2001).

3. GRAPHICAL ILLUSTRATION OF THE WPCA METHOD

Before formally presenting the wPCA method for Wiener system identification, let us first illustrate the main idea with a simple noise-free 2D example, i.e., $\varphi(t) = [u(t), u(t-1)]^T \in \mathbb{R}^2$ (the 2D restriction is only for the purpose of graphical illustration). For this purpose, define

$$g(\varphi) \triangleq f(\varphi^T\theta) \tag{5}$$

as a function of φ. An example of $g(\varphi)$, corresponding to a the square function f, is illustrated (in the 2D case) by the surface in Figure 1-(a) over the bottom φ_1—φ_2 plane, on which are superposed the level lines (also known as isolines or contour lines) of $g(\varphi)$ and the blue dots representing the vectors $\varphi(t)$ formed from an i.i.d. input sequence $u(t)$. The level lines and the blue dots are also illustrated in Figure 1-(b) for a better visibility.

In this example, $f(\cdot)$ is a (quadratic) convex function, hence the tighter are the level lines, the higher are the local values of $g(\varphi(t))$.

In the φ_1—φ_2 plane, the direction of the vector θ characterizing the Wiener system is perpendicular to the level lines. This fact can be checked by noticing that all points falling on a straight line perpendicular to θ has the same orthogonal projection on the line along θ, hence these points all have the same value of $\varphi^T(t)\theta$, belonging to the same level line of $g(\varphi)$.

The covariance matrix $\mathrm{cov}[\varphi(t)]$ as expressed in (3) has equal eigenvalues, hence the data points of $\varphi(t)$ shown in Figure 1-(b) have no principal component direction, in the sense of the principal component analysis (PCA).

The main idea of the wPCA method is to modify $\varphi(t)$ by some weighting factor depending on $y(t)$, so that the principal component of the modified $\varphi(t)$ will appear along the direction of θ. Let $w(y(t))$ be some non-negative valued weighting function and define the weighted vector

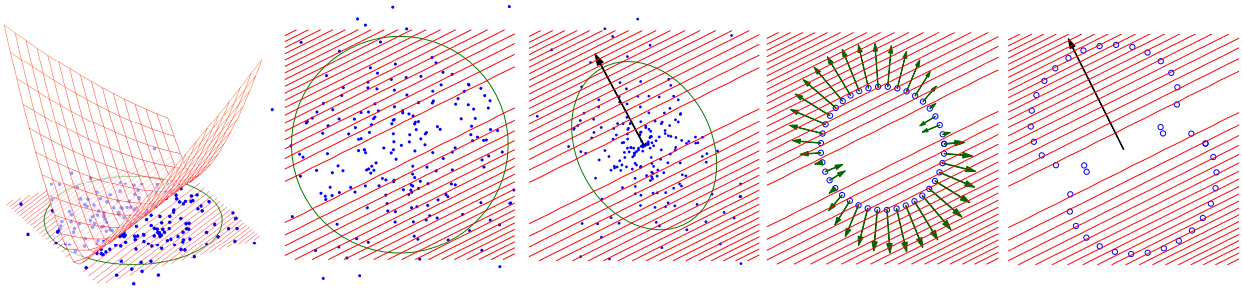
$$\varphi_w(t) \triangleq w(y(t))\varphi(t). \tag{6}$$

The weighting function $w(\cdot)$ should be appropriately chosen, so that the principal component of the modified vectors $\varphi_w(t)$ will coincide with the direction of θ. The conditions that should be satisfied by $w(\cdot)$ will be specified later (see Proposition 1).

For the 2D example where $f(\cdot)$ is a convex even function, let us choose a weighting function corresponding to an increasing function of $|y(t)|$, say

$$w(y(t)) = \alpha - e^{-\beta|y(t)|} \tag{7}$$

where $\alpha \geq 1$ and $\beta > 0$ are two chosen constants. In this considered noise-free example, $y(t) = g(\varphi(t))$, hence $\varphi_w(t)$ is weighted by a monotonically increasing function of $|y(t)| = |g(\varphi(t))|$. Hence, in Figure 1-(b), *virtually* all the points falling on the same level line would be modified by the same weight (the points in the available data set may all fall on different lines). In this example, the tighter are the level lines, the higher are the local values of $g(\varphi(t))$, hence the higher are the values of the weights. Consequently, the modification from $\varphi(t)$ to $\varphi_w(t)$ lead to displacements of the blue dots with a general trend



(a) (b) (c) (d) (e)

The direction of the vector θ is perpendicular to the level lines.

Fig. 1. 2D illustration with $\varphi(t) = [u(t), u(t-1)]^T$. (a): The function $g(\varphi(t)) = f(\varphi^T(t)\theta)$ represented by a surface over the bottom φ_1—φ_2 plane with level lines of $g(\varphi(t))$ and blue dots representing random data points (vectors $\varphi(t)$). (b): A better view of the bottom φ_1—φ_2 plane. (c): Weighted vectors $\varphi_w(t)$ represented by blue dots and their principal component direction represented by the black arrow. (d): To better illustrate the effect of weighting, the blue circles represent regularly placed data points and the green arrows indicate their displacements. (e): The regularly placed data points after weighting and their principal component direction represented by the black arrow.

perpendicular to the level lines, from the center to the outsides. After weighting, the positions of the weighted vectors $\varphi_w(t)$ represented by the blue dots are illustrated in Figure 1-(c).

At this step, an estimation of the vector θ is made by the principal component direction of the $\varphi_w(t)$ vectors, which can be found by computing the eigenvector corresponding to the largest eigenvalue of the following empirical covariance matrix

$$\Sigma_w = \frac{1}{N}\sum_{t=1}^{N} \varphi_w(t)\varphi_w^T(t), \tag{8}$$

as represented by the black arrow in Figure 1-(c).

To better illustrate the displacements of the data points caused by weighting, in Figure 1-(d) the blue circles represent regularly placed data points instead of random data points, and their displacements caused by weighting are indicated by the green arrows. The positions of the data points after weighting are shown in Figure 1-(e), with their principal component direction represented by the black arrow.

Due to its similarity to the standard PCA, this Wiener system identification method is called *weighted* PCA, or wPCA for short.

4. CONSISTENCY OF THE WPCA IN THE CASE OF GAUSSIAN INPUT

In this section the consistency of the wPCA method under the assumption of Gaussian input is shortly recalled.

Proposition 1. Assume that the input $u(t)$ is an i.i.d. sequence following the Gaussian distribution $\mathcal{N}(0, \sigma_u^2)$. If the weighting function $w(\cdot)$ satisfies

$$\mathrm{cov}[(\theta^T\varphi(t))^2, w^2(y(t))] \neq 0 \tag{9}$$

then the $n \times n$ covariance matrix of the weighted vector $\varphi_w(t)$ as defined in (6), namely $\mathrm{E}[\varphi_w(t)\varphi_w^T(t)]$, has a single eigenvalue λ_n with associated eigenvector $\pm\theta$. The remaining $n-1$ eigenvalues are all equal to each other and different from λ_n, *i.e.*, $\lambda_1 = \lambda_2 = \cdots = \lambda_{n-1} \neq \lambda_n$. □

This result proved in (Zhang and Laurain, 2014) ensures the consistency of the identification method based on the eigenvalue decomposition of the empirical covariance matrix Σ_w as in (8), when the data sample size N tends to infinity. Unlike the noise-free illustrative example of the previous section, this result has been established by taking into account the noise term $e(t)$ in (4). A typical example of the weighting function is as given in (7). See (Zhang and Laurain, 2014) for discussions about the choice of weighting functions.

5. REGULARIZATION FOR VARIANCE REDUCTION

Most dynamic systems encountered in practice have, in principle, an infinite impulse response (IIR). If an IIR system is stable, in the sense that its impulse response decays sufficiently rapidly, (high order) finite impulse response (FIR) models can be used as approximations. In this case, the identification method designed with a FIR model must be efficient enough for the estimation of high order FIR models. The wPCA method is numerically efficient for the estimation of FIR models with hundreds of FIR coefficients. However, the use of high order FIR models, with many FIR coefficients to be estimated, tends to increase the variance of the estimated coefficients. In principle, the variance of parameter estimates can be reduced by increasing data sample size, but in practice the available data may be limited. Alternatively, regularization techniques can help to reduce parameter estimation variances. A regularized wPCA method will be proposed in this section.

The eigenvalue problem as stated in Proposition 1 is asymptotically, when $N \to \infty$, equivalent to the optimization problem

$$\min_\theta \theta^T\Sigma_w\theta \quad \text{or} \quad \max_\theta \theta^T\Sigma_w\theta \tag{10}$$

depending on the sign of the correlation expressed in (9), under the constraint $\theta^T\theta = 1$. In what follows, only the case of minimization (negative correlation in (9)) will be considered, since the case of maximization can be studied similarly.

Regularization techniques are typically applied to system identification either for improving model smoothness or for obtaining sparse models (Chiuso, 2014). Here the purpose is to reduce parameter estimate variance, corresponding to the case of model smoothing. The concept of regularization

and the various techniques have been studied in many publications, for example in (Doan et al., 1984; Kitagawa and Gersh, 1984; Pillonetto and De Nicolao, 2010; Chen et al., 2012). For the particular problem considered in this paper, quadratic penalty will be added to the optimization criterion (10). The regularized optimization problem is then formulated as

$$\min_{\theta} \theta^T \Sigma_w \theta + \theta^T D\, \theta \text{ subject to } \theta^T \theta = 1 \qquad (11)$$

where D is a positive diagonal matrix containing the penalty coefficients. In practice, the solution of this constrained optimization problem is solved by computing the eigenvector of the matrix $\Sigma_w + D$ associated to its smallest eigenvalue.

Assume that the FIR coefficients contained in the vector θ as defined in (2) is an approximation of the IIR of a stable system, whose impulse response $h(t)$ decreases to zero when $t \to +\infty$. Accordingly, the diagonal entries of D, written as a vector $d = [\gamma(0),\ \gamma(1), \ldots, \gamma(n-1)]^T$, should have increasing components, so that smaller coefficients $h(t)$ (represented by the corresponding components of θ in (11)) are more penalized. For example, a possible choice is

$$\gamma(t) = \alpha(e^{\beta t} - 1) \qquad (12)$$

with some chosen values $\alpha > 0$ and $\beta > 0$. See (Pillonetto and De Nicolao, 2010; Chen et al., 2012) for discussions about the tuning of penalty coefficients.

To illustrate the ability of regularization for variance reduction, let us consider an example of IIR system, with its linear subsystem represented by the rational transfer function

$$G(q) = \frac{0.5q^{-1}}{1 - 1.69q^{-1} + q^{-2} - 0.096q^{-3}} \qquad (13)$$

It has reasonably stable poles equal to $0.786 \pm 0.444i$ and 0.118. The static nonlinearity following this IIR linear subsystem is the square function. Driven by a Gaussian input, a data sample of length $N = 1000$ is simulated, *with the true IIR transfer function* (13), followed by the square nonlinearity. The simulated output is corrupted by an additive Gaussian output noise of 20dB signal-to-noise-ratio (SNR). An FIR Wiener model of order $n = 100$ is then estimated by the regularized wPCA, with the weighing function

$$w(y) = 1 - e^{0.3|y|}$$

and the penalty coefficients

$$\gamma(t) = 0.25(e^{0.05t} - 1).$$

For the purpose of comparison, another FIR Wiener model of the same order is also estimated with the non-regularized wPCA method, without the penalty term. The results based on a simulated data sample are shown in Figure 2. The smoothness of the estimated (truncated) impulse response is clearly improved by the regularized method at the tail of the estimated part of the impulse response.

In practice, the improvement by regularization is significant only when the data sample is not sufficiently large compared to the number of parameters to be estimated. The results depend, of course, also on the SNR.

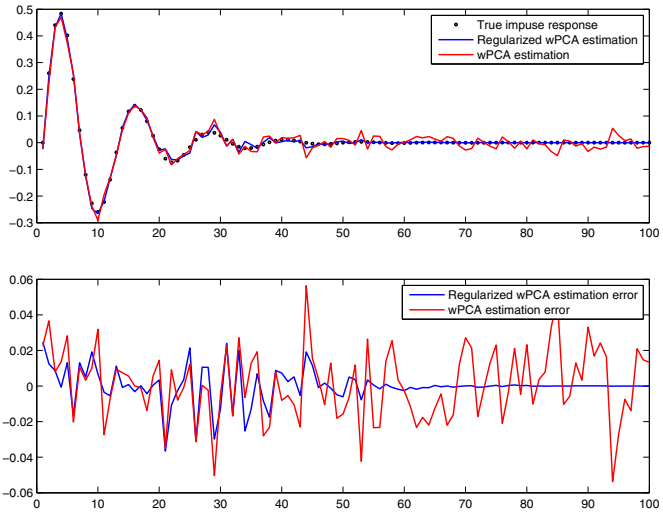


Fig. 2. Comparison between estimations by regularized wPCA (blue line) and wPCA without regularization (red line) of a truncated IIR Wiener model. The first 100 values of the true IIR impulse response are represented by dots. The data sample length $N = 1000$.

6. NON GAUSSIAN INPUTS

The Gaussian distribution condition was required for the consistency result of Proposition 1.

To deal with non Gaussian inputs, the method presented below is based on *importance sampling* (Robert and Casella, 2004, chapter 3). The main idea is to modify the computation of the empirical covariance matrix Σ_w, which was defined in (8), so that it converges to the covariance matrix defined under the Gaussian input distribution assumption, despite the fact that the modified empirical covariance matrix is based on a data set generated with a different input distribution. Importance sampling has already been applied to system identification in some cases, notably to the correlation-based method for Wiener system identification (Enqvist, 2007). See also (Wills et al., 2013).

Let $f(\varphi)$ be the probability density function (PDF) of the vector of random variables $\varphi \in \mathbb{R}^n$, and

$$\mathrm{E}_f[\omega(\varphi)] \triangleq \int_{\mathbb{R}^n} \omega(\varphi) f(\varphi) d\varphi \qquad (14)$$

be the mathematical expectation of $\omega(\varphi)$ where $\omega : \mathbb{R}^n \to \mathbb{R}$ is a function such that the integral in (14) is well defined. Accordingly define

$$\mathrm{E}_g[\omega(\varphi)] \triangleq \int_{\mathbb{R}^n} \omega(\varphi) g(\varphi) d\varphi \qquad (15)$$

where $g(\varphi)$ is the Gaussian PDF.

Notice that

$$\mathrm{E}_g[\omega(\varphi)] = \int_{\mathbb{R}^n} \omega(\varphi) g(\varphi) \frac{f(\varphi)}{f(\varphi)} d\varphi \qquad (16)$$

$$= \int_{\mathbb{R}^n} \omega(\varphi) \frac{g(\varphi)}{f(\varphi)} f(\varphi) d\varphi \qquad (17)$$

$$= \mathrm{E}_f \left[\omega(\varphi) \frac{g(\varphi)}{f(\varphi)} \right]. \qquad (18)$$

This result means that, according to the law of large numbers, given a sample sequence

$$\omega(\varphi(1)), \omega(\varphi(2)), \ldots, \omega(\varphi(N))$$

independently drawn following a given PDF $f(\varphi)$, the normalized sum

$$\frac{1}{N}\sum_{t=1}^{N}\omega(\varphi(t))\frac{g(\varphi(t))}{f(\varphi(t))} \tag{19}$$

is an empirical estimator of

$$\mathrm{E}_f\left[\omega(\varphi)\frac{g(\varphi)}{f(\varphi)}\right],$$

which turns out to be equal to $\mathrm{E}_g[\omega(\varphi)]$, according to (18).

Proposition 2. Let $u(t)$ be i.i.d. and denote with $f(\varphi)$ the corresponding PDF of $\varphi(t)$. Define

$$\Sigma_w^{\mathrm{IS}}(N) \triangleq \frac{1}{N}\sum_{t=1}^{N}\frac{g(\varphi(t))}{f(\varphi(t))}\varphi_w(t)\varphi_w^T(t) \tag{20}$$

$$= \frac{1}{N}\sum_{t=1}^{N}\frac{g(\varphi(t))}{f(\varphi(t))}w^2(y(t))\varphi(t)\varphi^T(t). \tag{21}$$

If the weighting function $w(\cdot)$ satisfies

$$\mathrm{cov}_g[(\theta^T\varphi(t))^2, w^2(y(t))] \neq 0 \tag{22}$$

then, when $N \to \infty$, $\Sigma_w^{\mathrm{IS}}(N)$ tends to an $n\times n$ matrix, which has $n-1$ equal eigenvalues, say $\lambda_1 = \lambda_2 = \cdots = \lambda_{n-1}$, and a different one $\lambda_n \neq \lambda_1$ with the associated eigenvector equal to $\pm\theta$. □

Proof of Proposition 2.

If $\varphi(t)$ was an i.i.d. vector sequence, then Proposition 2 would be a direct consequence of the above discussions, by treating each entry of the matrix $\varphi_w(t)\varphi_w^T(t)$ as $\omega(\varphi)$. However, though the input sequence $u(t)$ is i.i.d., the vector sequence composed of delayed $u(t)$ as defined in (2) is not an independent vector sequence. This detail is addressed below.

The basic law of large numbers assumes that the data sample is i.i.d.. To avoid the dependence between neighboring items of the sequence $\varphi(t)$, consider a sub-sum $\Sigma_w^{\mathrm{IS}}(N,i)$ involving one out every n consecutive items of the sum $\Sigma_w^{\mathrm{IS}}(N)$, namely

$$\Sigma_w^{\mathrm{IS}}(N,i) \triangleq \tag{23}$$

$$\frac{1}{m}\sum_{k=0}^{m-1}\frac{g(\varphi(kn+i))}{f(\varphi(kn+i))}\varphi_w(kn+i)\varphi_w^T(kn+i) \tag{24}$$

where

$$m \triangleq \left[\frac{N}{n}\right] \tag{25}$$

is the integer part of $\frac{N}{n}$ and

$$i \in \{0, 1, 2, \ldots, n-1\}.$$

According to the definition of $\varphi(t)$ in (2), the sequence $\varphi^T(kn+i)$ for $k = 1, 2, 3, \ldots$ is independent, and also i.i.d. Therefore, for each $i \in \{0, 1, 2, \ldots, n-1\}$, the partial sum $\Sigma_w^{\mathrm{IS}}(N,i)$ converges to $\mathrm{E}_g[\varphi_w(t)\varphi_w^T(t)]$, as explained before the statement of Proposition 2. When $N \to \infty$, in $\Sigma_w^{\mathrm{IS}}(N)$ the last terms after $t = nm$ can be neglected, hence

$$\Sigma_w^{\mathrm{IS}}(N) \approx \frac{m}{N}\Sigma_w^{\mathrm{IS}}(N,0) + \cdots + \frac{m}{N}\Sigma_w^{\mathrm{IS}}(N,n-1)$$

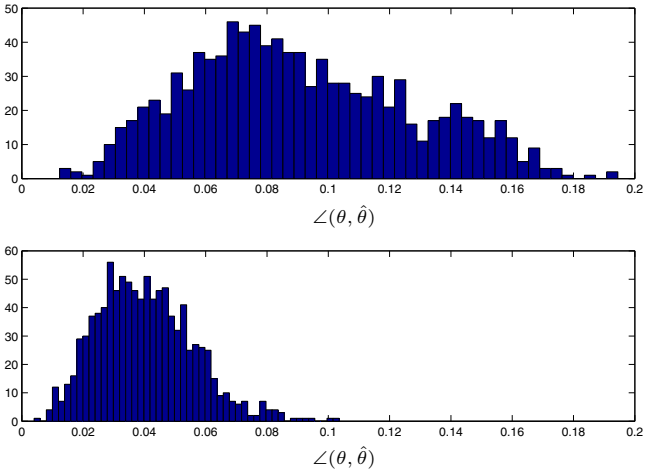


Fig. 3. Histograms of $\angle(\theta,\hat{\theta})$ (in radian) for the basic wPCA (top) and the wPCA with importance sampling (bottom) based on 1000 simulations. The smaller $\angle(\theta,\hat{\theta})$ is, the more accurate is the estimated model.

tends also to $\mathrm{E}_g[\varphi_w(t)\varphi_w^T(t)]$. The proof is then complete. □

By means of importance sampling, the Gaussian input assumption is then, in theory, relaxed. The current implementation of the algorithm based on the importance sampling is efficient for short FIR models only, since many of the weights $g(\varphi(t))/f(\varphi(t))$ for importance sampling become close to zero when n is large. Some further improvements are necessary for its application to systems approximated with high order FIR models.

As a numerical example, let us consider a FIR Wiener system with $n = 5$ and the square function static nonlinearity. A data sample of length $N = 5000$ is generated in every simulation, driven by an i.i.d. input $u(t)$ following the triangular PDF:

$$f(u) = \begin{cases} u+1 & \text{if } -1 \le u < 0 \\ 1-u & \text{if } 0 \le u < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

The simulation is repeated 1000 times, with different random realizations of the input, the output noise, and also the impulse response $h(t)$ with $n = 5$. The output noise is Gaussian and has a SNR of 20 dB.

In each simulation, to assess how close is the estimate $\hat{\theta}$ to the true vector θ, the angle between the two *normalized* vectors

$$\angle(\theta,\hat{\theta}) \triangleq \arccos(|\theta^T\hat{\theta}|) \tag{27}$$

measured in radian is used. Notice that $\angle(\theta,\hat{\theta}) \in [0, \pi/2]$, and that the 0 value means perfect fit, whereas $\pi/2 \approx 1.57$ corresponds to the worst fit.

The histograms of $\angle(\theta,\hat{\theta})$ based on the 1000 simulations, both for the basic wPCA and the wPCA with importance sampling, are illustrated in Figure 3. The improvement by importance sampling in this example is obvious.

7. CONCLUSION

The recently developed wPCA method for Wiener system identification has been improved in this paper in two aspects. The use of regularization techniques reduces effectively the error variance for the estimation of high order FIR models. On the other hand, Gaussian input assumption is relaxed by means of importance sampling. These results have been confirmed by numerical simulations.

APPENDIX – Correlation-based method

Define

$$z(t) \triangleq \theta^T \varphi(t) \tag{28}$$

$$\tilde{\varphi}(t) \triangleq (I_n - \theta\theta^T)\varphi(t) \tag{29}$$

then

$$\varphi(t) = \theta z(t) + \tilde{\varphi}(t) \tag{30}$$

It is easy to check that $\mathrm{E}[z(t)\tilde{\varphi}^T(t)] = 0$ by noting (3) and that $\|\theta\| = 1$. Under the Gaussian distribution assumption, this fact implies that that $z(t)$ and $\tilde{\varphi}(t)$ are independent.

The noise e is assumed independent of φ, then

$$\mathrm{E}(\varphi y) = \mathrm{E}[\varphi(f(z) + e)] \tag{31}$$

$$= \mathrm{E}[\varphi f(z)]. \tag{32}$$

Now apply (30),

$$\mathrm{E}(\varphi y) = \mathrm{E}[(\theta z + \tilde{\varphi}) f(z)] \tag{33}$$

$$= \mathrm{E}[\theta z f(z)] + \mathrm{E}[\tilde{\varphi} f(z)] \tag{34}$$

Because z and $\tilde{\varphi}$ are independent,

$$\mathrm{E}[\tilde{\varphi} f(z)] = \mathrm{E}(\tilde{\varphi})\mathrm{E}[f(z)] = 0, \tag{35}$$

hence

$$\mathrm{E}(\varphi y) = \mathrm{E}[\theta z f(z)] \tag{36}$$

$$= \mathrm{E}[z f(z)]\theta \tag{37}$$

where $\mathrm{E}[zf(z)]$ is an (unknown) scalar value. Therefore, the vector θ is determined by $\mathrm{E}(\varphi y)$, up to an unknown constant factor $\mathrm{E}[zf(z)]$. In practice the expectation $\mathrm{E}(\varphi y)$ is approximated by a finite sample average. This result is known as the *correlation-based method*.

The BLA in the least squares sense amounts to approximating $\left(\mathrm{E}(\varphi\varphi^T)\right)^{-1}\mathrm{E}(\varphi y) = \sigma_u^{-2}\mathrm{E}(\varphi y)$, which is equivalent to the correlation-based method under the assumption of white Gaussian input.

This method requires the condition that the unknown constant $\mathrm{E}[zf(z)] \neq 0$. If $f(z)$ is an even function, then $zf(z)$ is an odd function of z, therefore $\mathrm{E}[zf(z)] = 0$. Numerical examples (Zhang and Laurain, 2014) show that, when the nonlinearity is somehow close to an even function, the correlation-based method produces completely wrong results, whereas the wPCA-based method remains efficient.

The deficiency of the correlation-based method can also be avoided by simply adding an offset to the input signal. This solution assumes that the dissymmetrical input signal does not breach practical constraints (notably it may drive the system away from the desired working point).

REFERENCES

Bai, E.W. and Reyland Jr, J. (2008). Towards identification of Wiener systems with the least amount of a priori information on the nonlinearity. *Automatica*, 44(4), 910–919.

Billings, S.A. and Fakhouri, S.Y. (1982). Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18(1), 15–26.

Chen, T., Ohlsson, H., and Ljung, L. (2012). On the estimation of transfer functions, regularizations and gaussian processes–revisited. *Automatica*, 48, 1525–1535.

Chiuso, A. (2014). System identification techniques: Convexification, regularization, relaxation. In J. Baillieul and T. Samad (eds.), *Encyclopedia of Systems and Control*. Springer.

Doan, T., Litterman, R., and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3, 1–100.

Enqvist, M. and Ljung, L. (2005). Linear approximations of nonlinear FIR systems for separable input processes. *Automatica*, 41(3), 459–473.

Enqvist, M. (2007). A weighting method for approximate nonlinear system identification. In *46th IEEE Conference on Decision and Control (CDC)*. New Orleans.

Greblicki, W. (1992). Nonparametric identification of Wiener systems. *IEEE Transactions on Information Theory*, 38(5), 14871493.

Hu, X.L. and Chen, H.F. (2005). Strong consistence of recursive identification for Wiener systems. *Automatica*, 41(11), 1905 – 1916.

Kitagawa, G. and Gersh, H. (1984). A smothness priorsstate space modeling of time series with trends and seasonalities. *Journal of the American Statistical Association*, 79, 378–389.

Ni, B., Gilson, M., Zhang, Q., and Garnier, H. (2012). A Kalman pre-filtered IV-based approach to continuous-time Hammerstein-Wiener system identification. In *16th IFAC Symposium on System Identification*.

Pawlak, M., Hasiewicz, Z., and Wachel, P. (2007). On nonparametric identification of Wiener systems. *IEEE Transactions on Signal Processing*, 55(2), 482–492.

Pillonetto, G. and De Nicolao, G. (2010). A new kernel-based approach for linear system identification. *Automatica*, 46, 81–93.

Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York.

Voros, J. (2001). Parameter identification of Wiener systems with discontinuous nonlinearities. *Systems and Control Letters*, 44(5), 363–372.

Wills, A., Schön, T., Ljung, L., and Ninness, B. (2013). Identification of Hammerstein–Wiener models. *Automatica*, 49(1), 70–81.

Wong, H.K., Schoukens, J., and Godfrey, K.R. (2012). Analysis of best linear approximation of a Wiener-Hammerstein system for arbitrary amplitude distributions. *IEEE Transactions on Instrumentation and Measurement*, 61(3), 645–654.

Zhang, Q., Iouditski, A., and Ljung, L. (2006). Identification of Wiener system with monotonous nonlinearity. In *14th IFAC Symposium on System Identification (SYSID)*. Newcastle, Australia.

Zhang, Q. and Laurain, V. (2014). Wiener system identification by weighted principal component analysis. In *European Control Conference*. Strasbourg, France.

Zhu, Y. (1999). Distillation column identification for control using Wiener model. In *American Control Conference*, 3462–3466. San Diego.