Nonlinear system identification by means of SVMs: choice of excitation signals

Anna Marconato[1], Andrea Boni[1], Dario Petri[1], Johan Schoukens[2]

[1] *DISI – University of Trento, via Sommarive 14, 38100 Trento, Italy*
E-mail: anna.marconato@disi.unitn.it
[2] *ELEC – Vrije Universiteit Brussel, Pleinlaan 2,1050 Brussels, Belgium*

Abstract-In this work we discuss the application of Support Vector Machines to the problem of identifying a specific class of nonlinear systems, namely Wiener-Hammerstein systems. Only based on a set of Input/Output measurements, a model is built that approximates well the behavior of the considered system. However, care should be taken when designing suitable excitations, as the performance of the proposed approach turns out to be quite sensitive to the nature of the input signal. This sensitivity is studied here by using several datasets, characterised by different excitation signals, in terms of root mean square value, frequency band, spectrum shape and amplitude distribution.

I. Introduction

Nonlinear identification represents an essential tool in engineering, since it permits to describe the behavior of most real-life systems, both for control applications and for extracting a model that can be used to simulate the system under test. Here we deal with the identification of a specific class of systems, namely Wiener-Hammerstein systems. Wiener-Hammerstein systems are nonlinear systems characterized by a well defined structure: a cascade of two linear dynamic blocks G_1 and G_2 with a static nonlinearity sandwiched in between, as depicted in Fig.1.

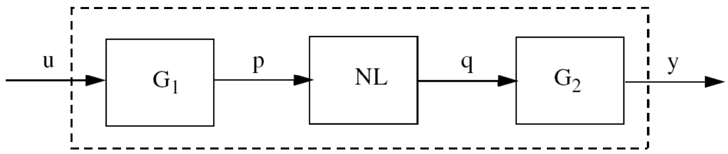

Fig.1 General block structure of a Wiener-Hammerstein system

The problem of identifying Wiener-Hammerstein systems has already been addressed in the literature, for example in [1] and [2].
Here instead we deal with the application of a well-known example of learning-from-examples algorithms, namely Support Vector Machines for Regression (SVRs), which provide a sound theoretical background, and have already shown very good performance in many applications [3], [4].
An important aspect that needs to be thoroughly discussed concerns the choice of the excitations in input to the system: we will show how the performance of an SVR model can be dependent on the specific kind of excitation signal. In particular, we will consider several datasets, built employing input signals of different nature, in order to study the sensitivity of SVR models to relevant properties of the chosen excitations.
This work is structured as follows: next section briefly presents the theory of SVRs applied to the problem of identifying nonlinear systems, while Section III describes more in details the methodology, with particular emphasis on the characteristics of the chosen excitation signals. Simulation results for the different cases are discussed in Section IV; some final comments conclude the paper in Section V.

II. Support Vector Machines for system identification

SVRs, as other learning-from-examples algorithms, are based on a very simple principle: once you have collected a set of Input/Output measurements $Z = \{\mathbf{u}_i, y_i\}_{i=1}^{m}$ a training phase is performed aimed at extracting all useful information from the available data. Thus, an estimating function $f(\mathbf{u}_i)$ is built that approximates the input-output relationship. Following the standard ε -SVR approach, errors are penalised, in a linear way, only outside a so-called insensitive zone, whose width is indicated by ε, while deviations smaller than ε are considered to be negligible [3]. The cost function is then defined by summing up all error contributions, and by taking into account a term describing the smoothness of the function.

This formulation of the SVR algorithm results in the definition of a constrained quadratic optimization problem where the function to be minimized turns out to be convex, therefore avoiding local minima problems (something that instead hampers other techniques, e.g. Artificial Neural Networks.)

The obtained SVR estimating function is the following:

$$f(\mathbf{u}, \boldsymbol{\beta}) = \sum_{i \in SV} \beta_i k(\mathbf{u}_i, \mathbf{u}) + b$$

Notice that not all the samples used during the training give a positive contribution, instead only a subset of the original dataset (the Support Vectors $\mathbf{u}_i$) is used for building the model [3]. Furthermore, suitable kernel functions are introduced in order to deal with the nonlinear case. Here we will consider the Gaussian kernel, expressed as $k(\mathbf{u}_i, \mathbf{u}) = e^{-\gamma \|\mathbf{u}_i - \mathbf{u}\|^2}$ [4].

The performance of the SVR function depends on a number of parameters (called hyperparameters) that need to be tuned in order to get the best possible model. Examples of these hyperparameters are the width ε of the insensitive zone, and parameter γ characterizing the Gaussian kernel function. Here, as optimality criterion, we will consider an accuracy index, namely the Root Mean Square Error (RMSE) of the model with respect to the true output value. The optimal model is then chosen as the one giving the smallest RMSE in the training phase, and can then be validated on a set of previously unseen samples.

To explore the hyperparameters space during the model selection phase, we decided to exploit a technique based on Genetic Algorithms, which provide a faster and more efficient search tool when compared with the traditional grid search approach [5].

III. Proposed methodology and considered excitation signals

As we have seen in the previous section, in general the data are formed by a vector of features **u** that characterise the samples, and a label y that represents the output value. In this work we consider, as features, the current value of the input signal of the Wiener-Hammerstein system, together with a number of past input values. This of course does not need to be the only choice: recursive models can be built by introducing also past output values in the feature vector. However, here we will deal only with the simple (non-recursive) case.

A. System

The proposed approach has been applied to a simulated example of nonlinear system with a Wiener-Hammerstein structure, as illustrated in Fig.1. In particular, the two linear dynamic blocks G_1 and G_2 were designed as two third order Chebyshev low-pass filters, with cut-off frequency equal to 0.11, and 1 dB of peak-to-peak ripple in the passband. The static nonlinear block NL was chosen as follows:

$$q = p + \alpha \cdot p^3$$

where the value of coefficient $\alpha = 0.05$ was tuned in order to have 90% of the power on the linear term and 10% on the nonlinear term. Notice here that in order to identify the system, only information about input values u and observed output values y is used, with no need of having direct access to signals p and q.

B. Excitation signals

As excitations, we decided to use, for the training phase, both a filtered Gaussian noise and a Random Phase Multisine signal (both made of N=4096 samples), the latter characterized by the following equation:

$$u(t) = \sum_{k=1}^{F} A_k \cos(2\pi f_0 k t + \varphi_k)$$

where $f_0 = 1/N$, $t = 1,...,N$ and φ_k is uniformly random distributed in $[0,2\pi)$ [6].

This choice was motivated by the idea of employing two signals, both belonging to the class of Gaussian excitations, but with different properties in the frequency domain.

For the Random Phase Multisine, we decided to consider both $F=F_1=500$ (excited frequencies up to

$F_1 \cdot f_0 = 500 \cdot 1/4096 = 0.12$) and *F=F₂=2048* (full multisine, excited frequencies up to $F_2 \cdot f_0 = 2048 \cdot 1/4096 = 0.5$). Amplitudes A_k were normalized in order to have a root mean square value of the signals equal to 1. To get rid of transient effects in the system response, two periods of the multisines were generated, but only the second ones were taken to build the datasets.
The second kind of excitation was generated by filtering a white Gaussian noise sequence by means of a Butterworth filter with cut-off frequency equal to $0.7 \cdot f_{\max}$, where $f_{\max} = F_1 \cdot f_0$ indicates the maximum frequency excited by the first multisine considered above. The signal was then normalized in order to have the same power of the multisines.
In all cases the feature vector **u** was built as $u(t), u(t-1), \ldots, u(t-d+1)$ with $d = 64$, given that, after 64 samples, the impulse response of the system as a whole is 1/1000 of the maximum value. The three considered SVR models show a number a Support Vectors in the order of *N/2*.
For the final validation of the three SVR models (obtained with the three excitations described above), we employed several examples of excitation signals: (i) Random Phase Multisines characterised by different frequency band (excited lines up to 0.12, 0.3 or 0.5), different root mean square value (equal to 0.5, 0.8, 1, 1.2, 2); (ii) Gaussian noise filtered with Butterworth filters of order 2 or 10; (iii) uniform noise characterised by different root mean square value (equal to 1 or 1.73); (iv) uniform noise filtered with Butterworth filters of order 2 or 10 (root mean square value equal to 1).

IV. Simulation results and comments

As already mentioned, we performed several simulations using different SVR models, and different validation sets. What we can observe from the results is a general trend that helps in pointing out examples of sensitivity to specific properties of the excitation signals. In the following we will analyse the influence on the performance of changes in root mean square values, frequency band, spectrum shape or amplitude distribution of the considered excitations. We will also try to provide hints of a possible explanation for the observed phenomena.

A. Root mean square value

Let us consider a model obtained by training an SVR with a multisine of root mean square value equal to 1 (and frequency range up to 0.12) as excitation signal. We validate this model on several datasets, built using different root mean square values for the multisine excitations. Results in terms of the RMSE and of the relative RMSE (RMSE/RMSY) are given in Table 1.

TRAINING			VALIDATION			
Type	**RMS**	**RMSE**	**Type**	**RMS**	**RMSE**	**RMSE/RMSY**
Multisine	1	0.0218	Multisine	0.5	0.0043	0.0100
			Multisine	0.8	0.0073	0.0101
			Multisine	1	0.0154	0.0164
			Multisine	1.2	0.0357	0.0301
			Multisine	2	0.4265	0.1638

Table 1. RMSE and relative RMSE results for multisine excitations characterised by different root mean square values (normalised band-width equal to 0.12).

Notice that, for a fair comparison of the accuracy performance, also results in terms of the relative RMSE are provided, since we are dealing with differences in the root mean square value of the inputs (and therefore also of the outputs). However, the difference in the performance (very good for small root mean square values and very bad for big root mean square values) is clearly due to something more than just a scale factor related to different root mean square values. A similar trend is observed when looking at the results in all those cases where uniform noise excitations (with different root mean square values) are used for the validation, as shown in Table 2 (further comments on the influence of changes in the amplitude distribution on the results are given in the last part of this section).
A reasonable explanation for this behavior can be found by having a closer look at the three SVR models obtained in the training phase. They are all characterised by a very small value (around 0.001) of coefficient γ of the Gaussian kernel, resulting in an estimating function which is almost linear. Since the considered nonlinear block shows an approximatively linear behavior for input values less than 1 in

absolute value, this kind of model will approximate well the behavior of excitations in this range, while it will difficultly follow excitations with root mean square values larger than 1.

TRAINING				VALIDATION			
Type	**RMS**	**BW**	**RMSE**	**Type**	**RMS**	**RMSE**	**RMSE/RMSY**
Multisine	1	0.12	0.0218	Uniform	1	0.0425	0.0992
				Uniform	1.73	0.1310	0.1678
Multisine	1	0.5	0.0121	Uniform	1	0.0105	0.0245
				Uniform	1.73	0.1350	0.1729
Gaussian	1	0.09	0.0240	Uniform	1	0.0307	0.0716
				Uniform	1.73	0.1280	0.1640

Table 2. RMSE and relative RMSE results for uniform noise excitations characterised by different root mean square values, for the three SVR models (normalised band-width for uniform noise equal to 0.5). For filtered Gaussian noise, the normalised band-width is the cut-off frequency of the Butterworth filter.

B. Frequency band

To check the influence of frequency band properties on the performance of the SVR models, we have taken into account several examples of Random Phase Multisines, with excited lines covering increasing frequency ranges, and an example of filtered Gaussian noise (with frequency band approximatively as for the first multisine). We used these excitation signals during training and validation phases, keeping the root mean square values of the inputs equal to 1. Some of the results are shown in Table 3.

TRAINING			VALIDATION		
Type	**BW**	**RMSE**	**Type**	**BW**	**RMSE**
Multisine	0.12	0.0218	Multisine	0.12	0.0154
			Multisine	0.3	0.0472
			Multisine	0.5	0.0424
			Gaussian	0.09	0.0252
Multisine	0.5	0.0121	Multisine	0.12	0.1143
			Multisine	0.3	0.0248
			Multisine	0.5	0.0126
			Gaussian	0.09	0.1217

Table 3. RMSE results for excitations characterised by different frequency band, for two different SVR models (all considered signals have root mean square value equal to 1). For filtered Gaussian noise, the normalised band-width is the cut-off frequency of the Butterworth filter.

Here the first aspect to be noticed is that best results are obtained when using the same kind of excitation for both training and validation. Furthermore, the first example of multisine (with band-width equal to 0.12) and Gaussian noise give more or less the same RMSE values. Finally, we observe that, in the considered example, when training an SVR model with an excitation characterised by frequencies in a limited range (multisine with band-width equal to 0.12), validation results for signals with a wider band (multisines with band-width equal to 0.3 and 0.5) are worse, but still reasonable. However, a model trained using a signal with large frequency band gives very poor results when validated on excitations with frequencies in a limited range. This phenomenon can be interpreted as follows: the non-negligible contributions at higher frequencies, which inevitably are present in a model trained employing excitations with large band, have a “disturbing” effect when validating data in a limited frequency range; in the opposite situation, instead, a model trained with a signal in a limited frequency range will not be able to reproduce contributions at higher frequencies, but will neither introduce undesired extra information.
However, we should underline the fact that these considerations are dependent on the band-width of the chosen filter G_1, and that a deeper analysis of this behavior needs to be performed before we can state that these conclusions have general validity.
As a final remark, we would like to underline the fact that in all cases, a uniform noise excitation with the same root mean square value, and excited lines in the whole range up to 0.5, gives the same RMSE results as a multisine with band-width equal to 0.5 (see Table 4 for details).

TRAINING			VALIDATION	
Type	**BW**	**RMSE**	**Type**	**RMSE**
Multisine	0.12	0.0218	Multisine	0.0424
			Uniform	0.0425
Multisine	0.5	0.0240	Multisine	0.0307
			Uniform	0.0307
Gaussian	0.09	0.0121	Multisine	0.0126
			Uniform	0.0105

Table 4. RMSE results for excitations characterised by same band-width (equal to 0.5) but different distribution, for the three SVR models (all considered signals have root mean square value equal to 1). For filtered Gaussian noise, the normalised band-width is the cut-off frequency of the Butterworth filter.

C. Spectrum shape

As introduced in Section II we decided to use, as validation sets, data obtained by employing Gaussian noise sequences and uniform noise sequences as excitation signals, filtered with Butterworth filters of two different orders: a second order filter, which of course gives contributions at higher frequencies, and a tenth order filter characterised by a much steeper spectrum. Table 5 summarises the obtained results. As usual, excitation signals with the same root mean square values are compared.
Here we see that differences in the shape of the spectrum do not influence much the performance (results corresponding to filter order 2 and 10 are very close). Moreover, excitation signals characterised by different amplitude distributions, but with identical spectrum shape, lead to very similar error values.

TRAINING				VALIDATION		
Type	**BW**	**filter order**	**RMSE**	**Type**	**filter order**	**RMSE**
Multisine	0.12	-	0.0218	Gaussian	2	0.0252
				Gaussian	10	0.0203
				Uniform	2	0.0231
				Uniform	10	0.0181
Gaussian	0.09	2	0.0240	Gaussian	2	0.0262
				Gaussian	10	0.0269
				Uniform	2	0.0235
				Uniform	10	0.0251

Table 5. RMSE results for excitations characterised by different spectrum shape (i.e. filters of different order), for two different SVR models (all considered signals have root mean square value equal to 1). For filtered Gaussian noise, the normalised band-width is the cut-off frequency of the Butterworth filter.

D. Amplitude distribution

As a last aspect to be taken into consideration, we tried excitation signals characterised by different amplitude distributions. In particular, we generated validation sets taking uniform noise data as input signal, in addition to multisines and Gaussian noise sequences (belonging, as we know, to the same kind of signals, namely the class of Gaussian excitations). Hence, as mentioned in the previous parts, we can compare the results obtained with uniform and multisine/Gaussian excitations, already shown in Tables 4 and 5. What really seems to matter, apart from aspects related to changes in the root mean square values (discussed in the first part of this section), is the frequency band of the chosen excitation signal. Differences in the amplitude distribution of the input do not influence significantly the accuracy performance. In fact, due to the Central Limit Theorem, since G_1 has a long impulse response, its output will become Gaussian, even for non-Gaussian (uniform in our case) inputs.

In our simulations, as far as Random Phase Multisines excitations were concerned, phase variations (due to different realizations of the multisines) did not affect much the results.
Finally, despite the differences discussed here, in all the considered examples our approach based on SVRs outperformed the estimate given by the Best Linear Approximation (BLA) [6]. Some details

about the compared validation performance are provided in Table 6, while Table 7 summarises the main results discussed in this section.

TRAINING				VALIDATION				
Type	**BW**	**filter order**	**RMSE**	**Type**	**BW**	**filter order**	**RMSE (SVR)**	**RMSE (BLA)**
Multisine	0.12	-	0.0218	Multisine	0.12	-	0.0154	0.0831
				Gaussian	0.9	2	0.0252	0.0762
				Gaussian	0.9	10	0.0203	0.0839
Gaussian	0.09	2	0.0240	Multisine	0.12	-	0.0326	0.0831
				Gaussian	0.9	2	0.0262	0.0762
				Gaussian	0.9	10	0.0269	0.0839

Table 6. RMSE results for different excitations obtained by the SVR-based approach and by the BLA, for two different SVR models.

Root mean square value	very low sensitivity if RMS < $RMS_{training}$
	high sensitivity if RMS > $RMS_{training}$
Frequency band*	low sensitivity if B > $B_{training}$
	high sensitivity if B < $B_{training}$
Spectrum shape	low sensitivity
Amplitude distribution	very low sensitivity

Table 7. Sensitivity of SVR accuracy performance to some properties of the excitation signals.
*Considerations about the frequency band of the excitations depend on the band-width of the system.

V. Conclusions

In this work we have discussed the application of SVRs for the identification of an interesting example of nonlinear systems, namely Wiener-Hammerstein systems. We have focused in particular on the critical task of choosing suitable input signals, showing the impact of changes in specific properties of the excitations on the performance of SVR models. The obtained results confirmed our intuition that care needs to be taken in order to design an appropriate excitation signal. More in details, validation results in the example of the considered Wiener-Hammerstein system seem to be quite sensitive to changes in root mean square values (for root mean square values bigger than in training) and frequency band (especially for signals with a band narrower than in training). Instead, variations of spectrum shape, amplitude distribution and phase of the signal do not affect the performance significantly. The lesson learned can be summarised as follows: when designing an excitation to be used for training a nonlinear SVR model, it is better to keep the root mean square value not too small, and a frequency band quite narrow. Of course, if any information about the validation data is available *a priori*, it should be used in order to design an appropriate excitation. In any case, deeper theoretical analysis and extensive simulations are needed for a satisfying validation of the above conclusions, so deriving design rules useful for SVM based system identification.

References

[1] S.A. Billings, S.Y. Fakhouri, “Identification of systems containing linear dynamic and static nonlinear elements”, *Automatica*, vol. 18, no. 1, pp. 15-26, 1982.

[2] M. Boutayeb, M. Darouach, “Recursive identification method for MISO Wiener-Hammerstein model”, *IEEE Transactions on Automatic Control*, vol. 40, pp. 287-291, 1995.

[3] V. Vapnik, *Statistical Learning Theory*, Wiley, 1998.

[4] B. Schölkopf and A. Smola, *Learning with kernels*, The MIT Press, 2002.

[5] A. Marconato, A. Boni, B. Caprile, D. Petri, “Model selection for power efficient analysis of measurement data”, *Instrumentation and Measurement Technology Conference*, Sorrento, Italy, 2006.

[6] R. Pintelon, J. Schoukens, *System Identification. A Frequency Domain Approach*, IEEE Press, New Jersey, 2001.