Machine Learning Applied to the Blind Identification of Multiple Delays in Distributed Systems

Post-print, author version

Machine Learning Applied to the Blind Identification of Multiple Delays in Distributed Systems

Felipe Treviso*, Riccardo Trinchero and Flavio G. Canavero
Politecnico di Torino, Torino, Italy

Abstract

This paper focuses on the application of the Least-Square Support Vector Machine (LS-SVM) regression for the modeling of frequency responses of complex interconnect structures. The goal is to obtain a delayed-rational model (DRM) for the structure accounting for multiple time-delays generated by wave propagation and reflections along the channel. A novel approach for the time-delays estimation based on the LS-SVM regression is introduced. The delays are estimated using the dual space formulation of the LS-SVM with an ad-hoc kernel that considers a possible delay interval. The results highlight the lower order of DRMs obtained using the delays identified through this method when comparing to the vector fitting approach by applying it to a high-speed cable link.

1 Introduction

Signal integrity is one of the limiting factors on the capacity of data transmission in a high-speed link. Electrical interconnects are responsible for a considerable part of signal degradation, due to propagation effects such as attenuation, ringing, signal delay, distortion, reflections and crosstalk [1]. Therefore, accurate models for the simulation of these structures are essential to predict signal integrity within a simulation framework during the design phase of a high-speed link. The standard approach to include an interconnect with a complex geometry in an electric circuit simulation tool is to approximate tabular data of their transfer function through a sum of rational functions, which can be identified with the popular vector fitting (VF) algorithm [2]. However, when the propagation delay between ports of the system is large, a VF model may require a large number of poles to accurately mimic the actual frequency-domain behavior of the interconnect. The resulting model eventually will not model exactly the propagation delay, which may appear in simulations as spurious signals transmitted in a time shorter than the actual propagation delay of the interconnect and lead to an inaccurate evaluation of the signal integrity of the channel [3].

A model that approximates accurately the interconnect structure described above, and also allows the transfer function to be equivalently represented by a simple circuit containing only ideal transmission lines and basic lumped elements (resistors, capacitors and inductors) [4, 5] can be written as:

$$H(j\omega) \approx \tilde{H}(j\omega) = \sum_{i=1}^{n_\tau} \left(\sum_{j=1}^{n_{p,i}} \frac{r_{ij}}{j\omega - p_{ij}} + r_{i,0} \right) e^{-j\omega\tau_i}. \quad (1)$$

$\tilde{H}(j\omega)$ approximates $H(j\omega)$ using rational functions with poles $p_{ij} \in \mathbb{C}$ and residues $r_{ij} \in \mathbb{C}$ in complex-conjugate pairs, with $p_{ij} = p'_{ij} + jp''_{ij}$, and a constant term $r_{i,0} \in \mathbb{R}$, in order to represent a real valued system in time-domain. Those rational functions are multiplied by an exponential term that accounts for the time-delay $\tau_i \in \mathbb{R}$, which is larger than zero for causal systems, resulting in a *delayed-rational model* (DRM). The model in (1) is valid for a single transfer function, however, a multiport system can be obtained by modeling each element of its transfer function matrix individually.

The estimation of all parameters in (1) at the same time is impractical, as suitable values are needed for the time-delay in the exponential term and for residues and poles in the numerator and denumerator of the rational term. Current methods to achieve the desired model format deal with the estimation of these parts independently, finding first the delay term and then the rational terms. The identification of multiple time-delays is a critical step of the process. The Gabor transform, which provides a time-frequency decomposition of $H(j\omega)$, turns out to be the most used method [4]. The underlying idea is to analyze the energy content as a function of the delay, and use the time-delays that have the largest relative contribution to the total energy as dominant propagation delays. The time-frequency decomposition can also be achieved by performing a wavelet transform on the time-domain impulse response [5]. An alternative time-domain based method considers that the analysis of the time-domain response of the system to a narrowband input pulse can be used as a way to estimate its corresponding propagation delays [6]. However, all these techniques estimate the time-delays without considering that they will be multiplied by a rational function in the final model. Given that an accurate set of delays is obtained, a robust method to obtain the rational part of the model is the delayed vector fitting (DVF) [4, 6], a modified version of VF that uses delayed basis functions and through iterations identify a common set of poles for all the delay terms. This method provides excellent results when an accurate estimate for the dominant time-delays is provided. Otherwise, an optimization to improve the initial set of time-delay values is needed [4].

In this work, we present an alternative scheme for the estimation of the dominant propagation delays of a distributed system based on the least-squares support vector machine (LS-SVM) regression, which is a flexible and powerful Machine Learning (ML) regression that has been recently adopted for the uncertainty quantification in complex electronic systems [7]. The feasibility and the strength of the proposed approach is then investigated by applying it to the modeling of a long high-speed cable [8].

2 The Least-Square Support Vector Machine

The adopted LS-SVM regression searches to approximate a set of training data pairs $\{\mathbf{x}_k, \mathrm{y}_k\}$ for $k = 1 \dots K$, with a nonlinear regression $\tilde{\mathrm{y}}(\mathbf{x})$, with the input $\mathbf{x} = [x_1, \dots, x_d]^T \in \mathbb{C}^d$ and output $\tilde{\mathrm{y}}(\mathbf{x}) \in \mathbb{C}$. The LS-SVM provides two ways to express its regression model [9]:

$$\tilde{\mathrm{y}}(\mathbf{x}) = \langle \mathbf{w}, \varphi(\mathbf{x}) \rangle + b, \tag{2}$$

where weights $\mathbf{w} = [\mathrm{w}_1, \dots, \mathrm{w}_D]^T \in \mathbb{C}^D$, $b = b_r + jb_i \in \mathbb{C}$, and the nonlinear map $\varphi(\cdot) : \mathbb{C}^d \to \mathbb{C}^D$ from the input space of dimension d to a feature space of dimension D. Equation (2) will be called *primal space representation*. The above formulation in the primal space is equivalent to the following dual space formulation, which writes:

$$\tilde{\mathrm{y}}(\mathbf{x}) = \sum_{k=1}^{K} \alpha_k \mathrm{k}(\mathbf{x}_k, \mathbf{x}) + b, \tag{3}$$

where $\alpha_k = \alpha_{k,r} + j\alpha_{k,i} \in \mathbb{C}$ and $\mathrm{k}(\mathbf{x}_k, \mathbf{x})$ is the so called kernel function $(\mathbb{C}^d \times \mathbb{C}^d) \to \mathbb{C}$ defined as:

$$\mathrm{k}(\mathbf{x}_i, \mathbf{x}_j) = \left\langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \right\rangle. \tag{4}$$

The duality between the formulation in (2) and (3) allows us to choose the more adequate model representation to work with, and when needed, the dual space coefficients α_k can be related to the primal space weights $\mathbf{w}$ by means of:

$$\mathbf{w} = \sum_{k=1}^{N} \alpha_k \varphi^*(\mathbf{x}_k). \tag{5}$$

The main difference between the two models is their dimensionality: the primal space model has D terms, which is the number of basis of the primal space, i.e., the dimension of $\mathbf{w}$ and $\varphi(\mathbf{x})$, which is a fixed and unbounded value, i.e., it can be infinite, while the dual space model is a non-parametric model with a number of terms equal to the number of training samples, K (plus the constant bias term present on both models).

The coefficients α_k and the bias term b of the dual space model are estimated by solving the following linear system of equations:

$$\begin{bmatrix} \Omega + \mathbf{I}^{2K}/\gamma & \mathbb{1} & \mathbb{0} \\ \mathbb{1}^T \quad \mathbb{0}^T & 0 & 0 \\ \mathbb{0}^T \quad \mathbb{1}^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_r \\ \alpha_i \\ b_r \\ b_i \end{bmatrix} = \begin{bmatrix} \mathbf{y}_r \\ \mathbf{y}_i \\ 0 \\ 0 \end{bmatrix}, \tag{6}$$

where $\mathbf{y}_r = [\mathrm{y}_{1,r}, \dots, \mathrm{y}_{K,r}]^T$ and $\mathbf{y}_i = [\mathrm{y}_{1,i}, \dots, \mathrm{y}_{K,i}]^T$ are the real and imaginary parts of $\mathbf{y}$, γ is a regularization parameter that should be large to minimize errors, $\mathbb{1} = [1, \dots, 1]^T$ and $\mathbb{0} = [0, \dots, 0]^T$ are vectors containing K equal elements, $\mathbf{I}^{2K}$ is an identity matrix with size $(2K \times 2K)$, and the $(2K \times 2K)$ kernel matrix $\Omega(\mathbf{x}, \mathbf{x})$ is defined as:

$$\Omega = \begin{bmatrix} \Omega^{(1,1)} & \Omega^{(1,2)} \\ \Omega^{(2,1)} & \Omega^{(2,2)} \end{bmatrix} = \begin{bmatrix} \Re\{\mathrm{k}(\mathbf{x}_i, \mathbf{x}_j)\} & \Im\{\mathrm{k}(\mathbf{x}_i, \mathbf{x}_j)\} \\ -\Im\{\mathrm{k}(\mathbf{x}_i, \mathbf{x}_j)\} & \Re\{\mathrm{k}(\mathbf{x}_i, \mathbf{x}_j)\} \end{bmatrix}, \tag{7}$$

The submatrices $\Omega^{(\cdot,\cdot)}$ are $(K \times K)$ square matrices in which the elements $\Omega^{(\cdot,\cdot)}_{i,j}$ in the i-th row and j-th column are defined from the input values of $\mathbf{x}_i$ and $\mathbf{x}_j$ by means of the kernel function in (4).

3 Time-Delay Identification via LS-SVM regression

The LS-SVM regression of the previous section can be applied to fit sampled pairs of a transfer function $\{(\omega_k, H(j\omega_k))\}_{k=1}^K$ by considering a scalar input $\mathbf{x}_k = [\omega_k]$ and complex output $\mathrm{y}(\mathbf{x}_k) = H(j\omega_k)$. The dual space model of this reads:

$$\tilde{H}(j\omega) = \sum_{k=1}^{K} \alpha_k \mathrm{k}(\omega_k, \omega) + b. \tag{8}$$

It requires the definition of the kernel function in (4). The most used kernel functions are the linear, polynomial and gaussian kernels, which are based on their specific feature space $\varphi(\omega)$, but it is possible to define an ad-hoc kernel function so that the basis in its feature space correspond to delayed-rational functions. Such kernel writes:

$$\mathrm{k}(\omega, \omega_k) = \mathrm{k}_p(j\omega, j\omega_k)\mathrm{k}_\tau(j\omega, j\omega_k), \tag{9}$$

with

$$\mathrm{k}_p(\omega_k, \omega_l) = \sum_{j=1}^{n_p} \left(\frac{|p'_j|^{1/2}}{(j\omega_k - p_j)} \frac{|p'_j|^{1/2}}{\left(-j\omega_l - p_j^*\right)} \right), \tag{10}$$

and

$$\mathrm{k}_\tau(\omega_k, \omega_l) = \begin{cases} \frac{\left(e^{-j\tau_m(\omega_l - \omega_k)} - e^{-j\tau_M(\omega_l - \omega_k)}\right)}{j(\omega_l - \omega_k)}, & \omega_k - \omega_l \neq 0 \\ \tau_M - \tau_m, & \omega_k - \omega_l = 0 \end{cases}. \tag{11}$$

The above kernel corresponds to a feature space with the following delayed-rational basis:

$$\varphi_{ij}(j\omega; p_{ij}, \tau_i) = \frac{|p'_{ij}|^{1/2}}{j\omega - p_{ij}} e^{-j\omega\tau_i}. \tag{12}$$

The kernel in (9) considers a feature space of delayed-rational basis with a set of n_p complex poles and the delay τ_i assuming all possible values from a minimum τ_m to a maximum τ_M. This feature space is infinite-dimensional, but we are able to use it due to the dual space regression model of the LS-SVM, from which we obtain a set of coefficients α and b, that represents the model in the dual space,

according to (8), which is not a DRM. However, we can use (5) to compute the value of $\mathrm{w}(\tau_i, p_j)$:

$$\mathrm{w}(\tau_i, p_j) = \sum_{k=1}^{K} \alpha_k \frac{|p'_j|^{1/2}}{j\omega_k - p_j} e^{-j\omega_k \tau_i}. \tag{13}$$

This value allows the representation of the model as in (2). As the basis used to build the LS-SVM kernel correspond to delayed rational functions, the primal space representation of the LS-SVM model is a DRM. The problem with the direct utilization of the LS-SVM DRM is the number of terms that it contains: it has all the basis with all the combinations of poles and delays included in the kernel. As an infinite number of delays was included, the model will have an infinite number of basis, which is not useful when the implementation of the model is considered. Nonetheless, the model contains information about the most relevant values of τ_i. As the energy density provided by each basis is proportional to $\|\mathrm{w}(\tau_i, p_j)\|^2$, independently of the chosen pole, we can compute the distribution of the energy associated with the system as a function of the time-delay value $\mathrm{w}_\tau(\tau_i)$:

$$\mathrm{w}_\tau(\tau_i) = \sum_{j=1}^{n_p} \|\mathrm{w}(\tau_i, p_j)\|^2. \tag{14}$$

$\mathrm{w}_\tau(\tau_i)$ shows at which values of τ_i the system can have a higher energy value (the system's dominant delays), which is observed by plotting and analyzing this function. We should remark that one of the terms that the LS-SVM problem seeks to minimize is $\sum |\mathrm{w}|^2$, making unlikely the occurrence of unnecessary high values for $\mathrm{w}(\tau_i, p_j)$. Algorithm 1 summarizes the procedure for the delay identification.

It is important to remark that the poles used in this ad-hoc kernel do not need to have a specific value. It is expected that the final DRM has a low-order rational part, and it is assumed that a higher number of poles randomly placed in the left half-plane is able to approximate the original rational part. For the time-delay interval, if we assume that the frequency points ω_k are uniformly spaced, τ_M should respect the relation $\tau_M < 2\pi/(\omega_{k+1} - \omega_k) = 1/\Delta f$.

Algorithm 1 Delay identification algorithm

1: Group the samples in the vectors $\mathbf{y}_r + j\mathbf{y}_i = [\mathrm{y}_{r,1} + j\,\mathrm{y}_{i,1}, \ldots, \mathrm{y}_{r,K} + j\,\mathrm{y}_{i,K}]$ and $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_K]$;
2: Randomly select a set of complex poles $\mathscr{P} \in \mathbb{C}^{n_p}$ in a reasonable region of the s-plane;
3: Define minimum and maximum delays τ_m and τ_M;
4: Compute the kernel for all frequency point pairs $\{\omega_i, \omega_j\}, \{i, j\} = \{1, \ldots, K\}$, using (9);
5: Assembly the matrix in (7);
6: Set a value for γ and solve (6);
7: Compute $\mathrm{w}_\tau(\tau_i)$ according to (13);
8: Find the peaks in $\mathrm{w}_\tau(\tau_i)$, which correspond to the set of relevant delays $\tau = \{\tau_1 \ldots \tau_{n_\tau}\}$.

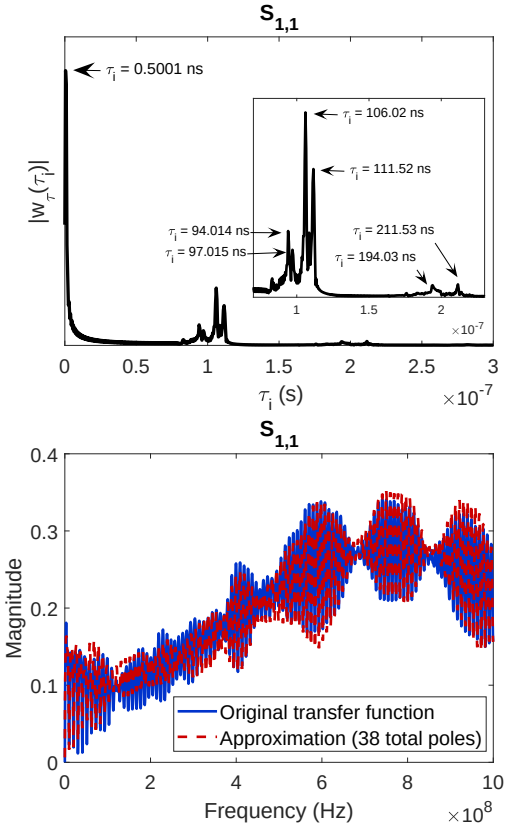

Figure 1. Top Panel: plot of $\mathrm{w}_\tau(\tau_i)$ for the transfer function of $S_{1,1}$. Bottom Panel: magnitude plot showing the accuracy of the model for $S_{1,1}$.

4 Application Example

The application of the presented method can be exemplified by analyzing simulated data of a single wire in a SpaceWire link [8], containing a 10 m long SpaceWire cable with compatible connectors and PCB adapters. The method is applied to the curves of a reflection $S_{1,1}(j\omega)$ and a transmission $S_{1,2}(j\omega)$ scattering parameters. A full multiport model is achieved by applying the method to each of the parameters in the scattering matrix individually, and then combining the individual results [4]. The method is applied first to $S_{1,1}(j\omega)$, which has its identified delays shown in Fig. 1. The first identified delay happens at $\tau = 0.5$ ns, representing a reflection that occurs a few centimeters into the propagation path. In the simulated setup, this reflection happens when the PCB and connector transitions into the SpaceWire cable. Further ahead, another 6 delays were chosen in the order of hundreds of ns. Furthermore, $\tau = 0$ also is considered as a possible delay because this is a scattering parameter representing the port reflection. The approximation for the magnitude of the $S_{1,1}$ transfer function performed with the identified delays is also presented. It shows a good accuracy by using 8 delays and 38 total poles distributed among them. A vector fitting approximation with an error of a similar order of magnitude requires around 180 poles.

The process is repeated for $S_{1,2}$. Figure 2 shows the iden-

tified delays for this component of the system. Here, we see that contrary to the previous reflection, in this transmission term it takes at least a certain amount of time until we can see the first peak. In total, 6 main delays were identified. Figure 2 shows also a delayed-rational approximation for the magnitude of $S_{1,2}$ performed with the identified delays. It shows a good accuracy by using 6 delays and 25 total poles distributed among them. A vector fitting approximation with an error of a similar order of magnitude requires the use of more than 100 poles. A summary of the L_2 and L_∞-norm error between the original data and the constructed model is shown in Table 1, with results of a VF model for comparison, remarking the reduced order achieved with this method while reaching similar or larger accuracy.

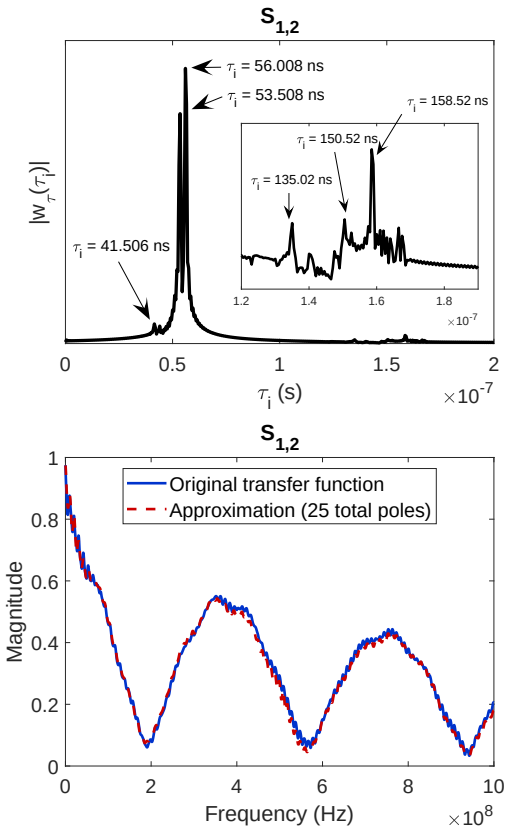

Figure 2. Top Panel: plot of $w_\tau(\tau_i)$ for the transfer function of $S_{1,2}$. Bottom Panel: magnitude plot showing the accuracy of the model for $S_{1,2}$.

Table 1. Summary of the error between available data for the link in the example and models used to approximate it.

	VF	Proposed method
$S_{1,1}$ error - L_2	0.776	0.780
$S_{1,1}$ error - L_∞	0.086	0.085
$S_{1,1}$ order	180	38
$S_{1,2}$ error - L_2	1.228	0.806
$S_{1,2}$ error - L_∞	0.177	0.041
$S_{1,2}$ order	119	25

5 Conclusion

This paper presents the application of the LS-SVM regression to the modeling of frequency responses generated by systems possessing multiple propagation delays based on the use of an ad-hoc kernel. This LS-SVM regression allows the estimation of the dominant propagation delays, which can be combined to rational fitting techniques to obtain models with a smaller order compared to VF models, leading to more compact behavioral models. By applying the method for transfer function samples of a 10 m SpaceWire cable link, the model extracted using the delays identified with the presented method used a number of poles around 4.7× smaller than VF models, while also having guaranteed causality.

References

[1] R. Achar and M. S. Nakhla, *Simulation of high-speed interconnects*, Proceedings of the IEEE, vol. 89, no. 5, pp. 693-728, May 2001.

[2] B. Gustavsen and A. Semlyen, *Rational approximation of frequency domain responses by vector fitting*, in IEEE Transactions on Power Delivery, vol. 14, no. 3, pp. 1052-1061, July 1999.

[3] R. Mandrekar, M. Swaminathan, *Causality Enforcement in Transient Simulation of Passive Networks through Delay Extraction*, 9th IEEE Workshop on Signal Propagation on Interconnects, 2005, pp. 25-28.

[4] A. Chinea, P. Triverio, S. Grivet-Talocia, *Delay-Based Macromodelling of Long Interconnects From Frequency-Domain Terminal Responses*, IEEE Trans. on Advanced Packaging, Vol. 33, No. 1, Feb. 2010.

[5] A. Charest *et al.*, *Time Domain Delay Extraction-Based Macromodeling Algorithm for Long-Delay Networks*, IEEE Trans. on Advanced Packaging, vol. 33, no. 1, pp. 219-235, Feb. 2010.

[6] A. Chinea *et al.*, *Signal Integrity Verification of Multichip Links Using Passive Channel Macromodels*, IEEE Trans. on Components, Packaging and Manufact. Tech., vol. 1, no. 6, pp. 920-933, Jun. 2011.

[7] R. Trinchero *et al.*, *Machine Learning and Uncertainty Quantification for Surrogate Models of Integrated Devices With a Large Number of Parameters*, in IEEE Access, vol. 7, pp. 4056-4066, 2019.

[8] F. Treviso, R. Trinchero and F. G. Canavero, "Validation of a Physical-Based Model for a Spacewire Cable," 2019 ESA Workshop on Aerospace EMC, Budapest, Hungary, 2019, pp. 1-6.

[9] J.A.K. Suykens *et al.*, *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002 (ISBN 981-238-151-1).