ON NONPARAMETRIC IDENTIFICATION OF WIENER SYSTEMS WITH DETERMINISTIC INPUTS

Simone Urbano(1)(4), *Eric Chaumette*(2)(4), *Philippe Goupil*(1) *and Jean-Yves Tourneret*(3)(4)

(1) Airbus Flight Control System Department, Airbus, Toulouse, France ([simone.urbano,philippe.goupil]@airbus.com)
(2) University of Toulouse/Isae-Supaero, 10 av. Edouard Belin, Toulouse, France (eric.chaumette@isae.fr)
(3) University of Toulouse/INP-ENSEEIHT/IRIT, 2 Rue Charles Camichel, Toulouse, France (Jean-Yves.Tourneret@enseeiht.fr)
(4) Cooperative research laboratory TéSA, 7 Boulevard de la Gare, Toulouse, France

ABSTRACT

The identification of nonlinear Wiener models (NWMs) for deterministic inputs and Gaussian noise is studied. We show that the nonparametric kernel regression estimation of the nonlinearity of a NWM (based on the Nadaraya-Watson kernel estimator) can be formulated as a parametric estimation problem leading to a Gaussian conditional observation model. This property allows us to derive the maximum likelihood estimators of the unknown parameters of the NWM, as well as the associated Cramér-Rao (CR) bounds. We finally derive a CR-like bound on the global mean squared error (MSE) of the estimated nonlinearity of a NWM. Numerical results obtained for a pulse wave input are presented and compared to the ones based on the Nadaraya-Watson kernel estimator.

Index Terms— Wiener model, non-parametric identification, Cramér-Rao bound, Maximum Likelihood Estimator, Mean Square Error.

1. INTRODUCTION

Many nonlinear models such as Wiener and Hammerstein models are composed by a combination of a linear filter and a static nonlinearity (see Fig. 1) . The combination of these very simple structures is known to approximate a wide range of nonlinear processes [1, 2, 3, 4]. In particular, these models become particularly attractive if one considers a general class of nonlinearities that are not assumed to be parametric and smooth, providing better results than a simple polynomial of finite order [5]. It is possible to extend even more their applicability to nonlinear system identification if one assumes a nonparametric model for the static nonlinearity, as introduced in [3][6] for nonlinear Wiener models (NWMs) and extended in [3][7] for noninvertible nonlinearities. A nonparametric identification algorithm was proposed in [7] for NWMs. The convergence of this algorithm relies on the following assumptions: (i) the input signal $\{x_n\}$ is a sequence of i.i.d. random variables with known probability density function (pdf) and finite first and second order moments, (ii) the noise process $\{z_n\}$ is an i.i.d. sequence with zero mean and finite but unknown variance σ_z^2, (iii) the noise $\{z_n\}$ and the input signal $\{x_n\}$ are mutually independent. The above basic assumptions imply that both the interconnecting signal $\{\omega_n\}$[1] and the output signal $\{y_n\}$ are second-order stationary stochastic processes.

However in many applications, the input signal x_n is not a sequence of i.i.d. random variables, but rather a deterministic time series, and the noise sequence $\{z_n\}$ is simply an additive i.i.d. Gaussian noise with zero mean and finite but unknown variance σ_z^2. In this setting, we show that the nonparametric kernel regression estimation of the nonlinear function $g(.)$ proposed in [7], i.e., the Nadaraya-Watson kernel estimator [10], can also be regarded as a parametric estimation problem, which belongs to the Gaussian conditional observation model [8][9]. Indeed, it amounts to estimating a parameter vector $\boldsymbol{\gamma}$ associated with a given nonparametric kernel estimator of the nonlinearity $g(.)$, as well as the weights $\boldsymbol{\lambda}$ associated with the filter relating x_n and ω_n and the unknown noise variance σ_z^2. By using the well-known Slepian-Bangs formula [16], the first contribution of this paper is to derive the deterministic Cramér-Rao (CR) bound (CRB) for the NWM parameters, i.e., $\boldsymbol{\gamma}$, $\boldsymbol{\lambda}$ and σ_z^2. Furthermore, we also derive an asymptotic CR-like bound on the global mean squared error (MSE) of the estimated nonlinearity $g(.;\boldsymbol{\gamma})$ for consistent and locally unbiased estimators of $\boldsymbol{\gamma}$. An interesting property of this bound is its relation with the mean integrated squared error (MISE) criterion introduced in [7]. Since we consider a conditional signal model, the maximum likelihood estimators (MLEs) of the NWM parameters converge to their associated CRBs at high signal-to-noise-ratio (SNR) [17]. Therefore we derive the associated MLEs and compare their performance with the estimators proposed in [7] (based on the Nadaraya-Watson kernel estimator), which are shown to be sub-optimal when the input signal x_n is not stationary.

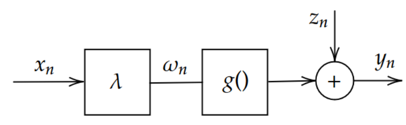

Fig. 1. Nonlinear Wiener model.

2. OBSERVATION MODEL FOR NONPARAMETRIC WIENER SYSTEM

The nonlinear Wiener model shown in Fig. 1 is defined as

$$y_n = g(\omega_n) + z_n,\ \omega_n = \sum_{p=0}^{P} \lambda_p x_{n-p}, \quad 1 \leq n \leq N \tag{1a}$$

where $g(.)$ is an unknown deterministic function of $\Omega \rightarrow \mathbb{R}$, $\Omega \subset \mathbb{R}$, and $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_P) \in \mathbb{R}^{P+1}$ is an unknown deterministic vector. It is important to observe that the pairs $(g(\omega), \boldsymbol{\lambda})$ and $(g(\lambda_0\omega), \boldsymbol{\lambda}/\lambda_0)$ generate the same observations. Indeed, the pair

[1] System identification algorithms assume that the input and output sequences $\{x_n\}$ and $\{y_n\}$ are available. However, the so-called interconnecting signal $\{\omega_n\}$ is not observed.

$(g(\omega), \boldsymbol{\lambda})$ can be identified up to an homothetic transformation affecting $g(.)$. This identifiability problem can be bypassed by assuming $\lambda_0 = 1$, leading to

$$y_n = g(\omega_n) + z_n,\ \omega_n = x_n + \sum_{p=1}^{P} \lambda_p x_{n-p}, \quad 1 \le n \le N \quad \text{(1b)}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_P) \in \mathbb{R}^P$. We introduce the following notations: $\mathbf{y} = (y_1, \ldots, y_N)^T$, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N)^T$, $\mathbf{g}(\boldsymbol{\omega}) = (g(\omega_1), \ldots, g(\omega_N))^T$, $\mathbf{z} = (z_1, \ldots, z_N)^T$, $\mathbf{x} = (x_1, \ldots, x_N)^T$, $\underline{\mathbf{x}} = ((x_{1-P}, \ldots, x_0), \mathbf{x}^T)^T$, and

$$\mathbf{T}_{\underline{\mathbf{x}}} = \begin{bmatrix} x_0 & \ldots & x_{1-P} \\ \vdots & \vdots & \vdots \\ x_{N-1} & \ldots & x_{N-P} \end{bmatrix}$$

where $\mathbf{y}, \boldsymbol{\omega}, \mathbf{g}(\boldsymbol{\omega}), \mathbf{z}, \mathbf{x} \in \mathbb{R}^N$, $\underline{\mathbf{x}} \in \mathbb{R}^{N+P}$, $\mathbf{T}_{\underline{\mathbf{x}}} \in \mathbb{R}^{N \times P}$. The nonparametric kernel regression estimation proposed in [7], based on the Nadaraya-Watson kernel estimator of the nonlinearity $g(.)$ [10], is defined as

$$\widehat{g}(\omega) = \widehat{g}\left(\omega; \widehat{\boldsymbol{\lambda}}\right), \quad \widehat{g}(\omega; \boldsymbol{\lambda}) = \frac{\sum_{i \in \mathcal{I}_1} y_i K_h(\omega - \omega_i(\boldsymbol{\lambda}))}{\sum_{i \in \mathcal{I}_1} K_h(\omega - \omega_i(\boldsymbol{\lambda}))}, \quad \text{(2a)}$$

$$K_h(\omega) = \frac{K(\frac{\omega}{h})}{h}, \widehat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}} \left\{ \sum_{n \in \mathcal{I}_2} (y_n - \widehat{g}(\omega_n(\boldsymbol{\lambda}); \boldsymbol{\lambda}))^2 \right\},$$

where $\omega_j(\boldsymbol{\lambda}) = x_j + \sum_{p=1}^{P} \lambda_p x_{j-p}$, $N = card(\mathcal{I}_1) + card(\mathcal{I}_2) + 2P$ and $K(\omega)$ is a positive symmetric function (kernel) such that

$$\int_{-\infty}^{\infty} K_h(\omega)\, d\omega = \int_{-\infty}^{\infty} K(u)\, du = 1. \quad \text{(2b)}$$

Let $\mathcal{G}_I(\boldsymbol{\gamma})$ be the set of parametric functions $g(.; \boldsymbol{\gamma})$ defined as

$$g(\omega; \boldsymbol{\gamma}) = \frac{\sum_{i=1}^{I} \alpha_i K_h(\omega - \beta_i)}{\sum_{i=1}^{I} K_h(\omega - \beta_i)},\ \boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} \in \mathbb{R}^{2I}. \quad \text{(3)}$$

From a broader perspective, (2a) can also be regarded as an estimator of $g(.; \boldsymbol{\gamma})$ defined in (3) where $I = card(\mathcal{I}_1)$, $\widehat{\alpha}_i = y_i$, $\widehat{\beta}_i = \omega_i\left(\widehat{\boldsymbol{\lambda}}\right)$. Therefore the nonparametric kernel regression estimation of the nonlinearity $g(.)$ defined in (2a) can be recast as a parametric estimation problem.

2.1. Gaussian Conditional Observation Model

The observation model (1b) can be rewritten as follows

$$z_n = y_n - g\left(x_n + \sum_{p=1}^{P} \lambda_p x_{n-p}\right), \quad 1 \le n \le N.$$

If $\underline{\mathbf{x}}$ is a known deterministic vector, the pdf of $\mathbf{y}$ conditionally on $\underline{\mathbf{x}}$ with parameters $\boldsymbol{\lambda} \in \mathbb{R}^P$ is

$$p(\mathbf{y}|\underline{\mathbf{x}}; \boldsymbol{\lambda}) = p_{\mathbf{z}}(\mathbf{y} - \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda})). \quad \text{(4a)}$$

If $p_{\mathbf{z}}(\mathbf{z})$ depends on a vector of unknown deterministic parameters $\boldsymbol{\mu}$, then $p_{\mathbf{z}}(\mathbf{z}) \triangleq p_{\mathbf{z}}(\mathbf{z}; \boldsymbol{\mu})$ and (4a) becomes

$$p(\mathbf{y}|\underline{\mathbf{x}}; \boldsymbol{\lambda}, \boldsymbol{\mu}) = p_{\mathbf{z}}(\mathbf{y} - \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda}); \boldsymbol{\mu}). \quad \text{(4b)}$$

At this point, if $g(.) \triangleq g(.; \boldsymbol{\gamma}) \in \mathcal{G}_I(\boldsymbol{\gamma})$ and if we consider $\boldsymbol{\theta}^T = (\boldsymbol{\mu}^T, \boldsymbol{\lambda}^T, \boldsymbol{\gamma}^T)$, then (4b) becomes

$$p(\mathbf{y}|\underline{\mathbf{x}}; \boldsymbol{\theta}) = p_{\mathbf{z}}(\mathbf{y} - \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda}; \boldsymbol{\gamma}); \boldsymbol{\mu}) \quad \text{(4c)}$$

where $g(.; \boldsymbol{\gamma})$ is an unknown parametric deterministic function. Finally, if $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma_z^2 \mathbf{I}_N)$ then (4c) is a Gaussian pdf as well and thus (1b) defines a Gaussian conditional observation model.

3. DETERMINISTIC CRAMÉR-RAO BOUNDS FOR A NONPARAMETRIC WIENER SYSTEM

The general theory about lower bounds on the MSE of estimators of deterministic parameters is detailed in [12, Section II & III][13] (and summarized in [14, Section II]). In particular, if $\underline{\mathbf{x}}$ is a known deterministic vector, the inverse CRB of $\boldsymbol{\theta}$ is [16]

$$\mathbf{CRB}_{\boldsymbol{\theta}}^{-1}(\underline{\mathbf{x}}) = \mathbf{F}_{\boldsymbol{\theta}}(\underline{\mathbf{x}}) = -E_{\mathbf{y}|\underline{\mathbf{x}};\boldsymbol{\theta}}\left[\frac{\partial^2 \ln p(\mathbf{y}|\underline{\mathbf{x}}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right] \quad \text{(5a)}$$

where $\mathbf{F}_{\boldsymbol{\theta}}(\underline{\mathbf{x}})$ is the Fisher information matrix (FIM). Under the hypothesis that $\mathbf{y} \triangleq \mathbf{y}|\underline{\mathbf{x}} \sim \mathcal{N}(\mathbf{m}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta}))$, the FIM (5a) is obtained from the Slepian-Bangs formula [16, (3.31)]

$$(\mathbf{F}_{\boldsymbol{\theta}})_{i,j} = \frac{\partial \mathbf{m}(\boldsymbol{\theta})^T}{\partial \theta_i} \mathbf{C}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \theta_j} + \frac{1}{2} tr\left(\mathbf{C}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{C}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_j}\right). \quad \text{(6)}$$

In the Gaussian case considered in this work, $\boldsymbol{\theta}^T = (\sigma_z^2, \boldsymbol{\lambda}^T, \boldsymbol{\gamma}^T)$, $\mathbf{C}(\boldsymbol{\theta}) = \sigma_z^2 \mathbf{I}_N$ and $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda}; \boldsymbol{\gamma})$. As a consequence, the FIM of $\boldsymbol{\theta}$ is

$$\mathbf{F}_{\boldsymbol{\theta}}(\underline{\mathbf{x}}) = \begin{bmatrix} \frac{1}{2}\frac{N}{\sigma_z^4} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{\boldsymbol{\lambda}}(\underline{\mathbf{x}}) & \mathbf{F}_{\boldsymbol{\lambda},\boldsymbol{\gamma}}(\underline{\mathbf{x}}) \\ \mathbf{0} & \mathbf{F}_{\boldsymbol{\lambda},\boldsymbol{\gamma}}^T(\underline{\mathbf{x}}) & \mathbf{F}_{\boldsymbol{\gamma}}(\underline{\mathbf{x}}) \end{bmatrix}$$

$$\mathbf{F}_{\boldsymbol{\lambda}}(\underline{\mathbf{x}}) = \frac{1}{\sigma_z^2}\left(\frac{\partial \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda}; \boldsymbol{\gamma})}{\partial \boldsymbol{\lambda}^T}\right)^T \frac{\partial \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda}; \boldsymbol{\gamma})}{\partial \boldsymbol{\lambda}^T}$$

$$\mathbf{F}_{\boldsymbol{\gamma}}(\underline{\mathbf{x}}) = \frac{1}{\sigma_z^2}\left(\frac{\partial \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda}; \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^T}\right)^T \frac{\partial \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda}; \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^T}$$

$$\mathbf{F}_{\boldsymbol{\lambda},\boldsymbol{\gamma}}(\underline{\mathbf{x}}) = \frac{1}{\sigma_z^2}\left(\frac{\partial \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda}; \boldsymbol{\gamma})}{\partial \boldsymbol{\lambda}^T}\right)^T \frac{\partial \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda}; \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^T}$$

which leads to

$$\begin{aligned} \mathbf{CRB}_{\boldsymbol{\lambda}}^{-1}(\underline{\mathbf{x}}) &= \mathbf{F}_{\boldsymbol{\lambda}}(\underline{\mathbf{x}}) - \mathbf{F}_{\boldsymbol{\lambda},\boldsymbol{\gamma}}(\underline{\mathbf{x}}) \mathbf{F}_{\boldsymbol{\gamma}}^{-1}(\underline{\mathbf{x}}) \mathbf{F}_{\boldsymbol{\lambda},\boldsymbol{\gamma}}^T(\underline{\mathbf{x}}) \\ \mathbf{CRB}_{\boldsymbol{\gamma}}^{-1}(\underline{\mathbf{x}}) &= \mathbf{F}_{\boldsymbol{\gamma}}(\underline{\mathbf{x}}) - \mathbf{F}_{\boldsymbol{\lambda},\boldsymbol{\gamma}}^T(\underline{\mathbf{x}}) \mathbf{F}_{\boldsymbol{\lambda}}^{-1}(\underline{\mathbf{x}}) \mathbf{F}_{\boldsymbol{\lambda},\boldsymbol{\gamma}}(\underline{\mathbf{x}}). \end{aligned} \quad \text{(7)}$$

With a few additional computations, it is easy to show that

$$\frac{\partial \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda}; \boldsymbol{\gamma})}{\partial \boldsymbol{\lambda}^T} = \left(\frac{\partial \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda}; \boldsymbol{\gamma})}{\partial \omega} \mathbf{1}_P^T\right) \odot \mathbf{T}_{\underline{\mathbf{x}}}$$

$$\frac{\partial g(\omega; \boldsymbol{\gamma})}{\partial \alpha_{i'}} = \frac{K_h(\omega - \beta_{i'})}{\sum_{i=1}^{I} K_h(\omega - \beta_i)}$$

$$\frac{\partial g(\omega; \boldsymbol{\gamma})}{\partial \beta_{i'}} = K_h^{(1)}(\omega - \beta_{i'}) \frac{\sum_{i=1}^{I} (\alpha_i - \alpha_{i'}) K_h(\omega - \beta_i)}{\left(\sum_{i=1}^{I} K_h(\omega - \beta_i)\right)^2}$$

where $\odot$ denotes the Hadamard product, $\mathbf{1}_P^T = (1, \ldots, 1) \in \mathbb{R}^P$, $K_h^{(1)}(\omega) = \partial K_h(\omega)/\partial\omega$ and

$$\frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\omega} = \left(\sum_{i=1}^{I} \alpha_i K_h^{(1)}(\omega-\beta_i)\right) / \left(\sum_{i=1}^{I} K_h(\omega-\beta_i)\right) - \left(\sum_{i=1}^{I} K_h^{(1)}(\omega-\beta_i)\right) / \left(\sum_{i=1}^{I} K_h(\omega-\beta_i)\right) g(\omega;\boldsymbol{\gamma}).$$

4. A LOWER BOUND ON THE GLOBAL ESTIMATION ERROR

The quality of the estimation of $g(.;\boldsymbol{\gamma}) \in \mathcal{G}_I(\boldsymbol{\gamma})$ based on the estimator $g(.;\widehat{\boldsymbol{\gamma}})$ can be measured via the global estimation error

$$\|g(.;\boldsymbol{\gamma}) - g(.;\widehat{\boldsymbol{\gamma}})\|^2 = \int_\Omega (g(\omega;\boldsymbol{\gamma}) - g(\omega;\widehat{\boldsymbol{\gamma}}))^2 d\omega. \quad (8)$$

From a theoretical point of view, (8) is a random variable whose distribution is difficult to determine in the general case. As a consequence, we consider a simpler performance criterion, i.e., its mean value which equals the global MSE defined as

$$\begin{aligned}\mathcal{C}(\boldsymbol{\gamma},\widehat{\boldsymbol{\gamma}}) &= E_{\mathbf{y}|\underline{\mathbf{x}};\boldsymbol{\theta}}\left[\|g(.;\boldsymbol{\gamma}) - g(.;\widehat{\boldsymbol{\gamma}})\|^2\right] \\ &= \int_\Omega E_{\mathbf{y}|\underline{\mathbf{x}};\boldsymbol{\theta}}\left[(g(\omega;\boldsymbol{\gamma}) - g(\omega;\widehat{\boldsymbol{\gamma}}))^2\right] d\omega. \quad (9)\end{aligned}$$

It is interesting to note that $\mathcal{C}(\boldsymbol{\gamma},\widehat{\boldsymbol{\gamma}})$ in (9) is the limiting value for $T, L \to \infty$ of the MISE performance criterion [7, (28)] (weak law of large numbers)

$$MISE(\widehat{g}(.)) = \frac{1}{LT}\sum_{l=1}^{L} \|\mathbf{g}(\boldsymbol{\omega}_T;\boldsymbol{\gamma}) - \mathbf{g}(\boldsymbol{\omega}_T;\widehat{\boldsymbol{\gamma}}_l)\|^2 \quad (10)$$

where L is the number of independent observations, $\Omega = [a,b]$, $\omega_t = a + \frac{b-a}{T}(t-1)$ is the compact interval containing the possible values of ω, and $\mathbf{g}(\boldsymbol{\omega}_T;\boldsymbol{\gamma}') = (g(\omega_1;\boldsymbol{\gamma}'), \ldots, g(\omega_T;\boldsymbol{\gamma}'))^T$. Under the assumption that $\widehat{\boldsymbol{\gamma}} \triangleq \widehat{\boldsymbol{\gamma}}(\mathbf{y}|\underline{\mathbf{x}})$ is a consistent estimator of $\boldsymbol{\gamma}$, i.e., provided that $\widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + d\widehat{\boldsymbol{\gamma}}$ with $d\widehat{\boldsymbol{\gamma}}^T d\widehat{\boldsymbol{\gamma}} \to 0$ when $\sigma_{\mathbf{z}}^2 \to 0$, then $g(\omega;\widehat{\boldsymbol{\gamma}}) - g(\omega;\boldsymbol{\gamma}) \to \frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T} d\widehat{\boldsymbol{\gamma}}$ when $\sigma_{\mathbf{z}}^2 \to 0$ leading to:

$$\begin{aligned}\mathcal{C}(\boldsymbol{\gamma},\widehat{\boldsymbol{\gamma}}) &\underset{\sigma_{\mathbf{z}}^2\to 0}{\longrightarrow} \int_\Omega \frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T} \mathbf{C}_{d\widehat{\boldsymbol{\gamma}}}(\underline{\mathbf{x}}) \frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}} d\omega \\ &= tr\left(\mathbf{C}_{d\widehat{\boldsymbol{\gamma}}}(\underline{\mathbf{x}}) \int_\Omega \frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}} \frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T} d\omega\right).\end{aligned}$$

Moreover, if $\widehat{\boldsymbol{\gamma}}$ is a locally unbiased estimator of $\boldsymbol{\gamma}$, then $\mathbf{C}_{d\widehat{\boldsymbol{\gamma}}}(\underline{\mathbf{x}}) \geq \mathbf{CRB}_{\boldsymbol{\gamma}}(\underline{\mathbf{x}})$ [16] (in the sense that the difference between the two matrices is positive) and

$$\frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T} \mathbf{C}_{d\widehat{\boldsymbol{\gamma}}}(\underline{\mathbf{x}}) \frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}} \geq \frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T} \mathbf{CRB}_{\boldsymbol{\gamma}}(\underline{\mathbf{x}}) \frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}$$

which allows us to define the following CR-like bound

$$\mathcal{C}(\boldsymbol{\gamma},\widehat{\boldsymbol{\gamma}}) \geq tr\left(\mathbf{CRB}_{\boldsymbol{\gamma}}(\underline{\mathbf{x}}) \int_\Omega \frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}} \frac{\partial g(\omega;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T} d\omega\right). \quad (11)$$

5. AN MLE FOR NONPARAMETRIC WIENER SYSTEMS

When $g(.)$ is an unknown parametric deterministic function, i.e., $g(.) \triangleq g(.;\boldsymbol{\gamma}) \in \mathcal{G}_I(\boldsymbol{\gamma})$, the analysis can be conducted by rewriting (1b) as

$$y_n = \sum_{i'=1}^{I} \frac{K_h(\omega_n(\boldsymbol{\lambda}) - \beta_{i'})}{\sum_{i=1}^{I} K_h(\omega_n(\boldsymbol{\lambda}) - \beta_i)} \alpha_{i'} + z_n$$

which leads to the well known conditional Gaussian linear model [8][9][16]

$$\mathbf{y} = \mathbf{H}_{\underline{\mathbf{x}}}(\boldsymbol{\beta},\boldsymbol{\lambda})\boldsymbol{\alpha} + \mathbf{z}, \quad \mathbf{H}_{\underline{\mathbf{x}}}(\boldsymbol{\beta},\boldsymbol{\lambda}) = \frac{\partial \mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda};\boldsymbol{\gamma})}{\partial\boldsymbol{\alpha}^T}, \quad (12)$$

for which the MLE of $\boldsymbol{\theta}^T = (\sigma_z^2, \boldsymbol{\lambda}^T, \boldsymbol{\gamma}^T)$ is

$$\begin{aligned}\widehat{\sigma_z^2}(\mathbf{y}|\underline{\mathbf{x}}) &= \frac{1}{N}\left\|\mathbf{y} - \mathbf{H}_{\underline{\mathbf{x}}}\left(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\lambda}}\right)\widehat{\boldsymbol{\alpha}}\right\|^2 \\ \left(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\lambda}}\right)(\mathbf{y}|\underline{\mathbf{x}}) &= \arg\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\lambda}}\left\{\frac{1}{N}\|\mathbf{y} - \mathbf{H}_{\underline{\mathbf{x}}}(\boldsymbol{\beta},\boldsymbol{\lambda})\boldsymbol{\alpha}\|^2\right\}.\end{aligned}$$

Straightforward computations lead to [8][9][16]:

$$\widehat{\boldsymbol{\alpha}}(\mathbf{y}|\underline{\mathbf{x}}) = \left(\mathbf{H}_{\underline{\mathbf{x}}}(\boldsymbol{\beta},\boldsymbol{\lambda})^T \mathbf{H}_{\underline{\mathbf{x}}}(\boldsymbol{\beta},\boldsymbol{\lambda})\right)^{-1} \mathbf{H}_{\underline{\mathbf{x}}}(\boldsymbol{\beta},\boldsymbol{\lambda})^T \mathbf{y} \quad \text{(13a)}$$

$$\left(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\lambda}}\right)(\mathbf{y}|\underline{\mathbf{x}}) = \arg\max_{\boldsymbol{\beta},\boldsymbol{\lambda}}\left\{\mathbf{y}^T \boldsymbol{\Pi}_{\mathbf{H}_{\underline{\mathbf{x}}}(\boldsymbol{\beta},\boldsymbol{\lambda})}\mathbf{y}\right\} \quad \text{(13b)}$$

where $\boldsymbol{\Pi}_{\mathbf{A}} = \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T$. We can observe that the MLE of $\boldsymbol{\alpha}$ (13a) is different from the "Nadaraya-Watson kernel estimator" (2a) [7, (11)]. In [17] it is shown that when $\sigma_{\mathbf{z}}^2 \to 0$, the MLEs $\left(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\lambda}}\right)(\mathbf{y}|\underline{\mathbf{x}})$ (13a-13b) are consistent, Gaussian, locally unbiased and efficient (minimum variance). As a consequence, when $\sigma_{\mathbf{z}}^2 \to 0$, for a given pair $\left(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\lambda}}\right)(\mathbf{y}|\underline{\mathbf{x}})$, (2a)[7, (11)] leads likely to a biased estimator and sub-optimal (in the MSE sense) compared to the MLE (13a). In a nutshell, the following results can be obtained asymptotically (when $\sigma_{\mathbf{z}}^2 \to 0$): (i) the proposed MLEs $\left(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\lambda}}\right)(\mathbf{y}|\underline{\mathbf{x}})$ are efficient; (ii) $g(.;\widehat{\boldsymbol{\gamma}}(\mathbf{y}|\underline{\mathbf{x}}))$ reaches (11).

6. RESULTS

We consider a synthetic scenario based on a pulse wave input $\mathbf{x}$ as displayed in Fig. 2 ($N = 100$), and a dynamical system defined by $\boldsymbol{\lambda} = (1/2, 1/2)^T$, $\boldsymbol{\alpha} = (6, -2)^T$, $\boldsymbol{\beta} = (-1/4, 1/4)^T$. A Gaussian kernel with bandwidth $h = 1$ is considered. The nonlinearity $g(.)$ resulting from this choice is shown in Fig. 2, where $\Omega = [a,b] = [-20, 20]$ and $T = 800$. Note that all the results presented in this paper have been obtained by averaging $L = 5000$ Monte Carlo runs. In Fig. 3 and 4 we compare the MSE of the MLEs (13a-13b) to the CRBs (7) as a function of the SNR defined as $SNR = \left(\frac{1}{N}\|\mathbf{g}(\mathbf{x} + \mathbf{T}_{\underline{\mathbf{x}}}\boldsymbol{\lambda};\boldsymbol{\gamma})\|^2\right)/\sigma_z^2$. Fig. 3 also compares the performance of two estimators of $\boldsymbol{\lambda}$, i.e., the MLE defined in (13b) and Pawlak's estimator defined in (2a) where $card(\mathcal{I}_1) = 51$ and $card(\mathcal{I}_2) = 47$. We can observe that the MLEs (13a-13b) converge to the CRBs (7) when the SNR increases, as expected [17]. Moreover, we can note that the MLE outperforms the kernel estimator of $g(.)$ proposed in [7]. Fig. 5 displays the estimated global estimation error, i.e. $MISE(\widehat{g}(.))$, of the two estimators versus SNR, which is compared with the proposed CR-like bound (11). As already mentioned for the estimation of $\boldsymbol{\lambda}$, the global estimation error of the MLE converges to the bound and outperforms the kernel estimator(2a)[7], which can also be observed in Fig. 6 showing the estimator of the nonlinearity $g(.)$ in both cases (for a given SNR), with a biased kernel estimator, as anticipated.

7. CONCLUSIONS

This paper addressed the nonlinear system identification problem for nonparametric Wiener models. The deterministic CRBs, the MLE

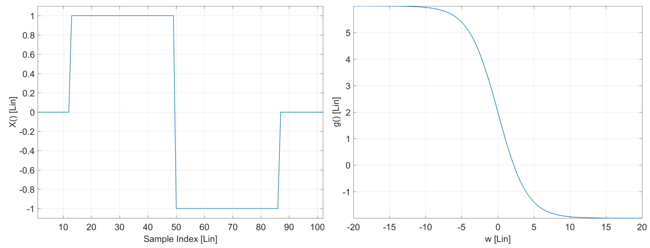

Fig. 2. Input signal x (left) and non linearity $g(.)$ (right)

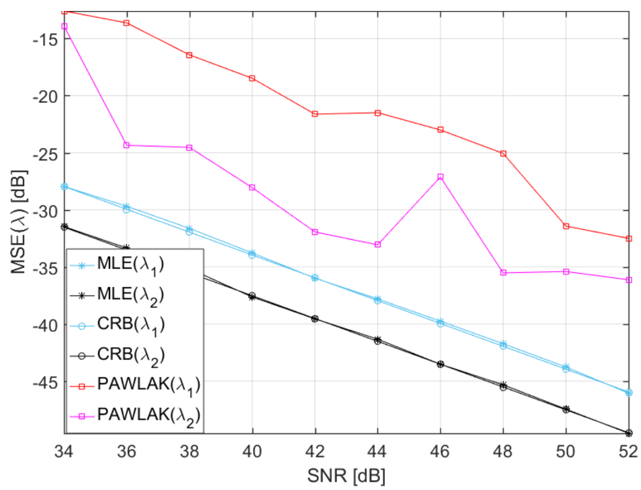

Fig. 3. MSEs of the MLE (13b) and of Pawlak's estimator (2a) for $\boldsymbol{\lambda}$ versus SNR, and the corresponding $\mathbf{CRB}(\boldsymbol{\lambda})$ (7).

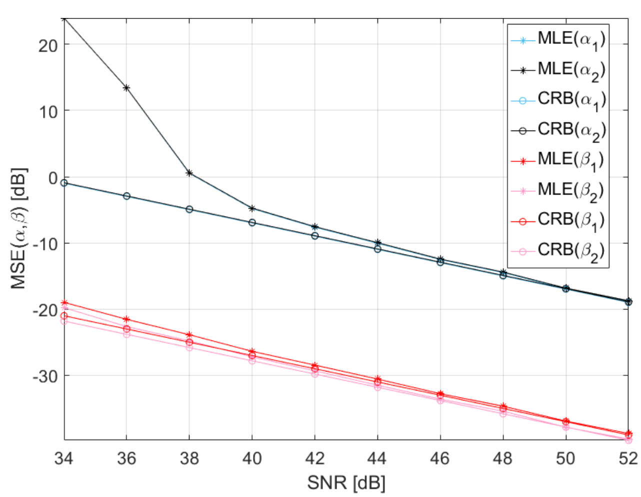

Fig. 4. MSE of the MLEs of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ (13a-13b) versus SNR, and the corresponding $\mathbf{CRB}\,(\boldsymbol{\alpha}, \boldsymbol{\beta})$ (7).

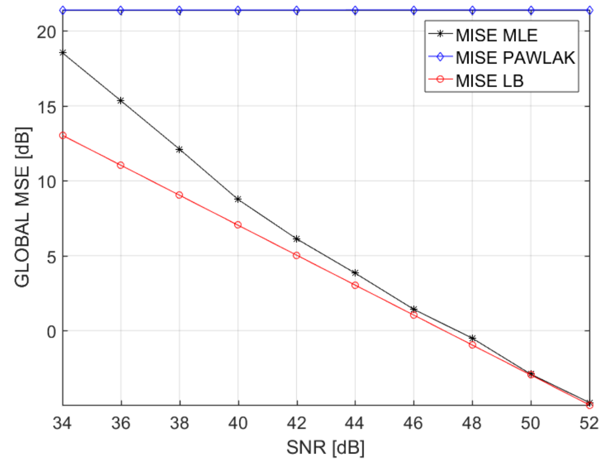

Fig. 5. $MISE(\widehat{g}\,(.))$ (10) of the MLE (13a-13b) and of Pawlak's estimator (2a) versus SNR, compared with the lower bound (11).

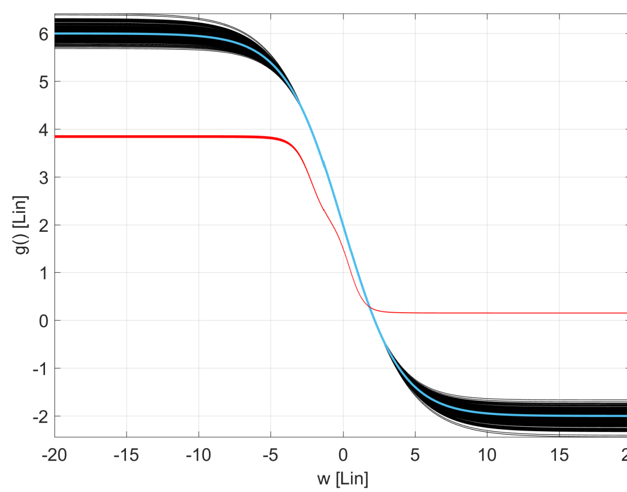

Fig. 6. Estimated nonlinearity $\widehat{g}\,(.)$ obtained with the MLEs (13a-13b, in black) and Pawlak's estimators (2a, in red) of $(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta})$, compared to the ground truth $g\,(.)$ (in blue) at $SNR = 52dB$.

and an asymptotic CR-like bound for the global estimation error of the estimated nonlinearity were derived for this problem. Some simulation results confirmed that the maximum likelihood estimator of the nonlinearity has a global estimation error closer to the corresponding Cramér-Rao bound than an existing kernel estimator, which was designed for i.i.d. random input signals [7]. Based on the obtained results, further studies can be carried out to evaluate the optimal input signal for Wiener system identification and the influence of the bandwidth parameter h and/or the kernel type on the MLE performance.

8. REFERENCES

[1] G. B. Giannakis and E. Serpedin, "A bibliography on nonlinear system identification," Signal Process., vol. 81, pp. 533-580, 2001.

[2] O. Nelles, *Nonlinear System Identification*, New York:Springer-Verlag, 2001.

[3] W. Greblicki and M. Pawlak, *Nonparametric system identification*, Cambridge university press, 2008

[4] F. Giri and E-W. Bai, *Block-oriented nonlinear system identification*, vol. 1, Springer, 2010

[5] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*, New York: Wiley, 2000

[6] W. Greblicki, "Nonparametric identification of Wiener systems," IEEE Trans. on IT, 38(5): 1487-1493, 1992

[7] M. Pawlak, Z. Hasiewicz, and P. Wachel, "On Nonparametric Identification of Wiener Systems", IEEE Trans. on SP, 55(2): 482-492, 2007

[8] P. Stoica and A. Nehorai, "Performances study of conditional and unconditional direction of arrival estimation," IEEE Trans. on ASSP, 38(10): 1783-1795, 1990

[9] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing," in Radar Array Processing, S. Haykin, J. Litva, and T. J. Shepherd, Eds., chapter 4, pp. 99-151. Springer-Verlag, Heidelberg, 1993.

[10] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London, U.K.: Chapman & Hall, 1995

[11] A. Quinlan, E. Chaumette, P. Larzabal, "A Direct Method to Generate Approximations of the Barankin Bound", in Proc. IEEE ICASSP, 2006

[12] E. Chaumette, J. Galy, A. Quinlan, P. Larzabal, "A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum Likelihood Estimators", IEEE Trans. on SP, 56(11): 5319-5333, 2008

[13] K. Todros and J. Tabrikian, "General Classes of Performance Lower Bounds for Parameter Estimation-Part I: Non-Bayesian Bounds for Unbiased Estimators", IEEE Trans. on IT, 56(10): 5064-5082, 2010

[14] N. Kbayer, J. Galy, E. Chaumette, F. Vincent, A. Renaux and P. Larzabal, "On Lower Bounds for Non-Standard Deterministic Estimation", IEEE Trans. on SP, 65(6): 1538-1553, 2017

[15] E. Chaumette, F. Vincent, J. Galy, P. Larzabal, "On the influence of detection tests on deterministic parameters estimation", in Proc. Eurasip EUSIPCO, 2006

[16] S.M. Kay, *Fundamentals of Statistical Signal Processing: estimation theory*, Prentice-Hall, 1993

[17] A. Renaux, P. Forster, E. Chaumette, P. Larzabal, "On the High-SNR Conditional Maximum-Likelihood Estimator Full Statistical Characterization", IEEE Trans. on SP, 54(12): 4840-4843, 2006

[18] S. Urbano, E. Chaumette, P. Goupil and J.Y. Tourneret, "On the high-SNR ROC of GLRT for the conditional signal model", in Proc. IEEE ICASSP, 2018