Continuous-time linear time-varying system identification with a frequency domain kernel based estimator

John Lataire[1,*], Rik Pintelon[1], Dario Piga[2], Roland Tóth[3]

[1]Vrije Universiteit Brussel (VUB), Dept. ELEC, Pleinlaan 2, 1050 Brussels, Belgium
[*]jlataire@vub.ac.be
[2]IMT School for Advanced Studies Lucca, Lucca, Italy
[3]Eindhoven University of Technology, Dept. of Electrical Engineering, The Netherlands

Abstract: A novel estimator for the identification of Continuous-time Linear Time-Varying systems is presented in this paper. The estimator uses kernel based regression to identify the time-varying coefficients of a linear ordinary differential equation, based on noisy samples of the input and output signals. The estimator adopts a mixed time- and frequency domain formulation, which allows it to be formulated as the solution of a set of algebraic equations, without relying on finite differences to approximate the time derivatives. Since a kernel based approach is used, the model complexity selection of the time-varying parameters is formulated as an optimization problem with continuous variables. Variance and bias expressions of the estimate are derived and validated on a simulation example. Also, it is shown that, in highly noisy environments, the proposed kernel based estimator provides more reliable results than an 'Oracle' based estimator which is deprived of regularisation.

Keywords – System Identification, Time-Varying systems, Kernel Based Regression, Frequency domain, Continuous-time

1. Introduction

This article considers the identification of Single-Input Single-Output (SISO) continuous-time Linear Time-varying (LTV) systems through a black-box approach. As a model structure, linear ordinary differential equations are considered, with coefficients which vary smoothly with time. These model structures accurately describe the behavior of many real-world systems, like electrical impedances in electrochemical processes [3], biological systems [18], or mechanical systems with time-varying set-points [21].

Different options for modelling the time variation of a system have been explored in the literature. In the introduction of [13], estimators of time-varying systems have been categorised based on whether the dynamics and the time variation are estimated parametrically or non-parametrically. The method proposed in the current paper can be categorised as *parametric in the dynamics* and *non-parametric in the time variation*. Specifically, the system behaviour is described by a continuous-time differential equation (i.e., parametric in the dynamics), where the coefficients are time-varying. These coefficients are estimated non-parametrically in the time. In essence, this means that the number of parameters to be estimated grows with the number of measured data points. Other examples of estimators which fall into the same category are i) the estimation of a finite impulse response model via recursive least squares in

[10], or via kernel based methods [5] ii) the estimation of a state-space model with a different state transition matrix at each time instant in [9], and iii) a time-varying state-space model where the system matrices depend on a non-parametrically estimated function [11].

In [7] an estimator has been proposed which is parametric in both the dynamics and the time variation. Namely, the system is described by an ordinary differential equation, the coefficients of which are written as expansions of basis functions that depend on the time. The estimator in the present article is inspired by [7]. The essential difference is that the time-varying coefficients will be estimated via kernel based regression, which is a non-parametric technique. A specific distinction between basis function expansion and kernel based regression is that, for the former, the selection of the number of basis functions is a discrete problem, of potentially combinatorial complexity, typically handled as the optimisation of Akaike's criterium [2] or cross-validation techniques. On the other hand, model complexity selection is a continuous optimisation problem for the kernel based method. Criteria to determine the optimal model complexity include the maximisation of the Marginal Likelihood [16, Chapter 5], and Leave-One-Out Cross-Validation (LOO-CV) [23]. The latter will be applied in this article to tune the level of smoothness of the time-varying coefficients of continuous-time dynamical systems. The other complexity measure that must be determined is the dynamical order of the system. This is assumed to be a priori fixed by the user in this paper. To tune it, regularisation methods have been proposed in [17].

Kernel based methods have been used in the context of system identification in the recent literature. In [4] and [12], the identification of linear time invariant (LTI) impulse responses has been discussed. Discrete-time linear parameter varying (LPV) systems have been estimated in [5]. In [11] and [8] continuous-time models are estimated in the time domain, respectively in LTV state-space form, and in LPV input-output form. The drawback of these continuous-time methods is that, either a matrix differential equation must be solved at each iteration step of the algorithm, or the derivatives must be approximated numerically.

One challenge of identifying continuous-time systems from sampled data is the correct handling of the derivatives of the signals in the differential equation. From [14], accurately computing the derivatives in the time domain involves the careful design of high order digital pre-filters, or requires oversampling the data. In this paper, the approach presented in [7] will be adopted, which uses a combined time- and frequency-domain formulation of the system equation. This circumvents the approximation of time derivatives by finite differences, and formulates the identification of a *differential* equation as the problem of solving a set of *algebraic* equations.

Both the input and the output signals will be assumed to be disturbed by additive Gaussian noise. Bias and variance expressions of the estimated parameters will be derived. The bias is due to the fact that i) the regressors are noisy and ii) the use of kernel based regression implies a regularisation in the cost function. However, a bias is not necessarily detrimental to the quality of the estimate. Namely, regularisation has been shown to balance the bias and the variance which often results in a lower Mean-Squared-Error, see [12] for the LTI case. We will demonstrate on a simulation example that, in highly noisy cases, the kernel based estimator provides more reliable estimates than an 'Oracle' estimator (i.e. which uses the true model structure) deprived of regularisation.

As has been shown in [15] for LTI systems and in [7] for LTV systems, when the noise appearing in the system equation is coloured (which is usually the case with dynamic systems), the estimates can still be made consistent, by including an appropriate weighting. Although

the Kernel Based Estimator presented here is not consistent, including such a weight will be shown to decrease the Mean Squared Error of the estimate.

The estimator presented in this paper resembles the one in [6], which considers discrete-time, time-varying systems and presents a limited discussion on its properties (no bias nor variance analysis is given). The main novel contributions of the present article are i) the formulation and rigorous derivation of a kernel based, frequency domain estimator of *continuous-time* LTV systems , ii) the construction of associated covariance and bias expressions, and iii) a comparison with a basis function expansion based 'Oracle' estimator.

The paper is organised as follows. **Section 2** introduces the considered model class and noise assumptions, and formulates the identification problem. In **Section 3** the differential equation describing the system is translated to the frequency domain and the kernel based estimator is formulated. In **Section 4**, the covariance and the bias of the kernel based estimates are computed, while **Section 5** discusses the tuning of the kernel. In **Section 6**, the practical implementation of the estimators is explained. In **Section 7**, the use of the kernel based estimator is demonstrated and compared to the parametric basis-function-based estimator.

2. Model assumptions and problem formulation

Definition 1 (Noiseless Model class)**.** The noiseless input and output signals, $u_\circ(t)$ and $y_\circ(t)$, satisfy a linear ordinary differential equation with time-varying coefficients, given by

$$\psi_r(p)y_\circ(t) = -\sum_{\substack{n=0\\n\neq r}}^{N_a} a_n(t)\psi_n(p)y_\circ(t) + \sum_{n=0}^{N_b} b_n(t)\psi_n(p)u_\circ(t), \tag{1a}$$

$$\psi_n(\bullet) = \begin{cases} jP_n(\bullet T_\mathrm{s}/(j\pi)) & \text{for } n \text{ odd},\\ P_n(\bullet T_\mathrm{s}/(j\pi)) & \text{for } n \text{ even}, \end{cases} \tag{1b}$$

with $a_n(t)$ and $b_n(t)$ smooth functions of t, P_n the n-th degree Legendre polynomial, see [1, Section 22], $j = \sqrt{-1}$, p^n the n-th derivative operator, and T_s the sample time.

Remark 2. The choice of Legendre polynomials in (1) is made without loss of generality, since they form a complete basis for polynomials. Other choices of polynomial basis functions are equally valid. However, Legendre polynomials in (1b) have been observed to usually yield a better numerical conditioning than using simple derivatives in (1), see [7], and will therefore be used in this article. Also, note that the polynomials ψ_n have real coefficients.

For notational convenience, the derivations will be done for $r = 0$ in (1). In this paper, the model orders N_a and N_b are assumed to be known. The sample time T_s is chosen such as to satisfy the Shannon-Nyquist theorem on the input and output signals, i.e. the signals are assumed to be band-limited. Sampled and windowed, noisy measurements of the input and output signals, denoted $u(t)$ and $y(t)$, are acquired, for $t \in \mathbb{T}$, with $\mathbb{T} = \{0, T_\mathrm{s}, \cdots, (N-1)T_\mathrm{s}\}$, and where N is the number of acquired samples. Denote $u, y \in \mathbb{R}^N$ the column vectors stacking $u(t)$ and $y(t)$, for $t \in \mathbb{T}$, and $Y = Fy$, $U = Fu$ their DFTs, consistent with the following definition.

Definition 3 (DFT)**.** The Discrete Fourier Transform (DFT) of a vectorized sampled signal

$x \in \mathbb{R}^N$ is given by $X = Fx$, where $F \in \mathbb{C}^{N\times N}$ is the DFT matrix, which concatenates

$$F(k,t) = N^{-\frac{1}{2}} e^{\frac{-j2\pi kt}{NT_\mathrm{s}}}, \qquad \text{for } k \in \mathbb{K},\ t \in \mathbb{T}. \tag{2}$$

where $\mathbb{K}$ is the ordered set of DFT bins:

$$\mathbb{K} = \{-\lfloor N/2 \rfloor, -\lfloor N/2 \rfloor + 1, \cdots, \lceil N/2 \rceil - 1\}. \tag{3}$$

The notations $\lceil . \rceil$ and $\lfloor . \rfloor$ represent the ceiling and floor functions, respectively. Note that F is a unitary transformation, such that $F^{\mathrm{H}}F = I$, where superscript $^{\mathrm{H}}$ denotes the conjugate transpose.

Assumption 4 (Noise assumptions)**.** The measured signals y and u are disturbed by additive, zero-mean, stationary noise, with DFTs V_Y and V_U, such that

$$Y = Y_\circ + V_Y, \qquad U = U_\circ + V_U. \tag{4}$$

V_Y and V_U are zero-mean and Gaussian distributed, and uncorrelated with $Y_\circ$ and $U_\circ$.

This corresponds for instance to the special case of the Errors-In-Variables noise assumption where the system operates in open-loop and does not produce any noise itself, but the measurement channels do. The noise covariance matrices are defined as

$$C_Y = \mathbb{E}\left\{V_Y V_Y^{\mathrm{H}}\right\},\ C_U = \mathbb{E}\left\{V_U V_U^{\mathrm{H}}\right\},\ C_{YU} = \mathbb{E}\left\{V_Y V_U^{\mathrm{H}}\right\} \tag{5}$$

Since the noise is stationary, we have that C_Y, C_U and C_{YU} are asymptotically ($N \to \infty$) diagonal matrices. This is because stationary noise can be modelled as filtered white noise, the DFT of which is asymptotically independent over the frequency [15, Theorem 16.25].

Problem formulation: The identification problem consists of estimating $a_n(t)$ and $b_n(t)$ in (1), based on sampled and windowed, noisy measurements of the input and output signals u and y, given that the noiseless signals obey (1) with known model orders N_a and N_b, and given Assumption 4 on the noise. We will distinguish the cases when the noise covariances are either known or not.

3. Methodology

3.1. Frequency domain system equation

Denote $\omega_k = \frac{2\pi k}{NT_\mathrm{s}}$ the angular frequency corresponding to the k-th DFT bin. We have that, if $x(t)$ is a periodic signal with period length NT_s, then $\psi_n(j\omega_k)X(k)$ is equal to the DFT of $\psi_n(p)x(t)$ at the kth bin. That is, for a periodic signal, a time-derivative corresponds to a multiplication of the DFT by (a power of) $j\omega_k$. This will give a convenient way of handling the time-derivatives in the system equation in (1). To handle non-periodic signals, a transient term will be taken into account, see further on.

Denote $\psi_n \in \mathbb{C}^N$ the column vector stacking $\psi_n(j\omega_k)$, $k \in \mathbb{K}$, see (3), and introduce the diagonal matrix

$$\phi_{n,x} \triangleq \mathrm{diag}\left(F^{\mathrm{H}}\left(\psi_n \odot X\right)\right), \tag{6}$$

with x to be substituted by y or u, X by Y or U, and $\odot$ the element-wise product. Introduce the following notations

$$c_n \triangleq \begin{cases} b_n & \text{for } 0 \leqslant n \leqslant N_b \\ a_{n-N_b} & \text{for } N_b < n \leqslant N_a + N_b + 1 \end{cases} \tag{7}$$

$$\phi_n \triangleq \begin{cases} \phi_{n,u} & \text{for } 0 \leqslant n \leqslant N_b \\ -\phi_{n-N_b,y} & \text{for } N_b < n \leqslant N_a + N_b + 1 \end{cases} \tag{8}$$

$$\Phi \triangleq \begin{bmatrix} \phi_0 & \cdots & \phi_{N_c} \end{bmatrix} \in \mathbb{R}^{N \times N(N_c+1)}, \tag{9}$$

$$C \triangleq [c_0^{\mathrm{T}} \ \cdots \ c_{N_c}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{N(N_c+1)}, \quad N_c = N_a + N_b, \tag{10}$$

where $a_n, b_n \in \mathbb{R}^N$ are equal to $a_n(t)$ and $b_n(t)$ vectorized with $t \in \mathbb{T}$.

Theorem 5 (DFT of the system equation)**.** Equation (1) is equivalent to

$$Y_\circ = F\Phi_\circ C + \Psi\gamma, \qquad \text{with} \quad \Psi \triangleq \begin{bmatrix} \psi_0 & \cdots & \psi_{N_\gamma} \end{bmatrix}, \tag{11}$$

such that $\Psi\gamma$ is a vectorised polynomial in $j\omega_k$ of degree $N_\gamma = \max(N_a, N_b) - 1$. The coefficient vector $\gamma \in \mathbb{R}^{(N_\gamma+1)}$ depends on the initial and end conditions $u_\circ^{(n)}(0)$, $y_\circ^{(n)}(0)$, $u_\circ^{(n)}(NT_\mathrm{s})$, $y_\circ^{(n)}(NT_\mathrm{s})$, and on the polynomial expansions of $c_n(t)$, with $n = 0, \cdots, \max(N_a, N_b) - 1$.

Proof. This has been proven for $c_n(t)$ a polynomial in t by [7] (see equations (12) and (23) in this reference). Since $c_n(t)$ is assumed to be smooth, it can be approximated arbitrarily well by a polynomial, which validates (11). □

Equation (11) expresses the system behaviour – the solution set of a *differential* equation – as the solution of a set of *algebraic* equations, without having to approximate the time derivatives by finite differences. The polynomial $\Psi\gamma$ takes into account the transient effects, such that (11) is valid, even if the signals $u(t)$ and $y(t)$ are non-periodic. The problem of identifying a system given by (1) is reformulated as the estimation of C and γ from the measured input and output DFTs U and Y, given that the noiseless DFT spectra satisfy (11).

Remark 6 (Residual alias error)**.** As explained in [15, Appendix 6.F] (for LTI systems) and in [7] for the system class given by (1), the conversion of the differential equation (1) into the algebraic equations (11) introduces a small residual alias error, when the signals are not periodic (even though perfectly band-limited). Nevertheless, this alias error is smooth and, thus, can be captured by an additional polynomial in $j\omega_k$. In practice, this alias error will be captured by the transient polynomial $\Psi\gamma$, the degree of which can be increased (by a few units, typically) until no significant decrease of the estimation residuals is detected (i.e. $N_\gamma \geqslant \max(N_a, N_b) - 1$).

Remark 7. For simplicity, and in view of Remark 6, the knowledge that γ depends on the initial and end conditions, on the model parameters, and on the residual alias error will not be used during the identification. That is, γ will be estimated alongside C as if it were an independent vector of parameters. Given the limited dimension of γ, this is unlikely to have a significant impact on the performance of the estimation.

3.2. Kernel based estimator (KBE)

The Kernel based estimator defined in this section, and further discussed in the next sections, constitutes the main contribution of this article.

Definition 8 (KBE)**.** The kernel based estimate of the system in (11) is

$$\hat{C}, \hat{\gamma} = \underset{C,\gamma}{\operatorname{argmin}}\, E^{\mathrm{H}} W^{-1} E + \breve{C}^{\mathrm{T}} \mathbf{K}^{-1} \breve{C}, \tag{12a}$$

$$E(C,\gamma) \triangleq Y - F\Phi C - \Psi\gamma, \qquad \mathbf{K} \triangleq I_{N_c+1} \otimes K, \tag{12b}$$

$$\mathring{C} \triangleq [\mathring{c}_0 \ \cdots \ \mathring{c}_{N_c}]^{\mathrm{T}} \in \mathbb{R}^{N_c+1}, \qquad C \triangleq \breve{C} + \mathring{C} \otimes \mathbf{1}, \tag{12c}$$

with $\otimes$ denoting the Kronecker product, I_{N_c+1} the identity matrix, and $\mathbf{1} \in \mathbb{R}^{N\times 1}$ a vector of ones.

Note that the vector of time-varying coefficients C is written as a time-varying part $\breve{C}$, and a constant part $\mathring{C}$. The matrix $W \in \mathbb{C}^{N\times N}$ is a symmetric, positive definite weighting matrix, and $K \in \mathbb{R}^{N\times N}$ is a kernel matrix (symmetric and semi-positive definite) which regularises the estimated time-varying part $\breve{C}$. In this paper, the same kernel is used for all time-varying coefficients for notational and computational simplicity. The method can straightforwardly be extended to allow a different kernel for each coefficient. The choice of the kernel and weighting matrices W and K have a great impact on the resulting estimate, see Sections 5 and 6. The solution of the KBE in Definition 8 is computed explicitly as follows (proof in Appendix A).

Theorem 9 (KBE solution)**.** The minimizers in (12a) are

$$\hat{\mathring{C}} = \mathbf{K}\Phi^{\mathrm{H}} F^{\mathrm{H}} \alpha \tag{13a}$$

$$\begin{bmatrix} \alpha \\ \hat{\mathring{C}} \\ \hat{\gamma} \end{bmatrix} = \left(\Lambda + \begin{bmatrix} W & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} F\Phi\mathbf{K}\Phi^{\mathrm{H}} F^{\mathrm{H}} & F\mathring{\Phi} & \Psi \\ \mathring{\Phi}^{\mathrm{H}} F^{\mathrm{H}} & 0 & 0 \\ \Psi^{\mathrm{H}} & 0 & 0 \end{bmatrix} \tag{13b}$$

with $\mathring{\Phi} = \begin{bmatrix} \mathring{\phi}_0 & \cdots & \mathring{\phi}_{N_c} \end{bmatrix}$, and $\mathring{\phi}_n \in \mathbb{R}^N$ a column vector with the diagonal elements of ϕ_n.

All the 0's in (13b) and in further equations are zero vectors or matrices with appropriate dimensions. Note that Theorem 9 corresponds to [25, Theorem 1.3.1], applied to the specific problem of this paper. It can be shown that $\hat{C}$ is real, such that the associated estimate of the differential equation is real as well.

Remark 10 (Selection of the frequency band)**.** The data fitting part $E^{\mathrm{H}} W^{-1} E$ in the cost function in (12a) is formulated in the frequency domain. As a consequence the estimation can be limited to a selected frequency band of interest. This is important in the context of continuous-time systems: the frequency band close to half the sampling frequency (where aliasing is likely to occur) can be discarded. Also, omitting the frequency band where the measured spectra do not contain information (e.g., where the spectrum of the input signal is zero) decreases the computational load of the estimator.

3.3. Basis function based estimator (BFE)

The KBE will be compared with the *Basis function based Estimator* (BFE), proposed in [7]. The BFE parameterises the time-varying coefficients $c_n(t)$ as linear combinations of N_p basis functions $f_p(t)$, $p = 0, \cdots, N_p$.

Definition 11 (BFE)**.** The Basis functions based estimate is given by

$$\hat{\theta}, \hat{\gamma} = \underset{\theta,\gamma}{\operatorname{argmin}}\, E(\theta,\gamma)^{\mathrm{H}} W^{-1} E(\theta,\gamma), \tag{14a}$$

$$E(\theta,\gamma) = Y - F \sum\nolimits_{n=0}^{N_c} \phi_n \hat{c}_n - \Psi\gamma, \tag{14b}$$

$$\hat{c}_n(t) = \sum\nolimits_{p=0}^{N_p} f_p(t)\hat{\theta}_{n,p}, \qquad n = 0, \ldots, N_c \tag{14c}$$

The basis functions $f_p(t)$ are assumed to be known. For this estimator, choosing W as the diagonal of $\operatorname{cov}(E)$ (see further) was proven in [7] to achieve consistency under some mild assumptions. Note that $\operatorname{cov}(E)$ depends on θ, making the optimisation problem non-convex.

4. Covariance and bias of the estimated parameters

Expressions for the noise covariance and bias are important because they provide a quality measure of the estimate. These expressions will be derived for the KBE in the following subsections. First, the covariance of the equation error E is introduced, as required subsequently.

4.1. Equation error and its covariance

For a given $\{a_n\}_{n=1}^{N_a}$ and $\{b_n\}_{n=0}^{N_b}$, the equation error defined in (12b) for the KBE and in (14b) for the BFE can be written as

$$E = \mathscr{A}Y - \mathscr{B}U - \Psi\gamma, \tag{15a}$$

$$\mathscr{A} = I_N + \sum\nolimits_{n=1}^{N_a} F \operatorname{diag}(a_n) F^{\mathrm{H}} \operatorname{diag}(\psi_n), \tag{15b}$$

$$\mathscr{B} = \sum\nolimits_{n=0}^{N_b} F \operatorname{diag}(b_n) F^{\mathrm{H}} \operatorname{diag}(\psi_n). \tag{15c}$$

Since $\Psi\gamma$ does not contain the noisy signals, we have

$$\begin{aligned} \operatorname{cov}(E) &\triangleq \mathbb{E}\left\{ \left(E - \mathbb{E}\{E\}\right) \left(E - \mathbb{E}\{E\}\right)^{\mathrm{H}} \right\} \\ &= \begin{bmatrix} \mathscr{A} & -\mathscr{B} \end{bmatrix} \begin{bmatrix} C_Y & C_{YU} \\ C_{UY} & C_U \end{bmatrix} \begin{bmatrix} \mathscr{A}^{\mathrm{H}} \\ -\mathscr{B}^{\mathrm{H}} \end{bmatrix}. \end{aligned} \tag{16}$$

4.2. Covariance of the estimated parameters

Theorem 12 (Covariance)**.** The noise covariance of the estimated parameters $\hat{C}$ and $\hat{\gamma}$ of the KBE can be approximated by the following expression:

$$\operatorname{cov} \begin{bmatrix} \mathring{\hat{C}} \\ \hat{\gamma} \end{bmatrix} \approx \begin{bmatrix} \mathbf{L} & 0 \\ 0 & I_{N_c+N_\gamma+2} \end{bmatrix} \left(JJ^{\mathrm{H}}\right)^{-1} R^{\mathrm{H}} \operatorname{cov}(E)\, R \left(JJ^{\mathrm{H}}\right)^{-1} \begin{bmatrix} \mathbf{L}^{\mathrm{H}} & 0 \\ 0 & I_{N_c+N_\gamma+2} \end{bmatrix} \tag{17a}$$

$$R \triangleq W^{-1} \begin{bmatrix} F\Phi\mathbf{L} & F\mathring{\Phi} & \Psi \end{bmatrix} \tag{17b}$$

$$J^{\mathrm{H}} = \begin{bmatrix} W^{-1/2} & 0 \\ 0 & I_{N(N_c+1)} \end{bmatrix} \begin{bmatrix} F\Phi\mathbf{L} & F\mathring{\Phi} & \Psi \\ I_{N(N_c+1)} & 0 & 0 \end{bmatrix} \tag{17c}$$

with $\mathbf{L} = I_{N_c} \otimes L$, and L the lower triangular Cholesky decomposition of K (s.t. $K = LL^{\mathrm{H}}$).

The evaluation of this covariance expression requires the noise covariances C_Y, C_U and C_{YU}. The proof is in Appendix B and explains why Theorem 12 is an approximation. Note that $(JJ^{\mathrm{H}})^{-1}$ in (17a) can be computed in a numerically stable fashion via the QR decomposition of J.

4.3. Bias of the estimated parameters

The KBE is biased for two reasons: i) the regressors ϕ are noisy (see the discussion in [20, p. 186]), and ii) the cost function includes a regularisation term. The sum of both bias contributions is given in the following theorem.

Theorem 13 (Bias)**.** Assuming that the true parameters (denoted $C_\circ$ and $\gamma_\circ$) exist, the bias is given by:

$$\begin{bmatrix} \breve{C}_{\mathrm{b}} \\ \mathring{C}_{\mathrm{b}} \\ \gamma_{\mathrm{b}} \end{bmatrix} = -\,\mathbb{E}\left\{ \mathbf{H}^{-1} \left(\begin{bmatrix} \mathbf{K}\Phi^{\mathrm{H}} F^{\mathrm{H}} \\ \mathring{\Phi}^{\mathrm{H}} F^{\mathrm{H}} \\ \Psi^{\mathrm{H}} \end{bmatrix} W^{-1} E_\circ + \begin{bmatrix} \breve{C}_\circ \\ 0 \\ 0 \end{bmatrix} \right) \right\}, \tag{18}$$

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{K}\Phi^{\mathrm{H}} F^{\mathrm{H}} \\ \mathring{\Phi}^{\mathrm{H}} F^{\mathrm{H}} \\ \Psi^{\mathrm{H}} \end{bmatrix} W^{-1} \begin{bmatrix} F\Phi & F\mathring{\Phi} & \Psi \end{bmatrix} + \begin{bmatrix} I_{N(N_c+1)} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where $E_\circ \triangleq E(C_\circ, \gamma_\circ)$ is the equation error, see (12b), evaluated in the true parameters.

This theorem is exact, see Appendix B. The following proposition provides an implementable approximation, and will be used in Section 7.3 on a simulation example.

Proposition 14 (Approximate bias)**.** The bias of the estimated parameters of the KBE can be approximated as

$$\begin{bmatrix} \breve{C}_{\mathrm{b}} \\ \mathring{C}_{\mathrm{b}} \\ \gamma_{\mathrm{b}} \end{bmatrix} \approx -\,\mathbf{H}^{-1} \left(\begin{bmatrix} \mathbf{K}\mathbb{E}\left\{\breve{B}\right\} \\ \mathbb{E}\left\{\mathring{B}\right\} \\ 0 \end{bmatrix} + \begin{bmatrix} \breve{C}_\circ \\ 0 \\ 0 \end{bmatrix} \right), \tag{19a}$$

$$\breve{B} = \begin{bmatrix} \vdots \\ (F\phi_n)^{\mathrm{H}} W^{-1} E_\circ \\ \vdots \end{bmatrix}, \qquad \mathring{B} = \begin{bmatrix} \vdots \\ (F\mathring{\phi}_n)^{\mathrm{H}} W^{-1} E_\circ \\ \vdots \end{bmatrix}, \tag{19b}$$

$$n = 0, \cdots, N_c.$$

The i-th element of $\mathbb{E}\left\{(F\phi_n)^{\mathrm{H}} W^{-1} E_\circ\right\}$ is given by

$$F^{\mathrm{H}}[i,:] W^{-1} (\mathscr{A}_\circ C_{YX} - \mathscr{B}_\circ C_{UX}) \operatorname{diag}(\psi_r^{\mathrm{H}}) F[:,i], \tag{19c}$$

and $\mathbb{E}\left\{(F\mathring{\phi}_n)^{\mathrm{H}} W^{-1} E_\circ\right\}$ is computed as

$$\operatorname{trace}\left\{ (\mathscr{A}_\circ C_{YX} - \mathscr{B}_\circ C_{UX}) \operatorname{diag}\left(\psi_r^{\mathrm{H}}\right) W^{-1} \right\}, \tag{19d}$$

$$\text{with} \begin{cases} r = n,\ X = U & \text{for } 0 \leqslant n \leqslant N_b \\ r = n - N_b,\ X = Y & \text{for } N_b < n \leqslant N_c \end{cases}$$

$\mathscr{A}_\circ$ and $\mathscr{B}_\circ$, from (15), are evaluated for the true parameters.

Proof. The approximation in (19a) results from the fact that the noise in $\mathbf{H}$ is discarded. Then, (19a) follows immediately from (18). The expressions (19c) and (19d) are obtained from simple algebraic operations, taking into account (5), (6), (8), the fact that $\mathbb{E}\{E_\circ\} = 0$, and that the noise is uncorrelated with the noiseless signals (Assumption 4).

Remark 15. Given that the estimate of $\breve{C}$ is not constrained to be zero-mean, a constant term can be exchanged between the estimates of $\mathring{C}$ and $\breve{C}$, such that they cannot be estimated uniquely. However, this is not a problem since we are only interested in the sum $C = \mathring{C} + \breve{C}$, which is estimated uniquely. The covariance and bias of $\hat{C}$ can straightforwardly be obtained from the covariances of $\hat{\mathring{C}}$ and $\hat{\breve{C}}$ in Theorem 12 and Theorem 13.

5. Kernel selection

5.1. Kernel structure selection

The use of semi-positive definite kernels for regularising a regression problem is an extensively studied topic in the literature, see [12, 16, 24, 25] just to name a few. The resulting estimate $\hat{C}$ highly depends on the choice of the kernel matrix K. A lot of work has been devoted to designing kernels and for an extensive discussion on this matter, the reader is referred to [19]. Typically, a kernel structure is selected, which is expected to be appropriate for the application at hand (i.e. in our case the kernel should capture the time function of the coefficient variations). In the simulation section, the squared exponential (SE) kernel will be used. That is, the elements of K are given by

$$\rho e^{\frac{-(t-t')^2}{l^2}}, \qquad \text{for } t, t' \in \mathbb{T} \tag{20}$$

which can be shown to favour smooth solutions for c_n. The kernel parameters l and ρ quantify, respectively, the length scale of the smoothness, and the inverse of the amount of regularisation applied. These kernel parameters must be tuned, as explained in the next section.

5.2. Tuning the kernel parameters with LTO-CV

We propose to tune the kernel parameters based on the Leave-two-out cross validation (LTO-CV) criterion, which is a frequency-domain modification of the Leave-One-Out cross-validation (LOO-CV). This tuning criterion aims at minimising the out-of-sample error, while using only the estimation data. The LTO-CV criterion is formulated in the frequency domain because, as has been shown in [6], this is more robust to time-correlated errors. The LTO-CV criterion is defined as

$$E_{\mathrm{LTO-CV}}(\Theta) = \sum\nolimits_{k \in \mathbb{K}} \left| Y(k) - \hat{Y}(k, \Theta)_{\setminus k} \right|^2, \tag{21}$$

where Θ denotes the vector of kernel parameters (given by $\Theta = [\rho\ l]$ for the kernel in (20)) and $\hat{Y}(k, \Theta)_{\setminus k}$ is defined as the estimated output spectrum at frequency bin k, computed using the parameters which have been estimated with neither the use of the system equation at that frequency nor that at the conjugate frequency $-k$. The frequencies at k and $-k$ are omitted simultaneously because the spectra at those frequencies are complex conjugates of each other.

5.3. Continuous tuning of the model complexity

It follows from the above discussion that the kernel parameters determine the *model complexity*. More importantly, the model selection criterion $E_{\mathrm{LTO-CV}}(\Theta)$ is *continuous* in Θ, such that it can be minimized via a *gradient based optimisation* algorithm. This is an attractive property of the KBE (and *kernel based regression methods* in general). In contrast to that, for the BFE the selection of *the number of basis functions* is a *discrete* optimisation problem of potentially combinatorial complexity.

Note that, being a nonlinear optimisation problem, the LTO-CV criterion is prone to have local optima. This is partially mitigated via an initial coarse grid search. Computationally efficient implementations of the LTO-CV criterion in (21) and its gradient are obtained from a minor modification of the expressions for LOO-CV in [23].

6. Algorithmic implementation

6.1. Choice of the frequency weighting W

Inspired by the fact that the BFE (Definition 11) is consistent (for known basis functions $f_p(t)$ and model orders), using the diagonal of the covariance of E as a weighting W in the KBE (Definition 8) is a sensible choice. In this case, consistency is not applicable, because the KBE is biased. Nevertheless, if the system lies in the considered model class (1) and Assumption 4 is satisfied, it can easily be shown that the optimiser of the expected value of the data fit term $E^{\mathrm{H}}W^{-1}E$ in (12a) is independent of the noise properties, shown by following the reasoning in [7]. It will be shown on simulations that this weighting reduces the *Mean-Squared-Error* (MSE) on the estimate.

6.2. Implementation of the Kernel based estimator

Algorithm 16. The KBE is implemented as follows.

1. Acquire the measured signals u and y, determine the model orders N_a and N_b (assumed to be known), and choose the kernel structure (for instance, a SE kernel).
2. Initialise the weighting matrix: $W \leftarrow I$
3. Tune the kernel parameters: $\Theta \leftarrow \operatorname{argmin}_{\Theta} E_{\mathrm{LTO-CV}}(\Theta)$
4. $\hat{C} \leftarrow$ from (13)
5. **if** the noise covariances are available, **repeat**
   (a) $V_{-} \leftarrow E^{\mathrm{H}}(\operatorname{diag}\operatorname{cov} E)^{-1}E$
   (b) $W \leftarrow (\operatorname{diag}\operatorname{cov} E)/\|\operatorname{diag}\operatorname{cov} E\|_1$
   (c) $\hat{C} \leftarrow$ from (13)
   (d) $V \leftarrow E^{\mathrm{H}}(\operatorname{diag}\operatorname{cov} E)^{-1}E$

   until $V_{-} \leqslant V$
6. $\hat{C} \leftarrow$ the parameter values that gave the smallest value for V in (5d).

The iterative procedure in Step 5 is a convex relaxation of the problem where $W = \operatorname{diag}(\operatorname{cov} E)$ would explicitly depend on the parameters C. This is inspired by the Iterative Quadratic Maximum Likelihood (IQML) estimator, see [22]. In Steps 5a, 5b and 5d, E is given by (12b), with the current values for $\hat{C}$. In step 5b, W is normalised to its 1-norm such that the relative balance in (12a) between the data fit $E^{\mathrm{H}}W^{-1}E$ and the regularisation term $\breve{C}^{\mathrm{T}}K^{-1}\breve{C}$ remains unaltered. Two different estimates are extracted from Algorithm 16:

KBE$_I$: defined as the estimate obtained at Step 4, and doesn't require the noise covariance information.

KBE$_W$: defined as the estimate obtained at Step 6, and relies on the availability of the noise covariances. The kernel parameters Θ obtained in Step 3 are retained.

6.3. Implementation of the Basis function based estimator

Two versions of the BFE are considered based on Definition 11, depending on whether the noise information is available: BFE$_W$ with $W = \operatorname{diag}(\operatorname{cov} E)$ and BFE$_I$ with $W = I$. The implementation is given in [7].

7. Simulation results

The main goals of the simulations in this section are the following:

- To show that the KBE is more robust to noise than the BFE. To this end, a low SNR of $10\,\mathrm{dB}$ on the input and output signals will be used. For the BFE, the true model orders and basis functions will be used. Therefore, the BFE can be seen as an 'Oracle' estimator and, in that sense, is put in a privileged position w.r.t. the KBE.
- To show the validity of the variance and bias expressions in Section 4 for the KBE.

7.1. Simulated and frozen system

The simulated signals satisfy (1), with $r = 2$, and where $a_n(t)$ and $b_n(t)$ are 8-th order polynomials in t, specified in Appendix C. The noise satisfies Assumption 4, with

$$C_Y = \operatorname{diag}\left|H_y(e^{-j\omega_k T_\mathrm{s}})\right|^2, \qquad C_U = \operatorname{diag}\left|H_u(e^{-j\omega_k T_\mathrm{s}})\right|^2$$

for $k \in \mathbb{K}$, where $H_y(z^{-1})$ and $H_u(z^{-1})$ are LTI filters (z^{-1} is the 1-sample backward-shift operator), see Appendix C. These are required to compute $W = \operatorname{diag}(\operatorname{cov} E)$, for KBE$_W$ and BFE$_W$. In this simulation example, $C_{YU} = 0$.

Definition 17 (Frozen system)**.** The frozen system associated with the model equation at the time instant $t \in [0, NT_\mathrm{s}]$ is the LTI system, whose frequency response function is parameterized by the instantaneous time t, *viz.*:

$$G_\mathrm{f}(j\omega, t) = \frac{b_0(t)\psi_0(j\omega) + b_1(t)\psi_1(j\omega)}{a_0(t)\psi_0(j\omega) + a_1(t)\psi_1(j\omega) + a_2(t)\psi_2(j\omega)}.$$

This frequency response function will be called the Frozen Frequency Response Function (FFRF) at time instant t.

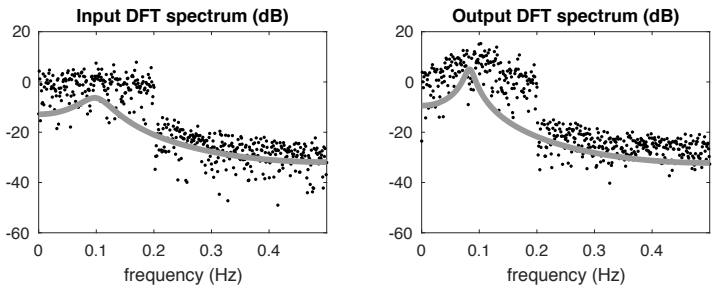

Fig. 1. *Noisy input and output DFT spectra (black dots •), and noise power spectrum (grey full line).*

Note that the frozen system at a given t is uniquely determined by the differential equation. The amplitude of the true FFRF as a surface plot is given in the top left of Fig. 3. The Mean-Square Error (MSE) on the estimated FFRF will be used as a performance indicator:

$$\mathrm{MSE}_{\mathbb{T}}(j\omega) \triangleq N^{-1} \sum\nolimits_{t\in\mathbb{T}} \left| G_{\mathrm{f}}(j\omega,t) - \hat{G}_{\mathrm{f}}(j\omega,t) \right|^2 . \tag{22}$$

The sample time is chosen as $T_{\mathrm{s}} = 1\mathrm{s}$. The noiseless input signal $u_{\circ}(t)$ is band-limited random noise with a flat amplitude spectrum in the frequency band $[0, 0.2]\,\mathrm{Hz}$. The signal-to-noise ratios of the noisy input and output signals is 10 dB. A total of 100 realisations of the disturbing input and output noise v_u and v_y are generated, and successively added to the noiseless input and output signals, giving 100 Monte Carlo experiments. For each noise realisation, a total of $N = 1014$ time domain samples of the input and output signals are acquired. An example of the noisy input and output DFT-spectra, together with the associated noise power spectra (i.e., the main diagonals of C_U and C_Y) are plotted in Fig. 1.

7.2. Higher robustness of the KBE to noise

The $\mathrm{MSE}_{\mathbb{T}}$, see (22), is shown in Fig. 2 for the noiseless signals (left), and for individual realisations of the noisy signals (right). For both estimators (KBE and BFE), $N_\gamma = 9$ in the noiseless case, and $N_\gamma = 4$ for the noisy case, in agreement with Remark 6. We observe the following.

Noiseless data The BFE and KBE give very accurate results. MSEs of about $90\,\mathrm{dB}$ and $80\,\mathrm{dB}$ below the true FFRFs are obtained, respectively. This observation confirms the validity of Theorem 5. For the KBE, the SE kernel in (20) is used, and the tuned kernel parameters (see Section 5.2) are $\rho = 10^9$, $l = 185\,\mathrm{s}$. The residual MSE of the BFE is due to the limited precision of the Ordinary Differential Equation (ODE) solver used to generate the data.

Noisy data For the KBE (black dots) and KBE_W (black squares) the MSEs lie respectively about $10\,\mathrm{dB}$ and $12\,\mathrm{dB}$ below the true FFRFs. The MSE of the BFE_W (grey squares), for one realisation of the noise, lies a few dB below that of the KBE. This estimate will be tagged the 'good' estimate of the BFE_W. For another noise realisation, the MSE lies much higher. This estimate will be tagged the 'bad' estimate of the BFE_W. The MSE of the BFE_I (grey dots) is high as well. The corresponding estimated FFRFs, obtained using the noisy signals, are shown in Fig. 3.

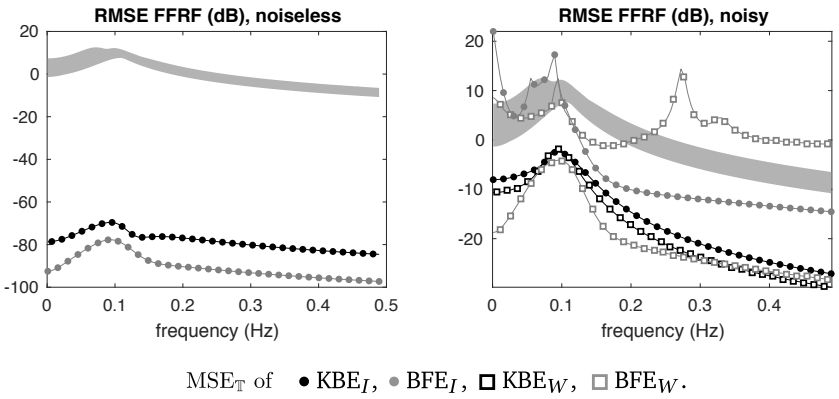

Fig. 2. *True frozen FRFs at $t \in [0, NT_s]$ (grey band), and RMS errors. Left: on noiseless data. Right: on noisy data.*
For the noisy data, two results are given for the BFE$_W$ (grey squares): one bad result (high RMSE) and one good result (low RMSE).

In Fig. 4, the MSEs, averaged over both the instantaneous time and the frequency, are given for all 100 noise realisations. This plot confirms the following.

- The KBE$_W$ performs slightly (a few dB) better than the KBE$_I$. As expected, it makes sense to include the noise variance information as prior knowledge.
- For a significant amount of noise realisations, the BFE$_W$ gives a much higher MSE than the KBE$_W$.
- For an input/output SNR of $10\,\mathrm{dB}$, the BFE$_I$ does not result in any useful estimate.

The 'bad' results of the BFE are probably due to its finite sample behaviour. As shown in Fig. 5, the variation of the estimated coefficients (black full lines) has a significantly larger amplitude than the true time-varying coefficients (dashed black-grey line). This phenomenon was not observed in the case of the KBE, which can be attributed to the associated regularisation. Note that, for the BFE$_W$, the estimator was assured not to be stuck in a local optimum, by using the true coefficients as initial estimates.

7.3. Verification of the bias and variance expressions

In Fig. 6, the estimated time-varying coefficients (black full lines) from the KBE$_I$ are given, averaged over the 100 Monte Carlo runs. The results for the KBE$_W$ (not shown) are qualitatively very similar. We observe the following.

The estimated standard deviation of $\hat{C}$, computed via Theorem 12 for *a single realisation of the data*, is denoted by the grey area. It has a very good agreement with the sample variance from the 100 Monte Carlo runs (black dashed line). Note that the kernel parameters were tuned for the first noise realisation only, and were set to $l = 183.6\,\mathrm{s}$ and $\rho = 1.56$. The same kernel parameters were used for all the Monte Carlo runs, because the expressions for the variance and the bias in Theorems 12 and 13 do not take into account the variability of the kernel parameters.

The estimated coefficients are clearly biased, as discussed in Section 4.3. The bias is approximated via Proposition 14 and by using the estimated coefficients $\hat{C}$ instead of $C_\circ$.

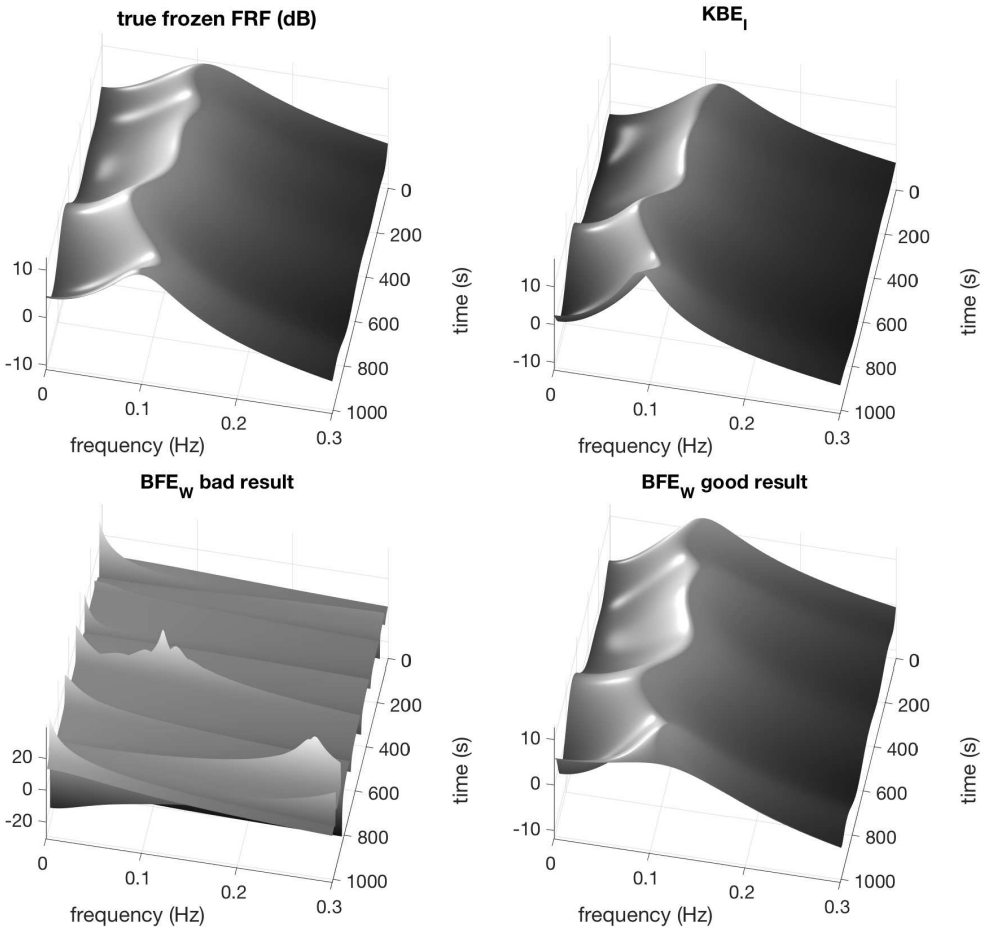

Fig. 3. *True and estimated Frozen FRFs. The result for KBE_W (not shown in the figure) is very similar to the KBE_I.*

This is used to compensate the averaged estimated time-varying coefficients, giving the black dotted line, which lies closer to the true parameters. This shows that Proposition 14 can be used to obtain an order of magnitude of the bias. Besides, the bias expression in Theorem 13 is validated by averaging the argument of the expected value in (18) by using the 100 Monte Carlo realisations. Then, the true time-varying coefficients are recovered (red full line).

The estimated coefficients are smoother than the true ones. This is the bias contribution due to the regularisation: a smoother estimate with a smaller variance is preferred to a less smooth estimate with a higher variance.

8. Conclusions

A kernel based estimator formulated in the frequency domain has been derived for identifying linear time-varying systems. The complexity of the time variation of the coefficients has been tuned via a Leave-Two-Out Cross-Validation technique. This has the advantage – over classical model order selection problems – of being a continuous optimisation problem. It has been

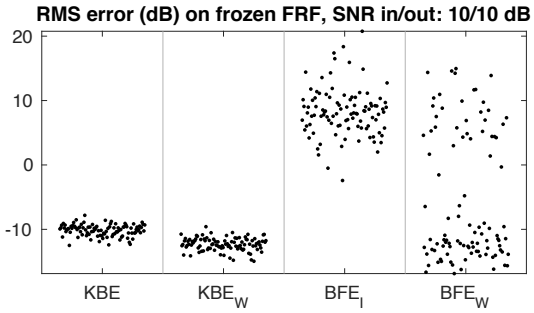

Fig. 4. *Root-mean-squared errors on the estimated frozen FRF, for 100 noise realisations (each dot corresponds to a realisation).*

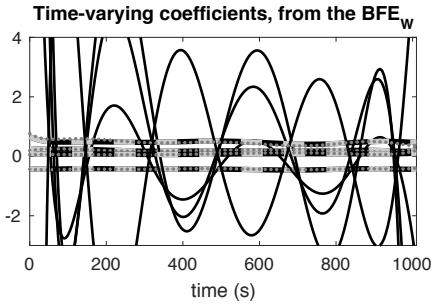

Fig. 5. *True (black/grey dashed lines) and estimated time-varying coefficients from BFE, with 'bad' results (black full lines), and 'good' results (grey) dots.*

shown on simulations that, in very noisy environments especially, the kernel based estimator manages to trade off the bias and the variance, resulting in more reliable results than a non-regularised 'Oracle' estimator (i.e. which uses the true model orders and structure). Bias and variance expressions of the kernel based estimator have been constructed, and verified on a simulation example. As such, this estimator combines system theoretic fundamentals with machine learning concepts, resulting in potentially attractive properties when linear time-varying models need to be estimated.

9. Acknowledgement

This work has been sponsored by the Research Foundation Flanders (FWO-Vlaanderen), the Flemish Government (Methusalem Fund, METH1), the Belgian Federal Government (Interuniversity Attraction Poles programme, DYSCO), and by the Netherlands Organization for Scientific Research (NWO, grant no. 639.021.127).

10. References

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, Mineola, N.Y., 1972.

[2] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, Dec 1974.

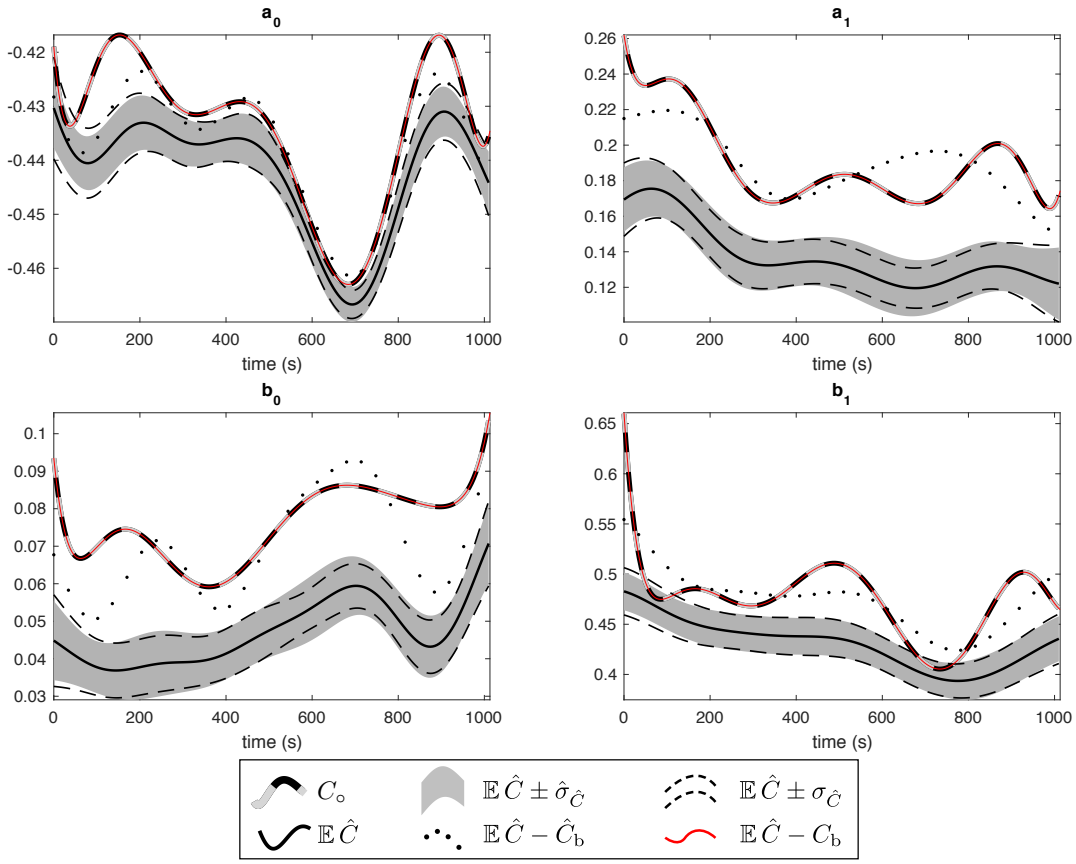

Fig. 6. *Estimated time-varying coefficients of KBE$_1$. $C_\circ$: true parameters. $\mathbb{E}\,\hat{C}$: estimated parameters using the noisy signals, averaged over 100 Monte Carlo runs. $\hat{C}_\mathrm{b}$: bias, approximated by Proposition 14, using the estimated coefficients. C_b: bias obtained via* (18) *by averaging over 100 Monte Carlo runs. The standard deviations of $\hat{C}$ are estimated with Theorem 12 from a single data set ($\hat{\sigma}_{\hat{C}}$) and computed empirically from the 100 Monte Carlo runs ($\sigma_{\hat{C}}$).*

[3] T. Breugelmans, J. Lataire, T. Muselle, E. Tourwé, R. Pintelon, and A. Hubin. Odd random phase multisine electrochemical impedance spectroscopy to quantify a non-stationary behaviour. Theory and validation by calculating an instantaneous impedance value. *Electrochimica Acta*, 76:375–382, May 2012.

[4] T. Chen, H. Ohlsson, and L. Ljung. On the estimation of transfer functions, regularizations and Gaussian processes – revisited. *Automatica*, 48(8):1525 – 1535, 2012.

[5] A. Golabi, N. Meskin, R. Tóth, and J. Mohammadpour. A Bayesian approach for estimation of linear-regression LPV models. In *Proc. of the 53rd IEEE conf. on Decision and Control (CDC)*, pages 2555–2560, Dec 2014.

[6] J. Lataire, D. Piga, and R. Tóth. Frequency-domain least-squares support vector machines to deal with correlated errors when identifying linear time-varying systems. In *Proc. of the 19th IFAC World Congress*, pages 10024–10029, 2014.

[7] J. Lataire and R. Pintelon. Frequency-domain weighted non-linear least-squares estimation of continuous-time, time-varying systems. *IET Control Theory & Applications*, 5(7):923–933, May 2011.

[8] V. Laurain, R. Tóth, M. Gilson, and H. Garnier. Direct identification of continuous-time linear parameter-varying input/output models. *IET Control Theory & Applications*, 5(7):878–888, 2011.

[9] K. Liu. Identification of linear time-varying systems. *Journal of Sound and Vibration*, 206(4):487 – 505, 1997.

[10] M. Niedzwiecki and P. Kaczmarek. Identification of quasi-periodically varying systems using the combined nonparametric/parametric approach. *IEEE Transactions on Signal Processing*, 53(12):4588–4598, Dec 2005.

[11] G. Pillonetto. Identification of time-varying systems in reproducing kernel hilbert spaces. *IEEE Transactions on Automatic Control*, 53(9):2202–2209, Oct 2008.

[12] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3):657 – 682, 2014.

[13] R. Pintelon, E. Louarroudi, and J. Lataire. Nonparametric time-variant frequency response function estimates using arbitrary excitations. *Automatica*, 51:308 – 317, January 2015.

[14] R. Pintelon and J. Schoukens. Real-time integration and differentiation of analog signals by means of digital filtering. *IEEE Transactions on Instrumentation and Measurement*, 39(6):923–927, Dec 1990.

[15] R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. John Wiley, 2nd edition, Mar 2012.

[16] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[17] C.R. Rojas, R. Tóth, and H. Hjalmarsson. Sparse estimation of polynomial and rational dynamical models. *IEEE Transactions on Automatic Control*, 59(11):2962–2977, Nov 2014.

[18] B Sanchez, E Louarroudi, E Jorge, J Cinca, R Bragos, and R Pintelon. A new measuring and identification approach for time-varying bioimpedance using multisine electrical impedance spectroscopy. *Physiological Measurement*, 34(3):339, 2013.

[19] B. Schölkopf and A. Smola. *Learning with kernels*. Cambridge MA: MIT Press, 2002.

[20] T. Söderström and P. Stoica. *System identification*. Prentice Hall, Hertfordshire, 1989.

[21] M.D. Spiridonakos and S.D. Fassois. Parametric identification of a time-varying structure based on vector vibration response measurements. *Mechanical Systems and Signal Processing*, 23(6):2029 – 2048, 2009.

[22] P. Stoica, J. Li, and T. Söderström. On the inconsistency of IQML. *Signal Processing*, 56(2):185 – 190, 1997.

[23] S. Sundararajan and S.S. Keerthi. Predictive approaches for choosing hyperparameters in Gaussian processes. *Neural Computation*, 13(5):1103–1118, May 2001.

[24] J. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific Publishing, Singapore, 2002.

[25] G. Wahba. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, 1990.

Appendix

A. Proof of Theorem 9

The constrained optimisation problem (12) is solved by defining the Lagrangian, and applying the Karush-Kuhn-Tucker conditions for optimality

$$L(C,\gamma,E,\alpha) = \frac{1}{2}E^{\mathrm{H}}W^{-1}E + \frac{1}{2}\breve{C}^{\mathrm{T}}\mathbf{K}^{-1}\breve{C} + \alpha^{\mathrm{H}}\left(Y - F\Phi C - \Psi\gamma - E\right), \tag{23}$$

$$\begin{aligned}
\partial L/\partial E = 0 \quad &\Rightarrow \quad W\alpha = E && (24)\\
\partial L/\partial \breve{C} = 0 \quad &\Rightarrow \quad \breve{C} = \mathbf{K}\Phi^{\mathrm{H}}F^{\mathrm{H}}\alpha && (25)\\
\partial L/\partial \mathring{C} = 0 \quad &\Rightarrow \quad 0 = \mathring{\Phi}^{\mathrm{H}}F^{\mathrm{H}}\alpha && (26)\\
\partial L/\partial \gamma = 0 \quad &\Rightarrow \quad 0 = \Psi^{\mathrm{H}}\alpha && (27)\\
\partial L/\partial \alpha = 0 \quad &\Rightarrow \quad E = Y - F\Phi C - \Psi\gamma && (28)
\end{aligned}$$

By plugging (24) and (25) into (28) we have

$$Y = \left(F\Phi\mathbf{K}\Phi^{\mathrm{H}}F^{\mathrm{H}} + W\right)\alpha + F\mathring{\Phi}\mathring{C} + \Psi\gamma \tag{29}$$

which, combined with (25), (26) and (27), gives (13). Note that (12) and (29) are often called the *primal* and *dual* formulations of the optimisation problem respectively, see [24]. Since (12a) is a convex objective function and (12b) is affine in the *primal parameters* C and γ, the dual optimisation problem (29) provides the same optimum as (12) (i.e., there is no duality gap).

B. Covariance and bias of estimated coefficients

Denote the minimisers of the expected value of the cost function:

$$\left[\tilde{C},\tilde{\gamma}\right] \triangleq \underset{C,\gamma}{\operatorname{argmin}}\, \mathbb{E}\left\{V\left([C,\gamma]\right)\right\}, \tag{30}$$

$$\text{with } V([C,\gamma]) \triangleq E^{\mathrm{H}}W^{-1}E + \breve{C}^{\mathrm{T}}\mathbf{K}^{-1}\breve{C}, \tag{31}$$

and E given in (12b). The shorthand notation $[C,\gamma] \triangleq [\breve{C}^{\mathrm{T}}\ \mathring{C}^{\mathrm{T}}\ \gamma^{\mathrm{T}}]^{\mathrm{T}}$ is used. Denote $[\hat{C},\hat{\gamma}] = [\tilde{C},\tilde{\gamma}] + \delta$ the minimiser of V for a given noisy data set. Since $[\tilde{C},\tilde{\gamma}]$ is a deterministic

vector, we have that $\mathrm{cov}(\delta) = \mathrm{cov}([\hat{C}, \hat{\gamma}])$. By letting δ be the solution of $V' \left(\left[\tilde{C}, \tilde{\gamma} \right] + \delta \right) = 0$, we have $\delta^{\mathrm{H}} = -V' \left([\tilde{C}, \tilde{\gamma}] \right) V''^{-1}$, with

$$\delta = -V''^{-1} \left(\begin{bmatrix} \Phi^{\mathrm{H}} F^{\mathrm{H}} \\ \mathring{\Phi}^{\mathrm{H}} F^{\mathrm{H}} \\ \Psi^{\mathrm{H}} \end{bmatrix} W^{-1} E + \begin{bmatrix} \mathbf{K}^{-1} \tilde{\breve{C}} \\ 0 \\ 0 \end{bmatrix} \right), \tag{32}$$

$$V'' = \left(\begin{bmatrix} \Phi^{\mathrm{H}} F^{\mathrm{H}} \\ \mathring{\Phi}^{\mathrm{H}} F^{\mathrm{H}} \\ \Psi^{\mathrm{H}} \end{bmatrix} W^{-1} \begin{bmatrix} F\Phi & F\mathring{\Phi} & \Psi \end{bmatrix} + \begin{bmatrix} \mathbf{K}^{-1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right). \tag{33}$$

Then, by approximating $\mathbb{E}\{V''^{-1}\bullet\} \approx V''^{-1}\mathbb{E}\{\bullet\}$ and by neglecting the effect of the correlation between the noise in Φ and E compared with terms which include $\Phi_\circ$ and $\mathring{\Phi}_\circ$, we have

$$\delta - \mathbb{E}\{\delta\} \approx -V''^{-1} \begin{bmatrix} \Phi^{\mathrm{H}} F^{\mathrm{H}} \\ \mathring{\Phi}^{\mathrm{H}} F^{\mathrm{H}} \\ \Psi^{\mathrm{H}} \end{bmatrix} W^{-1} (E - \mathbb{E}\{E\}). \tag{34}$$

This approximation is valid as long as the noise is significantly smaller than the noiseless signal. The covariance expression in Theorem 12 is obtained by computing $\mathrm{cov}(\delta)$ with (34), where E is evaluated for $[\tilde{C}, \tilde{\gamma}]$, and by rearranging the expression such that $\mathbf{K}^{-1}$ and the matrix multiplications in V'' (leading to an important loss of precision) do not need to be computed. The bias expression (18) is obtained as $[C_{\mathrm{b}}, \gamma_{\mathrm{b}}] = \mathbb{E}\{\delta\}$, where $[\tilde{C}, \tilde{\gamma}]$ is replaced by $[C_\circ, \gamma_\circ]$ (the true parameters) in (32), rearranged such that $\mathbf{K}^{-1}$ should not be computed.

C. Simulation details

The noise filters are given by

$$H_y(z^{-1}) = s_y z^{-2} / \left(1 - 1.65 z^{-1} + 0.90 z^{-2}\right), \tag{35}$$

$$H_u(z^{-1}) = s_u z^{-2} / \left(1 - 1.38 z^{-1} + 0.72 z^{-2}\right), \tag{36}$$

where s_y and s_u are scaling factors chosen such that the specified SNRs are obtained. The time-varying coefficients in (1) of the simulated system are given by $a_n(t) = \sum_{p=0}^{8} a_{n,p} f_p(t)$, $b_n(t) = \sum_{p=0}^{8} b_{n,p} f_p(t)$, where $f_p(t) = P_p\left((2t)/(NT_{\mathrm{s}}) - 1\right)$, P_p is the pth degree Legendre polynomial, and $a_{n,p}$ and $b_{n,p}$ are given in Table 1. The system was simulated with the ode45 solver from Matlab.

Table 1 Coefficients of $a_n(t)$ and $b_n(t)$ of the simulated system.

	$a_{n,p} \times 1000$			$b_{n,p} \times 1000$	
	$n=0$	1	2	$n=0$	1
$p=0$	-433.92	190.98	-1000	75.989	477.47
1	-8.157	-27.02	0	10.441	-30.077
2	13.238	33.317	0	8.3018	31.981
3	15.994	-19.641	0	-10.203	3.3666
4	-8.5959	-13.264	0	3.9055	65.955
5	-19.79	-0.54219	0	10.916	-62.18
6	-13.706	-16.247	0	10.237	-29.306
7	-1.4546	-5.3577	0	-11.389	-46.701
8	22.217	32.973	0	7.9717	54.298