Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Frequency-domain least-squares support vector machines to deal with correlated errors when identifying linear time-varying systems ⋆

J. Lataire * **D. Piga** ** **R. Tóth** **

* *Dept. ELEC (Fundamental Electricity and Instrumentation), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium; e-mail: jlataire@vub.ac.be*
** *Control Systems Group, Dept. of Electrical Engineering, Eindhoven University of Technology, Postbus 513, 5600 MB Eindhoven, The Netherlands.*

Abstract: A Least-Squares Support Vector Machine (LS-SVM) estimator, formulated in the frequency domain is proposed to identify linear time-varying dynamic systems. The LS-SVM aims at learning the structure of the time variation in a data driven way. The frequency domain is chosen for its superior robustness w.r.t. correlated errors for the calibration of the hyper parameters of the model.
The time-domain and the frequency-domain implementations are compared on a simulation example to show the effectiveness of the proposed approach. It is demonstrated that the time-domain formulation is mislead during the calibration due to the fact that the noise on the estimation and calibration data sets are correlated. This is not the case for the frequency-domain implementation.

1. INTRODUCTION

In engineering applications and, more specifically, in the systems and control field, a lot of effort has been spent to extend the comfortable framework of Linear Time-Invariant (LTI) systems to the more hazardous Linear Time-Varying (LTV) systems, Linear Parameter-Varying (LPV) systems and, more generally, to nonlinear systems. The major reason is the need to achieve better performance in modeling and controlling systems with nonlinear and time-varying dynamics, than can be provided by the LTI framework.

The present paper focuses on the data-driven identification of LTV systems. Practical applications of LTV systems include:

- the dynamics of the wing of a plane, which vary with the flight speed and altitude [Fujimori and Ljung, 2006],
- the electrical impedance of a metal, subject to (pitting) corrosion [Tourwé et al., 2010],
- the mechanical impedance of a human muscle, under varying spinal signals [Groenewegen et al., 2012].

From a black-box identification point of view, it is important to determine the structure of the time variation directly *from the data*, without limiting ourselves to restrictive assumptions. The above mentioned applications involve seriously complex dependencies on time, which can not be determined explicitly in advance. An acceptable assumption made in this paper is that the variation is smooth, to some extent. Still, the "level of smoothness" must be determined. To do that in a convenient fashion, a Least-Squares Support Vector Machines (LS-SVM) based approach is proposed in this paper to model the time-varying system parameters.

LS-SVMs are versatile estimators, with a continuously tunable smoothness. This tuning is performed by optimizing a calibration criterion (cross-validation), which assesses that the model is able to predict the system behavior on a data set which was not used for estimation. The fact that the calibration criterion is continuous w.r.t. the parameter to be tuned prevents the complexity of the problem to be combinatorial.

Two specific difficulties pop up w.r.t. the identification of time-varying, dynamic systems:

(1) It is difficult, and sometimes impossible, to repeat experiments since the system varies from one experiment to the other.
(2) The equation-error in the system equation is prone to be correlated, because of the dynamic nature of the system.

As a consequence, it is impossible to split the available data into an estimation data set and a calibration data set, affected by *different noise realizations*. The estimation data and the calibration data will have to be interleaved, such that the errors on the one will be correlated with those on the other data set. This limiting problem will

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

be overcome by formulating the problem in the frequency domain.

A frequency-domain formulation of system identification problems has proven advantages, one of which is that the *spectrum* of stationary, colored noise is *not correlated.* In other words:

correlated, stationary noise in the time domain

$$\Updownarrow$$

not correlated noise spectrum

This gives the motivation to develop the calibration of the model in the frequency domain. This paper aims at demonstrating that the frequency-domain formulation of the LS-SVM to identify dynamic time-varying systems can yield an impressive improvement w.r.t. the time-domain formulation if cross-validation based calibration of the hyper parameters is applied.

1.1 State of the art

The state of the art which is in direct relevance with the current paper is summarized as follows.

- Exhaustive and very detailed authorities providing frequency domain techniques for system identification is Pintelon and Schoukens [2012], and discussing LS-SVM is Suykens et al. [2002].
- Frequency domain techniques dealing with correlated noise in an elegant way are discussed in Pintelon et al. [2006].
- A time-domain formulation to deal with correlated errors in the LS-SVM setting is discussed in De Brabanter et al. [2011] and Laurain et al. [2011]. More precisely, advanced tools are used in De Brabanter et al. [2011] to adapt the kernel function to the correlation of the errors, while an *instrumental-variable* based approach is introduced in Laurain et al. [2011].
- A frequency domain formulation of the LS-SVM to solve the regression problem of a *static* time function in the presence of correlated errors has been introduced in Lataire et al. [2012]. This has been done without making assumptions on the specific nature of the correlation.

1.2 Novel contributions

The novel contribution of this paper is that the LS-SVM is formulated in the *frequency domain* for *dynamic time-varying systems*, and implemented for a time-varying ARMAX structure. As explained in the paper, the frequency domain formulation intrinsically yields a different pattern for splitting the available data into estimation and validation data sets. It is more involved than simply taking the DFT of the signals. It is then demonstrated that the frequency domain formulation is a highly valuable alternative to its time domain counterpart, and is especially more robust in the presence of correlated errors.

2. PROBLEM FORMULATION

Consider a linear time-varying system described by the ordinary difference equation:

$$y(t) = -\sum_{n=1}^{N_a} a_n(t)y(t-n) + \sum_{n=0}^{N_b} b_n(t)u(t-n) + v(t) \quad (1)$$

where $u(t)$ and $y(t)$ denote the input/output signals respectively, and $a_n(t)$ and $b_n(t)$ are the time-varying system coefficients. The equation error $v(t)$ is discrete-time colored stationary noise, *viz.*:

Assumption 1. [Colored stationary noise] For any $t \in \mathbb{Z}$

$$\mathbb{E}\{v(t)\} = 0,$$
$$\mathbb{E}\{v(t)v(t+\tau)\} \equiv C_v(\tau). \qquad \blacksquare$$

The fact that C_v only depends on τ makes it *stationary.* The fact that $C_v(\tau) \neq 0$ for $\tau \neq 0$ makes the noise *correlated.* Denote $\mathbb{T} = \{0, 1, \ldots, N-1\}$ the considered measurement window, inside which a data set is available.

Assumption 2. (Available data set). A single contiguous input/output data set $\mathcal{D}_N = \{(y(t), u(t))\}_{t\in\mathbb{T}}$ is available.

The problem is formulated as follows. Determine the time-varying coefficients $a_n(t)$ and $b_n(t)$ for $t \in \mathbb{T}$ without a priori knowledge on their shape. In other words, write the system coefficients as function expansions:

$$a_n(t) = \rho_{a_n}^\top \phi(t), \quad b_n(t) = \rho_{b_n}^\top \phi(t), \quad \text{for } t \in \mathbb{T} \quad (2)$$

where $\rho_{a_n}, \rho_{b_n}, \phi(t) \in \mathbb{R}^{n_H}$. $\phi(t)$ are potentially infinite ($n_H \to \infty$) dimensional basis functions. ρ_{a_n} and ρ_{b_n} are called the primal parameters, $\phi(t)$ is a (potentially infinite) vector of basis functions. The fact that $n_H \to \infty$ will require a regularization on the primal parameter vectors.

3. SOLUTION FORMULATION

3.1 Reasoning

The following reasoning is made:

(1) To avoid overfitting, this model must be *tuned* based on a *calibration criterion.*
(2) Since one single data set is available, the calibration data set will have to be *interleaved* with the data set used for the estimation.
(3) The fact that $v(t)$ is correlated over time (Assumption 1) yields a *correlation* between the noise on the calibration and the estimation data sets. This will mislead the calibration of the model.
(4) Since the *Discrete Fourier Transform* (DFT) of stationary noise is not correlated, the calibration of the model in the frequency domain formulation will not be mislead by the time correlated noise.

3.2 Time domain estimation: implementation

Introduce the following notations:

$$\begin{aligned} \rho_n &\equiv \rho_{b_n}, & x_n(t) &\equiv u(t-n), & n &= 0, \ldots, N_b \\ \rho_{n+N_b} &\equiv \rho_{a_n}, & x_{n+N_b}(t) &\equiv -y(t-n), & n &= 1, \ldots, N_a \end{aligned} \quad (3)$$

with $u(t) = y(t) = 0$ for $t < 0$. In combination with (2), rewrite (1) for $t \in \mathbb{T}$ as

$$y(t) = \sum_{n=0}^{N_a+N_b} \rho_n^\top \phi(t) x_n(t) + I(t) + v(t), \quad t \in \mathbb{T} \quad (4)$$

where $I(t) = \sum_{n=0}^{N_{\text{tr}}} \delta(t-n) tr_n$ takes into account the initial conditions ($\delta(\bullet)$ is the Kronecker delta; $N_{\text{tr}} =$

$\max(N_a, N_b) - 1)$ [Lataire and Pintelon, 2011]. Note that the coefficients tr_n must also be estimated. The estimation problem is formulated as follows in the *time domain*:

$$\hat{\rho}_n = \underset{\rho_n, e, \theta_{\mathrm{tr}}}{\operatorname{argmin}} \frac{1}{2}\rho^\top \rho + \frac{\gamma}{2} \sum_{t \in \mathbb{T}_\mathrm{e}} e^2(t), \tag{5a}$$

$$\text{s.t. } e(t) = y(t) - \sum_{n=0}^{N_a+N_b} \rho_n^\top \phi(t) x_n(t) - I(t), \tag{5b}$$

where $\mathbb{T}_\mathrm{e}$ is the set of time samples used for estimation, θ_{tr} is a column vector stacking tr_n, for $n = 0, 1, \ldots, N_{\mathrm{tr}}$, and γ is a regularization parameter tuned by the user to balance the bias/variance trade-off in the final estimate of the time-varying parameters $a_n(t)$ and $b_n(t)$.

Problem (5) is derived from the standard 'primal formulation' of Least Squares Support Vector Machines (LS-SVM) [Suykens et al., 2002]. In order to compute the parameters $a_n(t)$ and $b_n(t)$, the *Lagrangian* $\mathcal{L}(\rho_n, e, \theta_{\mathrm{tr}}, \alpha)$ associated with (5) is constructed:

$$\mathcal{L}(\rho_n, e, \theta_{\mathrm{tr}}, \alpha) = \frac{1}{2}\rho^\top \rho + \frac{\gamma}{2} \sum_{t \in \mathbb{T}_\mathrm{e}} e^2(t) + \tag{6}$$

$$+ \sum_{t \in \mathbb{T}_\mathrm{e}} \alpha(t) \left(y(t) - \sum_{n=0}^{N_a+N_b} \rho_n^\top \phi(t) x_n(t) - I(t) - e(t) \right),$$

where $\alpha(t)$ are the so-called *Lagrangian multipliers* and they are stacked in the vector $\alpha = \{\alpha(t) : t \in \mathbb{T}_\mathrm{e}\}$. From the *Karush-Kuhn-Tucker* (KKT) optimality conditions, the Lagrangian multipliers α and the parameters θ_{tr} can be computed by solving the set of linear equations:

$$\begin{bmatrix} y \\ 0_{N_{\mathrm{tr}} \times 1} \end{bmatrix} = \begin{bmatrix} \Lambda + 1/\gamma & \psi \\ \psi^\top & 0_{N_{\mathrm{tr}}} \end{bmatrix} \begin{bmatrix} \alpha \\ \theta_{\mathrm{tr}} \end{bmatrix}, \tag{7}$$

where the vector y stacks $y(t)$ for $t \in \mathbb{T}_\mathrm{e}$, and ψ is the identity matrix, truncated to the first N_{tr} columns. Eq. (7) is known as the 'dual formulation'. The matrix Λ is obtained by stacking $\Lambda(t, t')$ for $t, t' \in \mathbb{T}_\mathrm{e}$, with

$$\Lambda(t, t') = \sum_{n=0}^{N_a+N_b} x_n(t) \underbrace{\phi^\top(t)\phi(t')}_{\equiv K(t,t')} x_n(t'), \tag{8}$$

In (8), $K(t, t')$ is a positive definite *Kernel function* $K : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfying the Mercer's condition [Mercer, 1909] and defining the inner product $\phi^\top(t)\phi(t')$. The choice of the kernel $K(t, t')$ is discussed further on. From the KKT conditions and (2), the system coefficients can be computed in terms of the kernel function $K(t, t')$ and α:

$$\hat{b}_n(t) = \sum_{t' \in \mathbb{T}_e} K(t, t') x_n(t') \alpha(t'), \quad n = 0, 1, \ldots, N_b \tag{9}$$

$$\hat{a}_n(t) = \sum_{t' \in \mathbb{T}_e} K(t, t') x_{n+N_b}(t') \alpha(t'), \quad n = 1, \ldots, N_a \tag{10}$$

The reader is referred to Suykens et al. [2002] for further details on the LS-SVM approach.

3.3 *Time domain: discussion*

- The parameters ρ_n have been eliminated from the problem and, thus, are not computed explicitly. The resulting $\hat{a}_n$ and $\hat{b}_n$ in (9) and (10) are non-parametric estimates of the system coefficients.
- As is clear from (8), $\phi(t)$ is not required explicitly to solve (7). Only the scalar product $\phi^\top(t)\phi(t') \equiv K(t, t')$ is needed. $K(t, t')$ is known as the kernel. It implicitly defines the basis functions used. Specification of the kernel $K(t, t')$ instead of the maps $\phi(t)$ is called the *kernel trick* [Vapnik, 1998], and allows the use of infinite dimensional bases.
- A commonly used kernel to model arbitrary signals is the Radial Basis Function (RBF), given by
$$K(t, t') = \exp\left(-\frac{\|t - t'\|_2^2}{\sigma^2}\right). \tag{11}$$
Although other kernels can be used with the presented methods, the RBF will be used throughout this paper.
- The RBF depends on the tunable parameter σ, which determines the smoothness of the estimated time-varying coefficients. The tunable parameter γ in (5a) defines a trade-off between a bias error and the variance of the estimated parameters. σ and γ will be calibrated as explained further on.
- As indicated in (5a), the estimation is done using only a subset $\mathbb{T}_\mathrm{e} \subset \mathbb{T}$ of the available data set. The complement of that data set, $\mathbb{T}_\mathrm{c} = \mathbb{T} \backslash \mathbb{T}_\mathrm{e}$, will be used for tuning the hyper-parameters γ and σ via cross-validation. The choice of $\mathbb{T}_\mathrm{e}$ and $\mathbb{T}_\mathrm{c}$ will be discussed later.

3.4 *Frequency domain estimation: implementation*

Denote $\mathcal{F}_t\{x(t)\}$ the Discrete Fourier Transform (DFT) of $x(t)$ w.r.t. t at frequency bin k:

$$X(k) = \mathcal{F}_t\{x(t)\}_k = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x(t) e^{-\frac{j2\pi kt}{N}}, \; k \in \mathbb{K}, \tag{12}$$

with $\mathbb{K} = \{0, 1, \ldots, N-1\}$ denoting the set of frequency bins. By following the reasoning in Lataire and Pintelon [2011] and Lataire [2011], and by using (3), the model equation (1) can be written in the spectral domain as

$$Y(k) = \sum_{n=0}^{N_a+N_b} \rho_n^\top \mathcal{F}_t\{\phi(t) x_n(t)\}_k + I(z_k) + V(k), \tag{13}$$

where $I(z_k) = \sum_{n=0}^{N_{\mathrm{tr}}} tr_n z_k^{-n}$ takes into account the initial conditions, $z_k = \exp(j2\pi k/N)$, and $k \in \mathbb{K}$. Analogously to the time domain formulation (5), the frequency domain estimation problem is formulated as the following constrained optimization

$$\hat{\rho}_n = \underset{\rho_n, E, tr_n}{\operatorname{argmin}} \frac{1}{2}\rho^\top \rho + \frac{\gamma}{2} \sum_{k \in \mathbb{K}_\mathrm{e}} |E(k)|^2, \tag{14}$$

$$\text{s.t. } E(k) = Y(k) - \sum_{n=0}^{N_a+N_b} \rho_n^\top \mathcal{F}_t\{\phi(t) x_n(t)\}_k - I(z_k),$$

where $\mathbb{K}_\mathrm{e}$ is the subset of frequency bins $\mathbb{K}_\mathrm{e} \subset \mathbb{K}$ at which the estimation is performed. The selection of $\mathbb{K}_\mathrm{e}$ will be discussed later. The additional difficulty compared with solving (5) is that (14) is a complex equation. Nevertheless, it can be rewritten as a real-valued optimization problem by decomposing (14) into its real and imaginary parts. Through considerations similar to the ones discussed in the time-domain case Section 3.2, the *Lagrangian* associated with the primal problem (14) can be constructed and the solution of the dual problem of (14) can be computed by solving the set of linear equations:

$$\begin{bmatrix} Y \\ 0_{N_{\mathrm{tr}}\times 1} \end{bmatrix} = \begin{bmatrix} \Lambda_{\mathrm{F}} + 1/\gamma & \Psi \\ \Psi^\top & 0_{N_{\mathrm{tr}}} \end{bmatrix} \begin{bmatrix} \alpha_{\mathrm{F}} \\ \theta_{\mathrm{tr}} \end{bmatrix}, \tag{15}$$

for α_{F} and θ_{tr}, where Y and α_{F} stack the output spectrum $Y(k)$ and the frequency domain *Lagrangian multipliers* $\alpha_{\mathrm{F}}(k)$ for $k \in \mathbb{K}_{\mathrm{e}}$, respectively. The proofs for deriving (15) are outside the scope of this paper. In Lataire et al. [2012], the proofs are provided for static systems. The complex valued matrix Λ_{F} is obtained by stacking $\Lambda_{\mathrm{F}}(k, k')$ for $k, k' \in \mathbb{K}_{\mathrm{e}}$, with

$$\begin{aligned} \Lambda_{\mathrm{F}}(k,k') &\equiv \sum_{n=0}^{N_a+N_b} \mathcal{F}_t^\top \left\{\phi(t)x_n(t)\right\}_k \overline{\mathcal{F}_t \left\{\phi(t)x_n(t)\right\}_{k'}} \\ &= \sum_{n=0}^{N_a+N_b} \mathcal{F}_t \left\{ \overline{\mathcal{F}_{t'} \left\{x_n(t)K(t,t')x_n(t')\right\}_{k'}}, \right\}_k \end{aligned} \tag{16}$$

(an overline denotes a complex conjugation). The estimated system coefficients are then given by:

$$\hat{b}_n(t) = \sum_{k\in\mathbb{K}_{\mathrm{e}}} \overline{\mathcal{F}_{t'} \left\{K(t,t')x_n(t')\right\}_k} \alpha_{\mathrm{F}}(k), \quad n = 0, 1, \ldots, N_b$$

$$\hat{a}_n(t) = \sum_{k\in\mathbb{K}_{\mathrm{e}}} \overline{\mathcal{F}_{t'} \left\{K(t,t')x_{n+N_b}(t')\right\}_k} \alpha_{\mathrm{F}}(k), \quad n = 1, \ldots, N_a.$$

3.5 Frequency domain: discussion

Similarities and differences with the time domain implementation (see Section 3.3) are listed.

Similarities with the time-domain implementation:

- In this frequency-domain implementation, the time-domain kernel $K(t, t')$ appears explicitly. Using the same kernel as in the time-domain case allows for a fair comparison between the two implementations.
- The hyper parameters γ and σ should also be tuned.
- By Parceval's theorem, for the same values of σ and γ, and if the estimation data sets consist of the whole available data set (i.e. $\mathbb{T}_{\mathrm{e}} = \mathbb{T}$ and $\mathbb{K}_{\mathrm{e}} = \mathbb{K}$), the time domain and frequency domain formulations (5) and (14) are exactly equivalent and, thus, they have the same solution.

Difference with the time domain implementation

- The available total data set ($k \in \mathbb{K}$) is subdivided into an estimation ($\mathbb{K}_{\mathrm{e}}$) and a calibration data set ($\mathbb{K}_{\mathrm{c}} = \mathbb{K} \setminus \mathbb{K}_{\mathrm{e}}$). This subdivision is done in the frequency domain. That is, instead of the available *time samples*, the available *frequency bins* are subdivided. This implies that the estimated parameters from the time domain estimation data set are different from the estimated parameters from the frequency domain estimation data set. This will be important in the next section.

4. CALIBRATION BASED ON CROSS-VALIDATION

The following reasoning is made:

- For both the time- and frequency-domain implementations, the data sets must be split into estimation ($\mathbb{T}_{\mathrm{e}}$ and $\mathbb{K}_{\mathrm{e}}$) and calibration data sets ($\mathbb{T}_{\mathrm{c}}$ and $\mathbb{K}_{\mathrm{c}}$).
- The hyper parameters γ and σ must be tuned, based on the capability of the *estimated* model to predict the output signal/spectrum at the *calibration* samples/bins.
- The optimization of this prediction capability is formulated as the minimization of the following calibration objective functions (time and frequency domain respectively)

$$\gamma_{\mathrm{T}}, \sigma_{\mathrm{T}} = \underset{\gamma,\sigma}{\operatorname{argmin}} \sum_{t\in\mathbb{T}_{\mathrm{c}}} \left(y(t) - \hat{y}(t,\gamma,\sigma)\right)^2 \tag{17a}$$

$$\gamma_{\mathrm{F}}, \sigma_{\mathrm{F}} = \underset{\gamma,\sigma}{\operatorname{argmin}} \sum_{k\in\mathbb{K}_{\mathrm{c}}} \left|Y(k) - \hat{Y}(k,\gamma,\sigma)\right|^2 \tag{17b}$$

  where $\hat{y}$ and $\hat{Y}$ were estimated from, respectively, the time domain and frequency domain estimation data sets.
- For the considered model class, and since only one data set is available, it wouldn't make sense to use, for instance, the first half of the samples/bins for the estimation and the second half for the calibration, since no structure is imposed on the time-varying coefficients $a_n(t)$ and $b_n(t)$. This is because the system behavior *cannot be extrapolated* from the first half to the second half. For that reason, the estimation and calibration data sets must be *interleaved*, assuming the smoothness of $a_n(t)$ and $b_n(t)$.
- The choice of the interleaved estimation and calibration samples/bins is done as depicted below

  $\cdots \circ \bullet \bullet \bullet \circ \bullet \bullet \bullet \circ \bullet \bullet \bullet \circ \bullet \bullet \bullet \circ \ldots$

  with $\bullet$ an estimation sample/bin
  $\circ$ a calibration sample/bin

  This can straightforwardly be extended to an n-fold cross-validation scheme (with $n = 4$ in this case) by constructing multiple shifted versions of the above estimation and calibration sets.
- In the calibration objective functions (17), terms will appear containing the following products (as shown in the Appendix of De Brabanter et al. [2011]):

$$\text{Time domain: } v(t)v(t'), \qquad t \in \mathbb{T}_{\mathrm{e}},\ t' \in \mathbb{T}_{\mathrm{c}}$$

$$\text{Frequency domain: } V(k)\overline{V(k')}, \quad k \in \mathbb{K}_{\mathrm{e}},\ k' \in \mathbb{K}_{\mathrm{c}}$$

  By Assumption 1, the product of two noise samples in the time domain is non-zero in expected value. Also by Assumption 1, and since the DFT of stationary noise is *not correlated* over the frequency [Pintelon et al., 2006], the expected value of the noise spectrum at different bins *is* zero. Summarized:

$$\mathbb{E}\left\{v(t)v(t')\right\} = C_v(t - t') \neq 0 \tag{18a}$$

$$\mathbb{E}\left\{V(k)V(k')\right\} = 0 \tag{18b}$$

The reasoning above leads to the conclusion summarized as the following theorem.

Theorem 1. The time-domain calibration objective function in (17a) contains non-zero terms which are purely due to the noise, and also *depend on* γ and σ. These terms are prone to *shift the global minimum* of the calibration objective function, yielding suboptimal results. Similar terms don't appear in the frequency-domain calibration objective function (17b).

As such, the frequency domain formulation of the LS-SVM provides a more efficient way of calibrating the hyper parameters. The effects of Theorem 1 will be demonstrated on a simulation example.

5. GENERAL DISCUSSION

5.1 Advantages

- LS-SVMs have the versatility to model any smooth function. Translated to the identification of LTV systems, any type of time-variation can be captured.
- To avoid under/overfitting, the hyper parameters γ and σ must be tuned. Unlike the model order selection of conventional basis functions (like polynomials, splines), the calibration of the kernel (RBF, σ) and the bias variance trade-off (γ) is *continuous* in the hyper parameters.
- The presented frequency domain implementation of the LS-SVM is robust to correlated errors in the equation error.
- The presented LS-SVM estimator and its calibration is meant to be applicable, even if only a single data set is available.

5.2 Limitations and future improvements

- The model structure in (1) is fairly artificial. A more realistic model structure, in particular w.r.t. the noise structure, would allow $v(t)$ to be *non-stationary*, in addition to be colored. This extension would make the spectrum $V(k)$ correlated as well, thus compromising the effect of Theorem 1, since (18b) would not be valid. However, for a *slow variation* of $a_n(t)$ and $b_n(t)$, the correlation in time is expected to exceed the non-stationarity, thus still favoring the frequency domain implementation. This will be the case in the simulation example.

  Note that the adverse is equally possible: if $v(t)$ is an uncorrelated sequence, but highly non-stationary, the time domain implementation will outperform the frequency domain one.
- Solving (7) and (15) requires the inversion of square matrices, with dimensions as large as the estimation data set. This is computationally very demanding. In addition, for the calibration of the hyperparameter, this inversion must be performed for every evaluation of the calibration objective function.
- In (13), the spectrum of the noise $V(k)$ is not identically distributed over k (that is, the noise is colored). This means that a further refinement of (14) can be performed by including a k-dependent weight to the terms $|E(k)|^2$. In other words, the residuals are to be whitened, thus decreasing the variance of the estimated parameters, analogously to the Markov estimator (e.g. Section 4.3 in Söderström and Stoica [1989]).

6. SIMULATION EXAMPLE

6.1 Simulated system

Unlike the system model in this paper, the system is in an output error framework (which is an extension of (1) and, as discussed in Section 5.2, the spectrum of the equation error might be slightly correlated). Still the improvement of the frequency-domain implementation over the time-domain one will be clearly visible. This is probably thanks to the slow variation of the system coefficients, w.r.t. the typical time constants of the system dynamics.

The true system is described by an output-error structure:

$$y_\circ(t) = -\sum_{n=1}^{N_a} a_n(t) y_\circ(t-n) + \sum_{n=0}^{N_b} b_n(t) u(t-n), \quad \text{(19a)}$$

$$y(t) = y_\circ(t) + v(t), \quad \text{(19b)}$$

with $v(t)$ a colored, stationary noise sequence, described by

$$v(t) = 1.6578 v(t-1) - 0.8464 v(t-2) + e(t), \quad (20)$$

($e(t)$ is white gaussian noise) and normalized such as to yield a Signal to Noise Ratio (SNR) of 20 dB of the measured output signal. One data set was acquired, counting $N = 2048$ input/output samples. The system coefficients $a_n(t)$ and $b_n(t)$ were cubic splines with 4 equidistant breakpoints (including the boundaries), with $N_a = 3$ and $N_b = 2$, yielding an evolution of the instantaneous poles and zeros [1] as depicted in Figure 1, left.

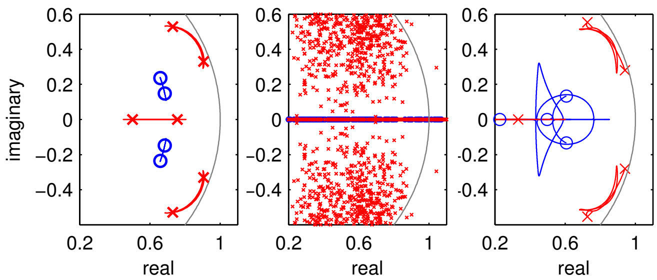

Fig. 1. Left: Evolution of the true instantaneous poles and zeros of the simulated system. The crosses denote the poles at the beginning and the end of the considered time window. The zeros are analogously given by the circles. Middle and right: Estimated pole zero evolution for $\sigma = 3$ (middle) and $\sigma = 1000$ (right), for $\gamma = 10^5$.

6.2 Estimation and calibration of the hyper parameters

The system is estimated by solving (7) for the time domain formulation and (15) for the frequency domain formulation. The hyper parameters γ and σ are obtained as the minimizers in (17).

The distinction of the time domain and frequency domain estimators is observed in the calibration objective function w.r.t. σ, as depicted in Figure 2. The following observations are made:

- An *ideal* calibration objective function is introduced (ideal TD/FD), given by the RMS of the true error on the estimated parameters. Ideally, the minimum of that objective function should be sought. In this case, that minimum lies at $\sigma = 5700$. That function is, of course, not known in practice, since the true system parameters are unknown.

[1] The instantaneous poles and zeros at a time instant t^* are defined as the poles and zeros of the LTI systems $G(t^*)$ that one obtains by freezing the system coefficients at that time instant $G(t^*) : y_\circ(t) = -\sum_{n=1}^{N_a} a_n(t^*) y_\circ(t-n) + \sum_{n=0}^{N_b} b_n(t^*) u(t-n)$.

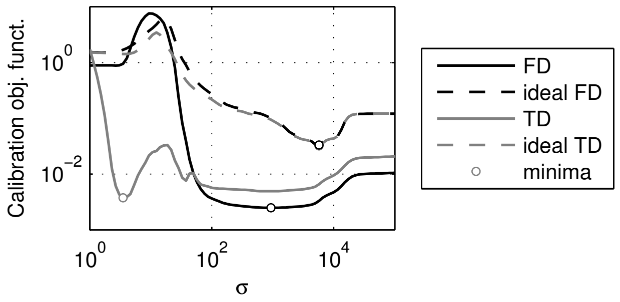

Fig. 2. Calibration objective functions, for the time-domain (TD) and frequency-domain (FD) formulations as a function of σ, with $\gamma = 10^5$.

- The calibration objective function of the time domain formulation (TD) has a global minimum at $\sigma_T = 3$. This shifted minimum, w.r.t. the ideal one is caused by the cross-correlation of the errors in the estimation and calibration data sets.
- The global minimizer $\sigma_F = 1000$ of the frequency domain calibration function (FD) is clearly much closer to the global minimum of the ideal one. The pronounced local optimum around $\sigma = 3$ is not present here. This reveals its robustness w.r.t. correlated errors.

6.3 Comparison of the resulting estimated models

Based on the global minimizers of the calibration objective functions in Figure 2, $\sigma_T = 3$ and $\sigma_F = 1000$ were selected, and the corresponding models were estimated. The calibration objective functions were observed to be very insensitive to the values of γ. This is, therefore, not further discussed. The values $\gamma_T = \gamma_F = 10^5$ were used.

Figure 1, middle and right, shows the evolution of the instantaneous poles and zeros of the estimated models, and are to be compared with the true ones, in Figure 1, left.

- *Figure 1, middle, $\sigma = 3$, from time domain calibration* it is clear that σ is too small, yielding a very roughly varying model. The instantaneous poles and zeros are scattered in a large area.
- *Figure 1, right, $\sigma = 1000$, from frequency domain calibration.* A very good agreement is obtained for the resonating poles, a discrepancy exists for the real pole and complex pair of zeros, probably due to the high noise level. Still, a tremendous improvement is observed w.r.t. the time domain calibration.

7. CONCLUSIONS

A new LS-SVM based estimator with auto-tuning has been introduced to identify linear, time-varying systems. Considering the fact that one single data set is present, and that the model can not be reliably used for extrapolation outside the measured time interval, the calibration had to be done via a validation data set which is interleaved with the estimation data set. For that reason, a frequency domain formulation has been chosen for its robustness w.r.t. correlated errors between these two data sets.

A simulation example has demonstrated that the calibration, when performed in the time domain, may be mislead by correlated errors, leading to useless models. The calibration in the frequency domain did not have that problem, making it more suitable to estimate dynamic systems.

REFERENCES

K. De Brabanter, J. De Brabanter, J. A. K. Suykens, and B. De Moor. Kernel Regression in the Presence of Correlated Errors. *Journal of Machine Learning Research*, 12:1955–1976, June 2011. ISSN 1532-4435.

A. Fujimori and L. Ljung. Parameter estimation of polytopic models for a linear parameter varying aircraft system. *Transactions of the Japan Society for Aeronautical and Space Sciences*, 49(165):129–136, 2006.

Jan S. Groenewegen, Jurriaan H. de Groot, Alfred C. Schouten, Andrea B. Maier, J. Hans Arendzen, and Carel G. M. Meskers. Spinal reflex properties in the long term after stroke. *Journal of electromyography and kinesiology: official journal of the International Society of Electrophysiological Kinesiology*, 22(2):234–242, 04 2012.

J. Lataire. *Frequency Domain Measurement and Identification of Linear, Time-Varying Systems.* PhD thesis, Vrije Universiteit Brussel, University Press, Zelzate, March 2011.

J. Lataire and R. Pintelon. Frequency domain errors-in-variables identification of a time-varying, discrete time system. In *Proc. of the 18th IFAC World Congress*, pages 13110–13114, Milan, Italy, Aug. 28 to Sept. 2 2011.

J. Lataire, D. Piga, and R. Tóth. A frequency-domain formulation of least-squares support vector machines to deal with correlated errors. Internal note, October 2012.

V. Laurain, W-X. Zheng, and R. Tóth. Introducing instrumental variables in the LS-SVM based identification framework. In *Proc. of the 50th IEEE Conf. on Decision and Control*, pages 3198–3203, Orlando, Florida, USA, Dec. 2011.

J. Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London, Series A*, 209:415–446, 1909.

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

R. Pintelon, J. Schoukens, and P. Guillaume. Continuous-time noise modelling from sampled data. *IEEE Trans. on Instrumentation and Measurement*, 55(6):2253–2258, Dec. 2006.

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

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

E. Tourwé, T. Breugelmans, J. Lataire, T. Hauffman, R. Pintelon, and A. Hubin. Estimation of the instantaneous impedance of time-varying systems. In *Proc. of 61th Annual Meeting of the International Society of Electrochemistry*, Nice, France, Sept. - Oct. 2010.

V. Vapnik. *Statistical Learning Theory.* Wiley-Interscience, 1998.