Optimal enforcement of causality in non-parametric transfer function estimation

Rodrigo A. González, Patricio E. Valenzuela, Cristian R. Rojas and Ricardo A. Rojas

***Abstract*— Traditionally, non-parametric impulse and frequency response functions are estimated by taking the ratio of power spectral density estimates. However, this approach may often lead to non-causal estimates. In this paper we derive a closed form expression for the impulse response estimator by smoothed empirical transfer function estimate (ETFE), which allows optimal enforcement of causality on non-parametric estimators based on spectral analysis. The new method is shown to be asymptotically unbiased and of minimum covariance among a broad class of linear estimators. Numerical simulations illustrate the performance of the new estimator.**

***Index Terms*— System identification, non-parametric estimation, spectral analysis, ETFE, causality.**

I. Introduction

The study of non-parametric methods for the estimation of dynamic systems has a long history [1], [2]. Its importance lies in that they can be used to verify the quality of an identification experiment, to get some insight about the dynamics of the system, and to validate an estimated parametric model. Among the non-parametric techniques commonly employed in practice, one of the most popular is spectral analysis [3], [4].

By spectral analysis, an empirical transfer function estimate (ETFE) is obtained from the ratio of power spectral density (PSD) estimates of its input and output time sequences [5]. Unfortunately this estimator is not consistent; i.e., the transfer function estimates do not converge in probability, because the variance of the estimate does not converge to zero as the number of data points tend to infinity [6], [7]. Therefore, the ETFE is usually modified by using non-rectangular windows in the time domain to obtain better PSD estimates [4], leading to the smoothed ETFE.

Recently, several new methods have been introduced to enhance the frequency response function estimates. As a way to increase the frequency resolution, the auto tuning of the window width has been studied in [8]. Other recently developed techniques to improve the statistical properties of the ETFE include the local polynomial method (LPM) [9], the local rational method (LRM) [10], and the transient impulse response modeling method (TRIMM) [11], [12]. These methods reduce the leakage effects in the non-parametric estimation of the frequency response function.

When estimating systems, it is common to know in advance some of their basic properties. A fundamental property of every physical system is *causality*, which means that effects cannot precede causes. However, most non-parametric estimators by spectral analysis do not explicitly enforce this important property to improve their performance.

In this paper, we introduce a method that optimally imposes causality in non-parametric estimation by spectral analysis. Based on Cholesky whitening [13], [14], we obtain an impulse response estimator that is asymptotically unbiased and of minimum covariance matrix among a broad class of linear impulse response estimators based on the smoothed ETFE. Numerical examples show that ETFE produces non-causal estimates of the impulse response, and that the new method ensures causality, while improving the statistical performance of the smoothed ETFE.

The remainder of this paper is organized as follows. Section II presents the spectral analysis method. Section III derives a closed form expression for the impulse response estimate of the smoothed ETFE. In Section IV, we analyze the mean and covariance matrix of the estimator described in the previous section, propose an optimal causal impulse response estimate, and prove some of its statistical properties. Section V illustrates the method with numerical examples and Section VI concludes this article.

Notation: $\mathbb{Z}$, $\mathbb{R}$ and $\mathbb{R}^{r\times s}$ denote the integer set, real set and the set of $r \times s$ matrices with real entries, respectively. Bold letters are used to denote vectors and matrices. $\mathbf{0}_N$ and $\mathbf{I}_N$ denote the null and identity matrix respectively, both of size $N \times N$. $\mathbf{A} \succeq \mathbf{B}$ means that the matrix $\mathbf{A} - \mathbf{B}$ is positive semi-definite, and $\mathrm{E}\{Y\}$ denotes the mean value of the random variable Y.

II. Non-parametric estimation by spectral analysis

Consider the discrete-time, single-input, single-output, stable, causal, linear time-invariant system described by

$$y_t = \sum_{k=-\infty}^{\infty} g_k^0 u_{t-k} + v_t, \tag{1}$$

where $\{g_k^0\}_{k\in\mathbb{Z}}$ denotes the impulse response sequence, $y_t \in \mathbb{R}$ is the measured output, $u_t \in \mathbb{R}$ is the input, and v_t is a zero-mean white noise of variance σ^2. We recall that a linear time-invariant system with impulse response $\{g_k^0\}_{k\in\mathbb{Z}}$

*The first author thanks the support of CONICYT, Chile. This work was supported by the Swedish Research Council under contract 2015-04393.

R. A. González and R. A. Rojas are with the Electronic Engineering Department, Universidad Técnica Federico Santa María (UTFSM), 2390123, Valparaíso, Chile (e-mails: rodrigo.gonzalezvi@alumnos.usm.cl; ricardo.rojas@usm.cl).

P. E. Valenzuela and C. R. Rojas are with the Department of Automatic Control and ACCESS Linnaeus Centre, KTH Royal Institute of Technology, 10044, Stockholm, Sweden (e-mails: pva@kth.se; crro@kth.se).

is *causal* if $g_k^0 = 0$ for all $k < 0$ [15]. The frequency response of the system (1) is defined as

$$G_0(e^{j\omega}) = \sum_{k=0}^{\infty} g_k^0 e^{-j\omega k}, \quad \omega \in [-\pi, \pi).$$

Based on a deterministic excitation signal $\{u_t\}_{t=1}^N$ and the measured output signal $\{y_t\}_{t=1}^N$, our objective is to build, by applying spectral analysis techniques, a non-parametric estimator of $\{g_k^0\}$ that is causal for every input sequence. Using the discrete Fourier transform (DFT)

$$X_N(e^{j\omega}) = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} x_t e^{-j\omega t},$$

we define the ETFE as the ratio of Fourier transforms of the output and input sequences:

$$\hat{G}_N(e^{j\omega}) := \frac{Y_N(e^{j\omega})}{U_N(e^{j\omega})}. \tag{2}$$

An exact frequency-domain formulation of (1) can be obtained, and it is given by

$$Y_N(e^{j\omega}) = G_0(e^{j\omega})U_N(e^{j\omega}) + V_N(e^{j\omega}) + M_N(e^{j\omega}), \tag{3}$$

where $Y_N(e^{j\omega})$, $U_N(e^{j\omega})$ and $V_N(e^{j\omega})$ are the DFTs of $\{y_t\}_{t=1}^N$, $\{u_t\}_{t=1}^N$ and $\{v_t\}_{t=1}^N$ respectively, and $M_N(e^{j\omega})$ is a leakage term that, under the stability condition $\sum_{k=1}^{\infty} |kg_k^0| < \infty$ of system (1) and the boundedness of $\{u_t\}$, satisfies $\lim_{N\to\infty} |M_N(e^{j\omega})| = 0$ for all $\omega \in [-\pi, \pi)$ [6]. This term is given by

$$M_N(e^{j\omega}) = \sum_{k=0}^{\infty} g_k^0 e^{-j\omega k} \times \left(\frac{1}{\sqrt{N}} \sum_{\tau=1-k}^{N-k} u_\tau e^{-j\omega\tau} - U_N(e^{j\omega}) \right).$$

A straightforward approach to obtain an impulse response estimate using (3) is to compute the inverse Fourier transform of the ETFE. The variance of the ETFE does not decrease with N [6], and therefore it may give poor estimates. A more commonly used approach is to estimate the frequency response of the system as the ratio of the smoothed spectral estimates [4]:

$$\hat{G}_N^s(e^{j\omega}) := \frac{\hat{\Phi}_{yu}^N(\omega)}{\hat{\Phi}_u^N(\omega)} = \frac{\int_{-\pi}^{\pi} W_\gamma(\xi-\omega) Y_N(e^{j\xi}) \overline{U}_N(e^{j\xi}) d\xi}{\int_{-\pi}^{\pi} W_\gamma(\xi-\omega) |U_N(e^{j\xi})|^2 d\xi}, \tag{4}$$

where $\overline{U}_N(e^{j\xi})$ denotes the complex conjugate of $U_N(e^{j\xi})$, and $W_\gamma(\xi)$ is an even weighting function in the frequency domain. Here, $\gamma > 0$ is the width of $w_\gamma(\tau)$, the inverse Fourier transform of $W_\gamma(\xi)$. By construction, $w_\gamma(\tau) = 0$ for $|\tau| > \gamma$. We recall that for the estimator (4) to be consistent, we must have $\gamma \to \infty$ but $\lim_{N\to\infty} \gamma/N = 0$. The optimal choice of the window width, as $N \to \infty$, is $\gamma = CN^{\frac{1}{5}}$, where C depends on several unknown quantities of the system [4, Section 6.4].

The problem treated in this paper is how to optimally impose causality on an estimator based on the smoothed version of the ETFE (4).

III. Impulse response estimate by smoothed ETFE

Consider the system (1) and its frequency domain equivalent (3) written in terms of $e^{j\xi}$. Multiplying (3) by $\overline{U}_N(e^{j\xi}) W_\gamma(\xi - \omega)$, integrating over $\xi \in [-\pi, \pi]$, and finally dividing by the denominator of (4), we obtain the smoothed frequency response estimate

$$\hat{G}_N^s(e^{j\omega}) = G_0(e^{j\omega}) + G_v(e^{j\omega}) + G_{bias}(e^{j\omega}), \tag{5}$$

where

$$G_v(e^{j\omega}) = \frac{\int_{-\pi}^{\pi} W_\gamma(\xi-\omega) V_N(e^{j\xi}) \overline{U}_N(e^{j\xi}) d\xi}{\int_{-\pi}^{\pi} W_\gamma(\xi-\omega) |U_N(e^{j\xi})|^2 d\xi}, \tag{6}$$

$$G_{bias}(e^{j\omega}) = \frac{\int_{-\pi}^{\pi} W_\gamma(\xi-\omega) M_N(e^{j\xi}) \overline{U}_N(e^{j\xi}) d\xi}{\int_{-\pi}^{\pi} W_\gamma(\xi-\omega) |U_N(e^{j\xi})|^2 d\xi} + \frac{\int_{-\pi}^{\pi} W_\gamma(\xi-\omega) (G_0(e^{j\xi}) - G_0(e^{j\omega})) |U_N(e^{j\xi})|^2 d\xi}{\int_{-\pi}^{\pi} W_\gamma(\xi-\omega) |U_N(e^{j\xi})|^2 d\xi}. \tag{7}$$

Note that from (5)-(7) the smoothed ETFE is biased, with bias given by $G_{bias}(e^{j\omega})$. The covariance of $\hat{G}_N^s(e^{j\omega})$ depends exclusively on $G_v(e^{j\omega})$, which is dependent on the Fourier transforms of the input and noise sequence. This term will be studied in the rest of this section. Since (6) is the ratio of the cross spectral density estimate $\hat{\Phi}_{vu}^N(\omega)$ and the PSD estimate $\hat{\Phi}_u^N(\omega)$,

$$\begin{aligned} G_v(z) &= \frac{\hat{\Phi}_{vu}^N(z)}{\hat{\Phi}_u^N(z)} = \frac{\sum_{k=-\gamma}^{\gamma} w_\gamma(k) \hat{R}_{vu}(k) z^{-k}}{\sum_{k=-\gamma}^{\gamma} w_\gamma(k) \hat{R}_u(k) z^{-k}} \\ &= \sum_{k=-\gamma}^{\gamma} w_\gamma(k) \hat{R}_{vu}(k) z^{-k} \frac{z^\gamma}{N_{\tilde{U}}(z)} \\ &= \sum_{k=-\gamma}^{\gamma} w_\gamma(k) \hat{R}_{vu}(k) z^{-k} \sum_{\tau=-\infty}^{\infty} \beta_\tau z^{-\tau}, \end{aligned}$$

where

$$\hat{R}_{vu}(k) := \begin{cases} \frac{1}{N} \sum_{t=k+1}^{N} v_t u_{t-k}, & k \geq 0 \\ \hat{R}_{vu}(-k), & k < 0 \end{cases}$$

$$\hat{R}_u(k) := \begin{cases} \frac{1}{N} \sum_{t=k+1}^{N} u_t u_{t-k}, & k \geq 0 \\ \hat{R}_u(-k), & k < 0 \end{cases}$$

$$N_{\tilde{U}}(z) := \sum_{l=0}^{2\gamma} w_\gamma(\gamma - l) \hat{R}_u(\gamma - l) z^l.$$

Also, $\{\beta_\tau\}$ are the Laurent coefficients of the rational function $z^\gamma (N_{\tilde{U}}(z))^{-1}$ around $z = 0$ [16], in case $N_{\tilde{U}}(z)$ does not have poles on $\{z \in \mathbb{C} : |z| = 1\}$. Therefore, if we denote by $\{g_t^{bias}\}$ and $\{\hat{g}_t^s\}$ the inverse Fourier transforms of $G_{bias}(e^{j\omega})$ and $\hat{G}_N^s(e^{j\omega})$ respectively, the smoothed ETFE can be written as

$$\hat{G}_N^s(e^{j\omega}) = \sum_{t=-\infty}^{\infty} \hat{g}_t^s e^{-j\omega t}$$

where

$$\hat{g}_t^s := g_t^0 + \sum_{k=-\gamma}^{\gamma} w_\gamma(k)\hat{R}_{vu}^N(k)\beta_{t-k} + g_t^{bias}. \tag{8}$$

Equation (8) gives an exact expression for $\hat{g}_t^s$, the smoothed ETFE estimator of the impulse response. For the purposes of the equation (8) in this paper, we do not require a closed form expression for the term g_t^{bias}. Note that even though the system (1) is causal, $\hat{g}_t^s$ is generally non-causal (i.e., $\hat{g}_t^s \neq 0$ for $t < 0$) due to both $\{\beta_\tau\}$ and $\{g_t^{bias}\}$.

IV. Causal spectral estimation

In this section we develop a causal estimator for the impulse response $\{g_k^0\}$. To this end, we compute the mean and covariance matrix of the estimator $\{\hat{g}_t^s\}_{t=-M}^N$, where the parameter $M \in \{1, \ldots, N\}$ determines the number of non-causal points being estimated. By (8), the mean value of $\hat{g}_t^s$ is

$$\begin{aligned}\mathrm{E}\{\hat{g}_t^s\} &= \mathrm{E}\left\{ g_t^0 + \sum_{k=-\gamma}^{\gamma} w_\gamma(k)\hat{R}_{vu}^N(k)\beta_{t-k} + g_t^{bias} \right\} \\ &= g_t^0 + g_t^{bias},\end{aligned}$$

where we use the fact that $\mathrm{E}\{\hat{R}_{vu}^N(k)\} = 0$. Note that since the system (1) is causal, $g_t^0 = 0$ for $t < 0$, but this is not necessarily so for g_t^{bias}.

For the covariance matrix of $\{\hat{g}_t^s\}_{t=-M}^N$, we define the estimation error

$$e_t := \hat{g}_t^s - \mathrm{E}\{\hat{g}_t^s\} = \sum_{k=-\gamma}^{\gamma} w_\gamma(k)\hat{R}_{vu}^N(k)\beta_{t-k}.$$

The error vector $\mathbf{e} := [e_{-M} \quad \ldots \quad e_N]^T$ can be written as

$$\mathbf{e} = \frac{1}{N}\mathbf{B}\,\mathrm{diag}(w_\gamma)\,\mathbf{U}\mathbf{v},$$

where $\mathrm{diag}(w_\gamma) \in \mathbb{R}^{(2\gamma+1)\times(2\gamma+1)}$ is a diagonal matrix with $w_\gamma(k-\gamma-1)$ as its (k,k) entry, $\mathbf{B} \in \mathbb{R}^{(N+M+1)\times(2\gamma+1)}$ and $\mathbf{U} \in \mathbb{R}^{(2\gamma+1)\times N}$ are Toeplitz matrices with elements

$$\begin{aligned}[\mathbf{B}]_{i,k} &= \beta_{i-k-M+\gamma}, \qquad (9)\\ [\mathbf{U}]_{i,k} &= \begin{cases} u_{-i+k+\gamma+1}, & -\gamma \leq -i+k \leq N-1-\gamma \\ 0, & \text{otherwise,}\end{cases}\end{aligned}$$

and $\mathbf{v} := [v_1 \quad \ldots \quad v_N]^T$. Hence, the covariance matrix $\mathbf{P}$ can be written as

$$\mathbf{P} = \mathrm{E}\left\{\mathbf{e}\mathbf{e}^T\right\} = \frac{\sigma^2}{N^2}\mathbf{B}\,\mathrm{diag}(w_\gamma)\,\mathbf{U}\mathbf{U}^T\,\mathrm{diag}(w_\gamma)\,\mathbf{B}^T. \tag{10}$$

Note that, if the noise variance σ^2 is known or if it can be estimated [17, Chapter 10], the covariance matrix $\mathbf{P}$ can be computed by forming $\mathbf{B}, \mathrm{diag}(w_\gamma)$, and $\mathbf{U}$ from the window function $w_\gamma(\tau)$ and the input sequence $\{u_t\}_{t=1}^N$. This computation is essential for the new estimator proposed next.

A. Optimal causal smoothed ETFE

Consider the impulse response estimator $\{\hat{g}_t^s\}_{t=-M}^N$ presented in (8), and define $\hat{\mathbf{g}}^s$ as the vector with elements $\{\hat{g}_t^s\}_{t=-M}^N$. We introduce the following causal impulse response estimator by smoothed ETFE:

$$\begin{aligned}\tilde{\mathbf{g}}^{cs} &:= \mathbf{C}\left[\begin{array}{c|c} \mathbf{0}_{M\times M} & \mathbf{0}_{M\times(N+1)} \\ \hline \mathbf{0}_{(N+1)\times M} & \mathbf{I}_{N+1}\end{array}\right]\mathbf{C}^{-1}\hat{\mathbf{g}}^s \qquad (11)\\ &= \left[\begin{array}{c|c} \mathbf{0}_{M\times M} & \mathbf{0}_{M\times(N+1)} \\ \hline -\mathbf{C}_{21}\mathbf{C}_{11}^{-1} & \mathbf{I}_{N+1}\end{array}\right]\hat{\mathbf{g}}^s,\end{aligned}$$

where

$$\mathbf{C} = \left[\begin{array}{c|c} \mathbf{C}_{11} & \mathbf{0}_{M\times(N+1)} \\ \hline \mathbf{C}_{21} & \mathbf{C}_{22}\end{array}\right]$$

is the Cholesky factor of $\mathbf{P}$ [18] (i.e., a lower triangular matrix with positive diagonal entries such that $\mathbf{P} = \mathbf{C}\mathbf{C}^T$), with $\mathbf{C}_{11} \in \mathbb{R}^{M\times M}$, $\mathbf{C}_{21} \in \mathbb{R}^{(N+1)\times M}$, and $\mathbf{C}_{22} \in \mathbb{R}^{(N+1)\times(N+1)}$. Denote the elements of $\tilde{\mathbf{g}}^{cs}$ as $\tilde{g}_t^{cs}, t \in \{-M, \ldots, N\}$.

For the study of some relevant properties of $\tilde{\mathbf{g}}^{cs}$, we assume that γ is chosen proportional to[1] N. For this choice of γ, we focus on improving the variance of $\tilde{\mathbf{g}}^s$.

An estimator that is linearly dependent on $\hat{\mathbf{g}}^s$ is an asymptotically unbiased estimator of $\{g_k^0\}$ if it has the form

$$\tilde{\mathbf{g}} = \mathbf{A}\hat{\mathbf{g}}^s = \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22}\end{array}\right]\hat{\mathbf{g}}^s,$$

where the matrices $\mathbf{A}_{12}$ and $\mathbf{A}_{22}$ converge weakly [19] to $\mathbf{0}_{M\times(N+1)}$ and $\mathbf{I}_{(N+1)\times(N+1)}$ respectively as $N \to \infty$. For Theorem 4.1, we say, with some abuse of notation, that $\tilde{\mathbf{g}}$ is a causal linear asymptotically unbiased estimator of $\{g_k^0\}$ if it has the form

$$\tilde{\mathbf{g}} = \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{0}_{M\times(N+1)} \\ \hline \mathbf{A}_{21} & \mathbf{I}_{N+1}\end{array}\right]\hat{\mathbf{g}}^s. \tag{12}$$

The following theorem establishes the optimality of the proposed estimator within the aforementioned class of estimators.

Theorem 4.1: The impulse response estimator (11) has the smallest covariance matrix among the class of causal linear asymptotically unbiased estimators of $\{g_k^0\}$.

Proof: Consider (12). The covariance matrices of $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{g}}^{cs}$ can be written as

$$\begin{aligned}\mathrm{Cov}(\tilde{\mathbf{g}}) &= \mathbf{A}\mathbf{P}\mathbf{A}^T = (\mathbf{A}\mathbf{C})(\mathbf{A}\mathbf{C})^T \\ &= \left[\begin{array}{c|c} \mathbf{A}_{11}\mathbf{C}_{11} & \mathbf{0} \\ \hline \mathbf{A}_{21}\mathbf{C}_{11} + \mathbf{C}_{21} & \mathbf{C}_{22}\end{array}\right] \\ &\quad \times \left[\begin{array}{c|c} \mathbf{C}_{11}^T\mathbf{A}_{11}^T & \mathbf{C}_{11}^T\mathbf{A}_{21}^T + \mathbf{C}_{21}^T \\ \hline \mathbf{0} & \mathbf{C}_{22}^T\end{array}\right], \\ \mathrm{Cov}(\tilde{\mathbf{g}}^{cs}) &= \mathbf{C}\left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_{N+1}\end{array}\right]\mathbf{C}^T = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{C}_{22}\mathbf{C}_{22}^T\end{array}\right] \qquad (13)\end{aligned}$$

[1] This assumption is based on the practical rule-of-thumb of starting with $\gamma = N/20$, and then increasing the value of γ until the emerging details are predominately spurious [4].

Therefore,

$$\mathrm{Cov}(\tilde{\mathbf{g}}) - \mathrm{Cov}(\tilde{\mathbf{g}}^{cs}) = \\ \left[\begin{array}{c|c} \mathbf{A}_{11}\mathbf{C}_{11} & \mathbf{0} \\ \hline \mathbf{A}_{21}\mathbf{C}_{11} + \mathbf{C}_{21} & \mathbf{0} \end{array}\right] \left[\begin{array}{c|c} \mathbf{A}_{11}\mathbf{C}_{11} & \mathbf{0} \\ \hline \mathbf{A}_{21}\mathbf{C}_{11} + \mathbf{C}_{21} & \mathbf{0} \end{array}\right]^T \succeq \mathbf{0},$$

which shows that $\tilde{\mathbf{g}}^{cs}$ has the smallest covariance matrix among all estimators of the form (12). ■

The next result shows that increasing the number of anti-causal terms for the computation of $\hat{\mathbf{g}}^s$ can only improve its variance.

Theorem 4.2: Consider $0 < M_1 < M_2 < N$ integers. Then, $\tilde{\mathbf{P}}_{M_1} \succeq \tilde{\mathbf{P}}_{M_2}$, where $\tilde{\mathbf{P}}_{M_1}$ and $\tilde{\mathbf{P}}_{M_2}$ are the covariance matrices of the $N+1$ causal impulse response estimators (11) based on $\{\hat{g}_t^s\}_{t=-M_1}^N$ and $\{\hat{g}_t^s\}_{t=-M_2}^N$, respectively.

Proof: We denote the covariance matrices of $\hat{\mathbf{g}}^s$ considering M_1 and M_2 non-causal data samples as $\mathbf{P}_{M_1}$ and $\mathbf{P}_{M_2}$ respectively, where

$$\mathbf{P}_{M_1} = \left[\begin{array}{c|c} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \hline \mathbf{P}_{12}^T & \mathbf{P}_{22} \end{array}\right], \quad \mathbf{P}_{M_2} = \left[\begin{array}{c|c|c} \mathbf{P}_{00} & \mathbf{P}_{01} & \mathbf{P}_{02} \\ \hline \mathbf{P}_{01}^T & \mathbf{P}_{11} & \mathbf{P}_{12} \\ \hline \mathbf{P}_{02}^T & \mathbf{P}_{12}^T & \mathbf{P}_{22} \end{array}\right],$$

with $\mathbf{P}_{00} \in \mathbb{R}^{(M_2-M_1)\times(M_2-M_1)}$, $\mathbf{P}_{11} \in \mathbb{R}^{M_1\times M_1}$ and $\mathbf{P}_{22} \in \mathbb{R}^{(N+1)\times(N+1)}$. From (13) it is possible to write

$$\tilde{\mathbf{P}}_{M_1} = \mathbf{C}_{22}\mathbf{C}_{22}^T = \mathbf{P}_{22} - \mathbf{P}_{12}^T\mathbf{P}_{11}^{-1}\mathbf{P}_{12}.$$

After some algebra we obtain

$$\tilde{\mathbf{P}}_{M_2} = \mathbf{P}_{22} - \left[\begin{array}{c|c} \mathbf{P}_{02}^T & \mathbf{P}_{12}^T \end{array}\right] \left[\begin{array}{c|c} \mathbf{P}_{00} & \mathbf{P}_{01} \\ \hline \mathbf{P}_{01}^T & \mathbf{P}_{11} \end{array}\right]^{-1} \left[\begin{array}{c} \mathbf{P}_{02} \\ \hline \mathbf{P}_{12} \end{array}\right].$$

Therefore, by a matrix inequality related to Schur complements [20, Theorem 2.1] applied to $\mathbf{P}_{M_2}$, we conclude that

$$\tilde{\mathbf{P}}_{M_1} - \tilde{\mathbf{P}}_{M_2} = \\ \left[\begin{array}{c|c} \mathbf{P}_{02}^T & \mathbf{P}_{12}^T \end{array}\right] \left[\begin{array}{c|c} \mathbf{P}_{00} & \mathbf{P}_{01} \\ \hline \mathbf{P}_{01}^T & \mathbf{P}_{11} \end{array}\right]^{-1} \left[\begin{array}{c} \mathbf{P}_{02} \\ \hline \mathbf{P}_{12} \end{array}\right] - \mathbf{P}_{12}^T\mathbf{P}_{11}^{-1}\mathbf{P}_{12} \succeq \mathbf{0},$$

which shows that involving more non-causal points of $\hat{g}_t^s$, the covariance matrix of the causal estimators of $\{g_k^0\}_{k\geq 0}$ based on the smoothed ETFE can only decrease. ■

Remark: The value of M cannot exceed the rank of the covariance matrix $\mathbf{P}$, due to the restriction of invertibility of $\mathbf{C}_{11}$. It is clear that

$$\begin{aligned} \mathrm{rank}\{\mathbf{P}\} &= \mathrm{rank}\left\{\frac{\sigma}{N}\mathbf{B}\mathrm{diag}(w_\gamma)\mathbf{U}\right\} \\ &\leq \min(\mathrm{rank}\{\mathbf{B}\}, \mathrm{rank}\{\mathbf{U}\}) \\ &\leq \min(N+M+1, 2\gamma+1, N), \end{aligned}$$

which implies that the parameter M must be not larger than $2\gamma+1$. Over extensive simulations, we have noticed that values of M close to γ usually give satisfactory results.

V. Numerical example

For the numerical example, the Cholesky factor $\mathbf{C}$ is computed from (10) as

$$\mathbf{C} = \frac{\sigma}{N}\mathbf{B}\,\mathrm{diag}(w_\gamma)\,\mathbf{U}\mathbf{Q},$$

where $\mathbf{Q}$ is an orthogonal matrix obtained by the QR decomposition of $\mathbf{P}^{1/2}$ [18]. The entries of matrix $\mathbf{B}$ in (9) are calculated by trapezoidal approximation of

$$\beta_k = \frac{1}{\pi}\mathrm{Re}\left\{\int_0^\pi \frac{e^{j\omega(\gamma+k)}}{N_{\tilde{U}}(e^{j\omega})}d\omega\right\}.$$

Consider the system (1) with

$$G_0(q) = \frac{0.41275(q^2 - 1.98q + 1)}{(q - 0.9704)(q^2 - 1.852q + 0.9231)},$$

where $\{u_t\}$ is unit variance Gaussian white noise and $\{v_t\}$ is Gaussian white noise of variance $\sigma^2 = 0.05$. Figure 1 shows non-parametric estimates with a Hamming window of width $\gamma = 180$, $N = 1000$ and $M = 150$. Note that the estimate $\{\hat{g}_t^s\}$ is non-causal, and that the causal frequency estimate $\hat{G}_N^{cs}(e^{j\omega})$ shows an improvement over the more erratic behavior of $\hat{G}_N^s(e^{j\omega})$ in the low and mid-frequency range. The mean square error (MSE) of the frequency response estimates are $0.796 \cdot 10^{-2}$ and $0.565 \cdot 10^{-2}$ for the smoothed ETFE and causal smoothed ETFE, respectively. For this example, the MSE is reduced by 29%. These results show that setting the non-causal estimates to zero contributes significantly to reducing the MSE of the frequency response estimate.

To statistically test the estimator, 500 Monte-Carlo simulations were performed under the previous setup, with a fixed input sequence. For this example, we consider $\hat{\mathbf{g}}$ as the vector with elements $\{\hat{g}_t\}_{t=-M}^N$ formed by the ETFE obtained by (2), and $\tilde{\mathbf{g}}^c$ as the vector with elements $\{\tilde{g}_t^c\}_{t=-M}^N$ formed by the causal standard ETFE, that is, the causal estimator derived with (11) using the estimations given by (2). The empirical variances of each sample are shown in Figure 2, and the MSE of the first 150 causal samples for each estimator are given in Table I.

TABLE I

Trace of the covariance matrix and mean square error of the first 150 samples for ETFE, causal ETFE, smoothed ETFE and causal smoothed ETFE respectively.

	$\hat{\mathbf{g}}$	$\tilde{\mathbf{g}}^c$	$\hat{\mathbf{g}}^s$	$\tilde{\mathbf{g}}^{cs}$
tr{cov(·)}	$1.609 \cdot 10^{-2}$	$1.329 \cdot 10^{-2}$	$0.365 \cdot 10^{-2}$	$0.345 \cdot 10^{-2}$
MSE	$1.789 \cdot 10^{-2}$	$1.476 \cdot 10^{-2}$	$1.029 \cdot 10^{-2}$	$1.076 \cdot 10^{-2}$

The first row of Table I shows that, for both ETFE and smoothed ETFE, the causal estimators improve the variance of each estimate. This is a direct consequence of Theorem 4.1. Also, the MSE of the causal estimates can be decreased in some cases, depending on the bias imposed by the causal estimation.

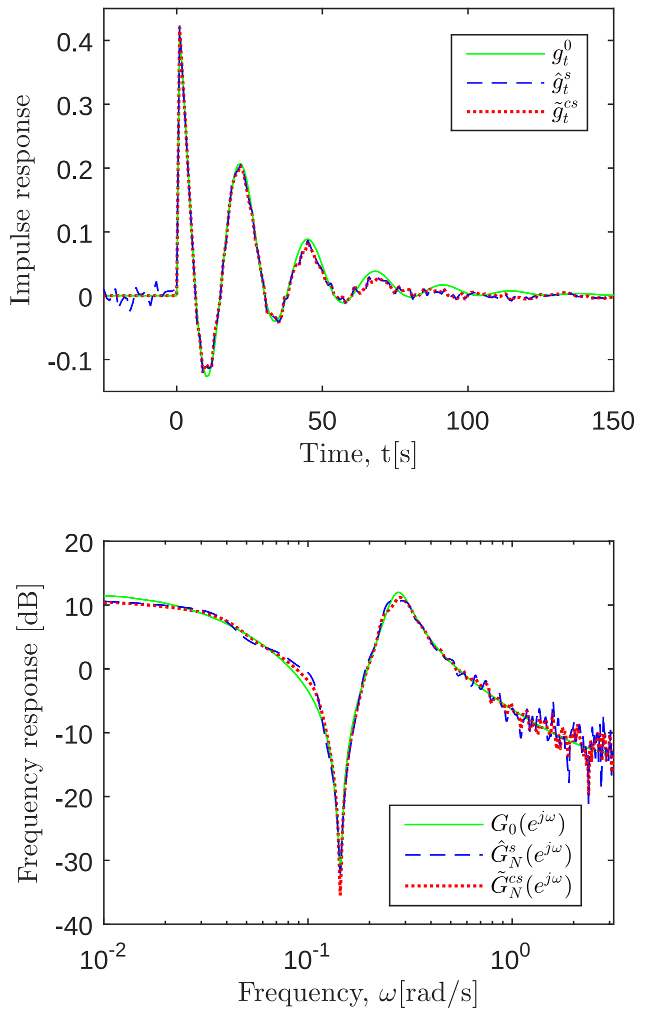

Fig. 1. Time (left) and frequency-domain (right) non-parametric estimations with smoothed ETFE (blue, dashed), and the proposed method (red, dotted).

TABLE II

MSEs of the first 150 samples for ETFE, causal ETFE, smoothed ETFE and causal smoothed ETFE considering $M = 50, 100, 150, 200$.

$M\backslash$MSE estimator	$\hat{\mathbf{g}}$	$\tilde{\mathbf{g}}^c$	$\hat{\mathbf{g}}^s$	$\tilde{\mathbf{g}}^{cs}$
50	$2.329{\cdot}10^{-2}$	$2.156{\cdot}10^{-2}$	$1.004{\cdot}10^{-2}$	$0.976{\cdot}10^{-2}$
100	$2.329{\cdot}10^{-2}$	$2.043{\cdot}10^{-2}$	$1.004{\cdot}10^{-2}$	$0.966{\cdot}10^{-2}$
150	$2.329{\cdot}10^{-2}$	$1.999{\cdot}10^{-2}$	$1.004{\cdot}10^{-2}$	$0.990{\cdot}10^{-2}$
200	$2.329{\cdot}10^{-2}$	$1.919{\cdot}10^{-2}$	$1.004{\cdot}10^{-2}$	$1.140{\cdot}10^{-2}$

Also, the relationship between M and the MSE of each estimator is studied. By considering 500 Monte-Carlo simulations with $N = 1000$ and $\gamma = 250$, the MSE for different values of M is shown in Table II.

Obviously, the MSE of the non-causal estimators $\hat{\mathbf{g}}$ and $\hat{\mathbf{g}}^s$ remains unchanged, because the causal estimates do not depend on the number of non-causal points considered. This

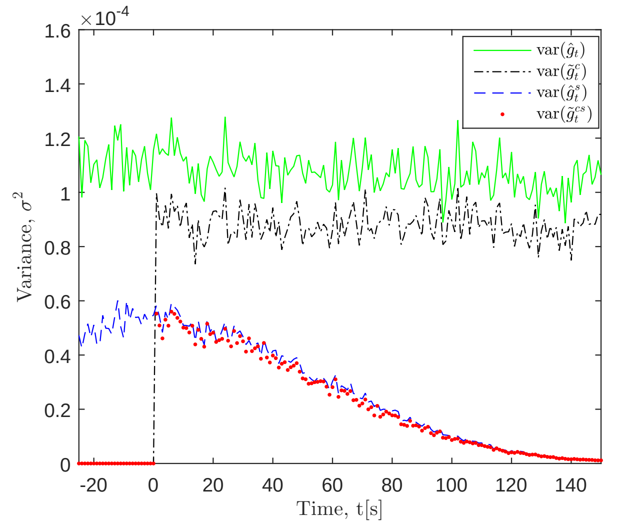

Fig. 2. Variance of each element of $\hat{\mathbf{g}}$ (green, solid), $\tilde{\mathbf{g}}^c$ (black, dash-dot), $\hat{\mathbf{g}}^s$ (blue, dashed), and $\tilde{\mathbf{g}}^{cs}$ (red, dotted).

example shows that larger values of M can have a positive effect on the MSE of causal estimators. Note that there may exist a trade-off between improving the variance, and the increase of bias. As seen in Table I, the variance always decreases, but this is not necessarily so for the bias, which also plays an important role in the MSE computations. The reduced gain in performance for the smoothed ETFE case is due to the improved accuracy on non-causal estimates, which reduces the magnitude of these values and therefore, decreases the corrections made to $\{\hat{g}_t^s\}_{t=0}^N$.

VI. Conclusions

We have proposed a novel non-parametric estimator that optimally enforces causality on the smoothed ETFE under a certain class of linear estimators. An explicit expression for the impulse response estimator by smoothed ETFE has been obtained, and an optimal causal estimator has been put forward and analyzed. Simulations have indicated that the proposed method reduces the variance of the traditional spectral analysis estimation, and that the size of the MSE reduction is dependent on the number non-causal samples used to estimate the ETFE.

References

[1] P. D. Welsh, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, IEEE Transactions on Audio and Electroacoustics 15 (2) (1967) 70–73.

[2] P. E. Wellstead, Non-parametric methods of system identification, Automatica 17 (1) (1981) 55–69.

[3] T. Söderström, P. Stoica, System identification, Prentice-Hall, 1988.

[4] L. Ljung, System Identification: Theory for the User, 2nd edition, Prentice-Hall, 1999.

[5] P. Stoica, R. L. Moses, Introduction to spectral analysis, Prentice-Hall, 1997.

[6] L. Ljung, On the estimation of transfer functions, Automatica 21 (6) (1985) 677–696.

[7] P. M. Broersen, Mean square error of the empirical transfer function estimator for stochastic input signals, Automatica 40 (1) (2004) 95–100.

[8] A. Stenman, F. Gustafsson, Adaptive smoothing methods for frequency-function estimation, Automatica 37 (5) (2001) 675–685.

[9] J. Schoukens, G. Vandersteen, K. Barbé, R. Pintelon, Nonparametric preprocessing in system identification: a powerful tool, in Proceedings of the 10th IEEE European Control Conference (ECC), Budapest, Hungary (2009) 1–14.

[10] T. McKelvey, G. Guérin, Non-parametric frequency response estimation using a local rational model, IFAC Proceedings Volumes 45 (16) (2012) 49–54.

[11] P. Hägg, H. Hjalmarsson, B. Wahlberg, A least squares approach to direct frequency response estimation, in Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference, Orlando, United States (2011) 2160–2165.

[12] M. Gevers, P. Hägg, H. Hjalmarsson, R. Pintelon, J. Schoukens, The transient impulse response modeling method and the local polynomial method for nonparametric system identification, IFAC Proceedings Volumes 45 (16) (2012) 55–60.

[13] A. Kessy, A. Lewin, K. Strimmer, Optimal whitening and decorrelation, arXiv preprint arXiv:1512.00809 (2015).

[14] Y. C. Eldar, A. V. Oppenheim, MMSE whitening and subspace whitening, IEEE Transactions on Information Theory 49 (7) (2003) 1846–1851.

[15] A. V. Oppenheim, A. S. Willsky, S. H. Nawab, Signals and systems, Prentice-Hall, 1983.

[16] W. Rudin, Real and complex analysis, 3rd edition, McGraw-Hill, 1987.

[17] R. Pintelon, J. Schoukens, System identification: A frequency domain approach, 2nd Edition, Wiley-IEEE Press, 2012.

[18] R. A. Horn, C. R. Johnson, Matrix analysis, Cambridge university press, 2nd ed., 2012.

[19] W. Rudin, Functional analysis, 2nd edition, McGraw-Hill, 1991.

[20] Y. Zi-Zong, Schur complements and determinant inequalities, Journal of mathematical inequalities 3 (2) (2009) 161–167.