L2: successes and pitfalls of frequency-domain identification methods

Pablo Santana González

***Abstract*— For a given dataset three techniques of system identification were put to test: Correlation Analysis, empyrical transfer function estimate (ETFE) and smoothing of the ETFE. It is found that it is easier to get an accurate low-order model with the correlation analysis than with the pure frequency-domain techniques for this case.**

I. Results

The correlation analysis is performed by obtaining the correlations of the input and output vectors, obtaining a vector h which holds the coefficients of the difference equation that models the system. We can choose on a system order - a length for h - to adjust better to our data. Figure 1 shows a plot with the mean square errors (MSEs) obtained for different system orders. For demonstration, a time-series with the real output and the predicted one is also included.

The empyrical transfer function estimate (EFTE) can be calculated in a straightforward manner as the division of the output and input in frequency domain. The ETFE function may get noisy in some region, so a smoothing algorithm can be performed, typically some kind of window averaging. A Bartlett window will be applied to the obtained ETFE in hopes of denoising it. The Bartlett window width is chosen to minimize the MSE when predicting the output. Both these methods carry the disadvantage that, at first, the model order is the size of the training data, which is usually big when facing this kind of problem.

The Bode plots of the three obtained transfer functions are shown in Figure 2. All of them are seen to be very noisy in higher frequencies, while they are stabler at lower

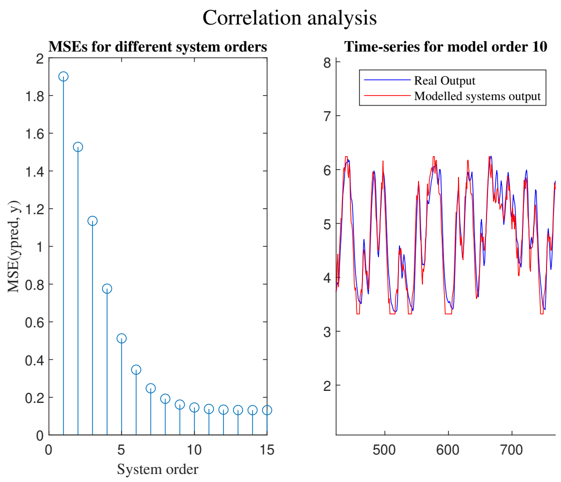

Fig. 1: Correlation analysis MSE for different system orders (left). Predicted vs measured output (right).

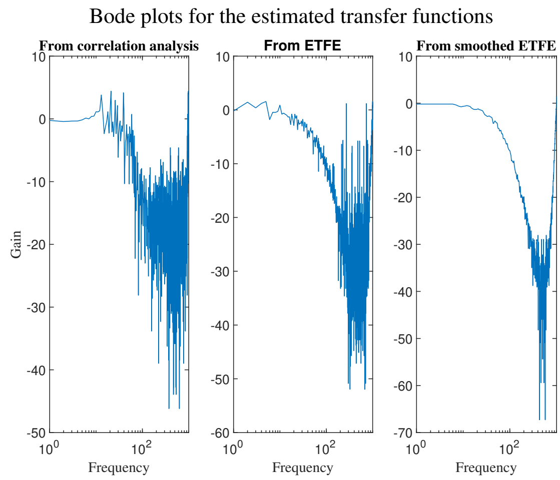

Fig. 2: Bode plots for the identified models.

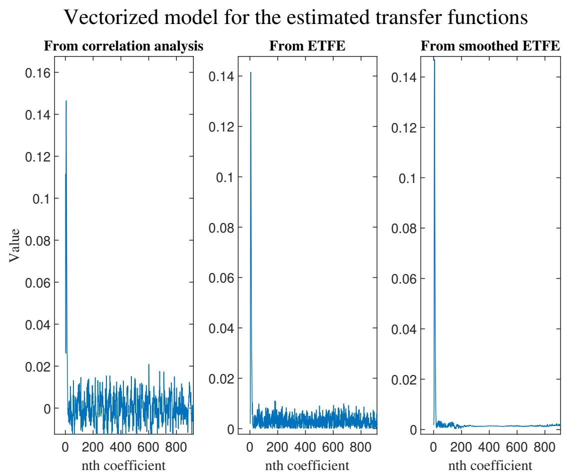

Fig. 3: Identified models in vector form through each method.

ones. In the Bode plots the difference between the raw and smoothed ETFE can be seen very clearly, being the latter quite noiseless compared to the others.

If the frequency-domain transfer functions are converted to time-domain the resulting vector will hold, again, the coefficients of the difference equation that models the system, which are plotted in Figure 3. This vector convolves the input to return the output. With the algorithms at full-power, the model order will have the size of the data. High orders are poison, because not only calculations take longer, but overfitting issues are stronger and the model is degraded.

From the plot, it can be seen that the coefficients can

Model order	Correlation	Raw ETFE	Smoothed ETFE
10th	0.1460	1.7291	1.3355
50th	0.1331	0.1440	0.1395
999th	0.1132	65.246	14.007

TABLE I: Mean squared errors for each method and various model orders.

get really low after a few vector indices, after which the coefficients are negligible white noise: these models can be greatly simplified by just assuming that the lowest values are zero. Cropping the vector is, for this case, a very effective way of reducing the model order. Table 1 holds the MSEs of the different models for a full and reduced system orders.

II. Conclusions

The correlation analysis is a clean way to obtain an accurate low-order model for the system. ETFE and its smoothed variant they require a high order to achieve a MSE comparable to that of the correlation analysis. Also, they must be calculated for the whole data then cropped if possible to try and reduce model order. Without cropping, the smoothed ETFE shows smaller MSEs than the original one.

Acknowledgments: Thanks to Martin and Robert.