IFAC Symposium on System Identification, SYSID 2006
March 29-31 2006, Newcastle, Australia

CORRELATION TEST OF RESIDUAL ERRORS IN FREQUENCY DOMAIN SYSTEM IDENTIFICATION

István Kollár[*], Rik Pintelon[**], Johan Schoukens[**]

[*]*Budapest University of Technology and Economics*
Department of Measurement and Information Systems
H-1521 Budapest, Magyar tudósok krt. 2. Hungary
Fax: +36 1 463-4112, email:kollar@mit.bme.hu
[**]*Vrije Universiteit Brussel, Dienst ELEC, Belgium*
Pleinlaan 2, B-1050 Brussel, Belgium, fax: +32 2 629-2850
Email: [Johan.Schoukens,Rik.Pintelon]@vub.ac.be

Abstract: Residual errors (deviations between measurements and system models) can be caused by several reasons: observation/process noise, nonlinear products, system transients, unmodelled dynamics, etc. The first two cannot be explained by linear models, the latter two can. Identification procedures can be stopped when the latter are not present in a reasonable size model. Therefore, we need to distinguish among these error sources.

This paper analyses components of the residual errors, and suggests a simple automatic method to effectively suppress system transients and unmodelled dynamics during variance analysis. This allows robust noise analysis, and then reliable model validation. The method is based on different properties of the error sequences in the frequency domain, without the need of separate observation noise analysis or of "full-blown" nonlinear analysis.

Keywords: system identification, residual errors, nonlinear errors, correlation test, uncorrelatedness, whiteness, parameter estimation, frequency domain.

1 INTRODUCTION

Model validation in system identification usually involves testing of the residuals: the difference between measurements and the system model. When the model properly describes system behavior, the residuals contain only observation/process noise. In this case, the properties of the residuals coincide with the theoretically expected properties. If tests verify this assumption, we consider the model validated, at least according to this test.

In practice, residual errors comprise not only observation/process noise, but also unmodelled dynamics, unexplained system transients, and/or nonlinear behavior of the system. Therefore, care must be taken when analyzing measured data for noise: unmodelled dynamics and/or unexplained system transients will yield erroneously large error levels,

[1] with support of the Flemish-Hungarian bilateral agreement B-1/04, and OTKA TS49743

leading to validation of models with large and otherwise unacceptable errors. On the other hand, nonlinear products cannot be modelled by a linear model, so they need to be incorporated into the error level to be tested, especially when the experiment is designed to have the nonlinearities act like "noise".

The noise contributions, and the nonlinearity products with excitation having some randomness (Gaussian noise or random phase excitation), form uncorrelated sequences in the frequency domain, while non-modeled dynamics and system transients yield strongly correlated sequences. Based on this, a highpass filtering operation can be used to remove the latter. This idea will be explored and executed in the paper.

2 DEFINITIONS

The usual framework applied in frequency domain system identification is as follows (Pintelon and Schoukens, 2001).

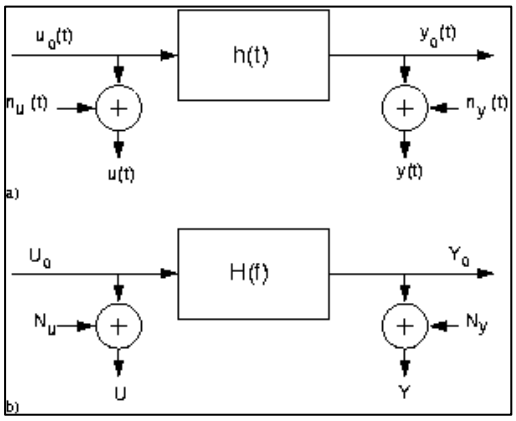

Fig. 1 The framework for frequency domain system identification (a) in time domain (b) in frequency domain (Fourier coefficients)

The unknown linear system is excited by $u_0(t)$, and the dynamic response is $y_0(t)$. Both input and output signals are measured, with observation noises $n_u(t)$ and $n_y(t)$, respectively.

For the sake of simple analysis, we assume here that the excitation is periodic, we measure in steady-state, and sampling is coherent (DFT of the time sequence contains harmonic components at just the corresponding lines). Note that these assumptions are not necessary, but make our considerations and discussion simpler.

At a given frequency f_k, we can write that

$$H(f_k) = \frac{B(f_k)}{A(f_k)} = \frac{Y_{0k}}{U_{0k}}. \qquad (1)$$

In theory, the residual at f_k can be determined by subtracting the transfer function from the ratio of the measured Fourier amplitudes:

$$e_{FRFth}(f_k) = \frac{Y_k}{U_k} - \frac{B(f_k)}{A(f_k)}, \qquad (2)$$

or, in a form linear in the system coefficients:

$$e_{th}(f_k) = A(f_k)Y_k - B(f_k)U_k. \qquad (3)$$

In practice, the transfer function coefficients are not known, we only estimate them. Therefore, the practically calculable residual series is:

$$e(f_k) = \hat{A}(f_k)Y_k - \hat{B}(f_k)U_k, \qquad (4)$$

which is, for good estimates of the coefficients of B and A, close to (3).

When the input and output observation noises are normally distributed, are independent, and additive, these residual values are complex-valued, normally distributed, have circular symmetry, with expected value zero, and variance

$$\begin{aligned}\mathrm{var}\{e_{th}(f_k)\} = |A(f_k)|^2 s^2_{Yk} + |B(f_k)|^2 s^2_{Uk}, \\ -2\,\mathrm{Re}\left(\mathrm{cov}\{U_k, Y_k\}\overline{B(f_k)}A(f_k)\right)\end{aligned} \qquad (5)$$

where $\mathrm{cov}\{U_k, Y_k\} = E\{\overline{(U_k - U_{0k})}(Y_k - Y_{0k})\}$.

The variance of the practically calculable sequence (4) is somewhat different, since in it, estimated quantities are involved. Similarly to the correction necessary for empirical variances (see Bendat and Piersol, 1986, Eq. (11.7)), the variances are slightly modified:

$$\mathrm{var}\{e(f_k)\} = s^2_e \approx \frac{F - n_p/2}{F}\mathrm{var}\{e_{th}(f_k)\} \qquad (6)$$

At different frequencies, these random variables $e(f_k)$ are uncorrelated. In other words, the sequence along the frequency axis is uncorrelated. The variances of the individual samples change with frequency, thus for simple testing, we usually normalize them, and investigate the sequence

$$\frac{e(f_k)}{std\{e(f_k)\}} = \frac{\hat{A}(f_k)Y_k - \hat{B}(f_k)U_k}{\sqrt{\mathrm{var}\{e(f_k)\}}}. \qquad (7)$$

These are, in theory, iid complex, circularly symmetric Gaussian variables, with variance 1. It is easy to construct tests for this, for example by checking the percentage of the absolute values of (7) being under 1 or another preselected value.

This all corresponds to theory, and works well in simulations. However, when working with measured data, the validation step often fails, even if the fit of the transfer function to the frequency response is seemingly very good. Excellent signal-to-noise

ratio can easily result in poor validation results – the better SNR, the poorer the validation. This seems to be puzzling, so it deserves consideration and requires improvement of the test.

When looking into the cause, one can quickly find that the reason is that the terms used in the validity test, shown in (7), have too large absolute values. The deviations between measurement and model are too large with respect to the noise we have in the model.

3 ANALYSIS OF THE RESIDUALS

In order to analyze the cause, we need to examine the factors which yield the residuals.

In general, the residuals of a fitted model ((2), (3), or (4)) are caused by the following factors:

- observation noise,
- nonlinearity products,
- unmodelled dynamics (e.g. undermodeling),
- transients during measurement.

The first two are deviations which cannot be modeled by a linear system. Their frequency domain sample sequences are uncorrelated (Pintelon and Schoukens, 2001). We wish that the test be not sensitive to them.

The second two components are related to poor modeling, thus we want to detect them in validation, and eliminate them by better models. Since they are usually described in frequency domain by low-order rational forms, their samples are highly correlated.

If we want to test the magnitudes of the residual samples, we need to select levels which correspond to the first two factors, but whose determination is insensitive to the last two factors.

A highly correlated sequence has lowpass nature, while the uncorrelated one is "white" (uncorrelated), that is, has constant spectrum. Therefore, discrimination is possible through high-pass filtering of the sequence. We need to apply a highpass filter the frequency domain sequence, then use the result to estimate the RMS value of the first two components. This will be more precisely described in the following section.

4 SELECTING A FILTER

In the frequency domain, we do filtering on the residual sequence. The specification is still quite loose – without knowing much about the system, we cannot design an optimal filter. There is however one more aspect which helps in the specification. Since we do not want to lose too much of the length of the sequence we have, we might say that we need a possibly short impulse response. From time domain signal processing, we know that from 3-coefficient harmonic windows, the one with best suppression at zero frequency is the sine wave which smoothly touches the frequency axis, the Hanning window. Therefore we will apply the following FIR filter (a kind of "inverse" Hanning window) to the frequency domain sequence:

$$h = [h(-1), h(0), h(1)] = [-0.5, 1, -0.5] \qquad (8)$$

We may notice that this is nothing else than a double differentiator in the discrete domain.

The "transfer function" is illustrated in Fig. 1. Note that this is in the "inverse transform domain" of the frequency-domain residual sequence.

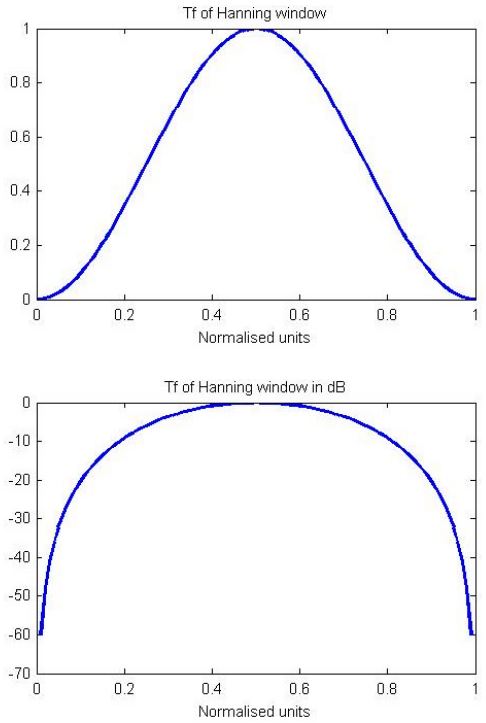

Fig. 1 Transfer function of the Hanning window

It can be seen that below 0.1, suppression is better than 20 dB. For longer impulse responses, even better suppression could be achieved, but we do not need this in practice. For a reasonable estimate of the standard deviation of the noise, this will be OK. We do not need a very precise value anyhow, only a good guess. By using this, the major part of the unmodelled dynamics and transients will be removed, and this is enough to make validation sensitive to such errors, since these are in excess to the measured RMS value.

The RMS value of the filtered sequence is different from the RMS value for the non-filtered one. For statistical tests, we need the latter. It is straightforward to calculate that application of (8) increases the variance of an input white noise by a factor of 0.25 + 1 + 0.25 = 1.5. Therefore,

$$RMS_e = \frac{RMS_{filterede}}{\sqrt{1.5}}. \qquad (9)$$

5 THE SUGGESTED BASIC ALGORITHM

According to the above thoughts, an algorithm for robust validation is as follows:

1. calculate the frequency domain residual sequence as in (4),
2. apply highpass filtering as in (8),
3. determine the RMS value from the filtered sequence,
4. correct the RMS value to remove the effect of filtering as in (9)
5. calculate the autocorrelation from the residual sequence, and apply statistical test for the presence of unmodeled dynamics and transients.

6 DETAILS AND EXTENSIONS

6.1 Statistical Properties of the Residual Sequence

Let us assume that in the residual sequence taken in the frequency domain (see (4)), no transients or unmodelled dynamics are present, that is, $e(n)$ contains only complex, approximately uncorrelated noise. This is theoretically not perfectly uncorrelated (removal of the estimated model introduces some covariance), but this effect is very small, so we may assume an almost uncorrelated sequence, with zero mean and variance (6). We use filter (8) on this sequence to remove the effect of transients and unmodelled dynamics, that is, to produce a filtered signal which contains primarily information on the noise and nonlinear products.

6.2 The Effect of the Finite Number of Samples in the Determination of the Standard Deviations

We normalize the sequence by dividing each sample by its standard deviations. This sounds very good, but in practice, we do not have access to the precise values of these standard deviations, so we approximate them by their estimates from a finite number of samples. For finite sample numbers, this slightly increases the variance of the normalized samples by the following factor (Pintelon and Schoukens, 2001):

$$\mathrm{var}\{e_M(n)\} = \frac{M-1}{M-2}\mathrm{var}\{e(n)\}, \qquad (10)$$

$$\mathrm{std}\{R_{e_M}(m)\} = \frac{M-11/12}{M-5/3}\mathrm{std}\{R_{e_M}(m)\} \qquad (11)$$

where e_M is the sample normalized by the standard deviation calculated from M samples, and R is the correlation.

For validity tests, the uncorrelatedness of the normalized sequence is tested. For this purpose, the RMS of the normalized sequence is determined from the sequence itself, so this increase is automatically taken into account.

6.3 The Effect of Highpass Filtering

We suppress lowpass contents by applying filtering with (8) as $e_h = h * e$. The filtered sequence is more or less free from unmodelled dynamics and transients. Its first moments are, assuming noise contents only

$$E\{e_h(n)\} = 0,$$
$$s_{eh}^2 = \mathrm{var}\{e_h(n)\} = \sum_{k=-1}^{1}\left(h^2(k)\right)s_e^2 = 1.5 s_e^2 \qquad (12)$$

Therefore the variance of the initial residual sequence can be estimated as

$$\hat{s}_e^2 = \frac{\hat{s}_{eh}^2}{1.5} = \frac{1}{1.5}\frac{1}{(F-2)}\sum_{k=2}^{F-1}\left|e_h(k)\right|^2 \qquad (13)$$

6.4 Validation Test

We want to test the residual sequence for uncorrelatedness. For this, we calculate the autocorrelation of the original sequence as

$$\hat{R}_{ee}(m) = \frac{1}{F-m}\sum_{k=1}^{F-m}\overline{e(k)}e(k+m) \qquad (14)$$

If the residual sequence is uncorrelated, the expected value is close to zero for $m \neq 0$. More precisely, the expected value at zero is equal to the variance of the sequence,

$$E\left\{\hat{R}_{ee}(0)\right\} = \mathrm{var}\{e(k)\} = \hat{s}_e^2, \qquad (15)$$

and the expected value at other lags is much smaller than this:

$$E\left\{\hat{R}_{ee}(m)\right\} = E\left\{\frac{1}{F-m}\sum_{k=1}^{F-m}\overline{e(k)}e(k+m)\right\} << E\left\{\hat{R}_{ee}(0)\right\} \qquad (16)$$

For completeness, we can determine the variance of the correlation estimate:

$$\mathrm{var}\left\{\hat{R}_{ee}(0)\right\} \approx \frac{2}{F-m} C_{ee}^2(0), \qquad (17)$$

but for testing purposes, we need the variance of the lag values:

$$\mathrm{var}\left\{\hat{R}_{ee}(m>0)\right\} \approx \frac{1}{2}\mathrm{var}\left\{\hat{R}_{ee}(0)\right\} \approx \frac{1}{F-m} C_{ee}^2(0) = \frac{1}{F-m} s_e^4 \qquad (18)$$

with $C_{ee}(0) = s_e^2$ being the autocovariance at zero lag (see Bendat and Piersol, 1986). From this, the standard deviation is

$$\mathrm{std}\left\{\hat{R}_{ee}(m>0)\right\} \approx \sqrt{\frac{1}{F-m}} C_{ee}(0) \approx \sqrt{\frac{1}{F-m}} s_e^2$$

$$= \sqrt{\frac{1}{F-m}} \cdot \frac{1}{1.5} \cdot \frac{1}{(F-2)} \sum_{k=2}^{F-1} \left|e_h(k)\right|^2 \qquad (19)$$

Validation tests can be executed by evaluating the correlation of the residual sequence, and checking if at nonzero lags it behaves like correlation of a sequence of zero-mean of random variables, with standard deviation as in (19). For this, consider also that $\hat{R}_{ee}(m>0)$ is zero mean, approximately complex Gaussian with circular symmetry, as a consequence of the central limit theorem. Therefore, $\left|\hat{R}_{ee}(m>0)\right|^2$ is exponential, with

$$l = 1/\mathrm{var}\left\{\hat{R}_{ee}(m>0)\right\},$$

and this can be tested for.

Percent of $\left|\hat{R}_{ee}(m>0)\right|$ values above the treshold $\sqrt{-\ln(1-P)}\,\mathrm{std}\left\{\hat{R}\right\}$:

Percent	Bound
50	$0.84\ \mathrm{std}\left\{\hat{R}\right\}$
95	$1.73\ \mathrm{std}\left\{\hat{R}\right\}$
99.5	$2.30\ \mathrm{std}\left\{\hat{R}\right\}$

A plot of the values and the limits gives direct information for the user.

6.5 Frequency Dependent Scaling

In the above considerations, we tacitly assumed that the normalized residual sequence is stationary: its variance is constant. However, since the behavior of the sequence may be different at different frequencies (colored noise), it makes sense to apply frequency dependent normalization. This can help us to have a constant-variance normalized residual sequence, even if the variance of the residual sequence (4) depends on the frequency (the time domain total noise is colored).

The theoretically correct way of doing this is to apply known standard deviation values. However, especially with the above approach, calculating the *combined effect* of noise and nonlinearities, these are a priori unknown. We need to determine the values from the sequence itself. But, for each frequency, we have one residual value only, which does not allow to make a reasonable estimate of the variance. Therefore, we chose a compromise.

When the number of residuals is sufficiently large, we evaluate (13) for segments. That is,

- (13) is used for $F < 23$.
- For $23 = F = 402$, we apply

$$\hat{s}_e^2(i) = \frac{\hat{s}_{eh}^2}{1.5} = \frac{1}{1.5}\frac{1}{21}\sum_{k=-10}^{10}\left|e_h(i+k)\right|^2, \qquad (20)$$
$$i = 11,2,...,F-10$$

- Last, when $403 = F$, we use

$$\hat{s}_e^2(i) = \frac{\hat{s}_{eh}^2}{1.5} = \frac{1}{1.5}\frac{1}{\sqrt{F-2}+1}\sum_{k=-\sqrt{F-2}/2}^{\sqrt{F-2}/2}\left|e_h(i+k)\right|^2,$$
$$i = \sqrt{F-2}/2+1,2,...,F-\sqrt{F-2}/2$$
$$\hat{s}_e^2(i) = \hat{s}_e^2(\sqrt{F-2}/2+1),\ i = 1,...,\sqrt{F-2}/2$$
$$\hat{s}_e^2(i) = \hat{s}_e^2(F-\sqrt{F-2}/2),\ i = F-\sqrt{F-2}/2+1,..F \qquad (21)$$

The rule seems to be a bit complicated. The rationale is that for large values of F, we chose to average about $\sqrt{F-2}$ values, and decreasing F, we want to provide smooth transition to the average of all values for $F < 23$.

7 IS NONLINEAR ANALYSIS AN ALTERNATIVE?

We would like to have a descriptor which accounts for nonlinearity and noise, but not for unmodelled dynamics. This is provided by the results of nonparametric nonlinear FRF analysis (Pintelon and Schoukens, 2002; Schoukens et al, 2005; Kollár et al, 2006). By using these techniques, validation of linear models will give correct results: indicate unmodelled dynamics in measurements using the best linear model, but will not be sensitive to nonlinearities, since these are incorporated into the "total variance".

This sounds very good, but may be applied with some care. Nonlinear analysis as presented in the above papers is somewhat signal dependent. Therefore, its results are perfectly applicable only when the excitation signal is similar in the nonlinear variance analysis and in the parametric system identification steps (random phase multisine with the same spectrum). Moreover, while the robust method (Kollár et al, 2006) correctly estimates the total variance, the non-robust one can slightly over- or underestimate it. Also, experiments for the evalua-

tion of nonlinearity limits are somewhat more laborious to execute than for simple linear-model identification.

8 EXAMPLE

Our practical example is linear analysis of a robot arm (the excitation is torque, the response is displacement at the other end). The cost function is about 30 times the theoretical value, that is, the RMS of the residuals is $\sqrt{30}$ times larger than expected from the residuals. This is not *very* large, but according to validation, the model is not acceptable.

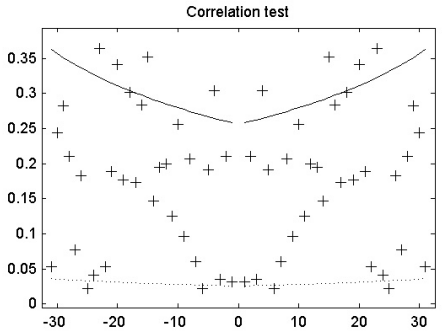

Fig. 2 Validation of the linear model by correlation test. The continuous line represents the 95% probability bound of the time domain lag correlation, based on the total variance (this paper). The dotted line is the 95% bound of linear analysis – this does not validate the model.

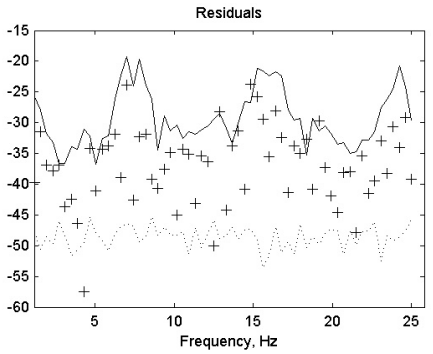

Fig. 3 Validation of the linear model. The continuous line represents the 95% probability bound based on the total variance (this paper), the dotted line the 95% bound of linear analysis. The method based on total noise variance validates the model, the one based on linear variance does not.

Applying the above algorithm, the correlation test (Pintelon and Schoukens, 2001) validates the model properly (Fig. 2).

The very same total variances can also be used to test the magnitude of the residuals (Fig 3). Again, the total variances yield model validation.

9 SUMMARY

An algorithm has been presented which discriminates between noise and nonlinear errors on the one side, and unmodeled dynamics and transients on the other side. Using this, model validation of a linear model becomes robust with respect to nonlinear errors.

REFERENCES

J. S. Bendat and A. G. Piersol (1986), "Random Data: Analysis and Measurement Procedures," John Wiley and Sons.

I. Kollár, R. Pintelon, and J. Schoukens (2006), "Frequency Domain System Identification Toolbox for Matlab: Characterizing Nonlinear Errors of Linear Models." *IFAC Symposium on System Identification, SYSID 2006*, Mar. 2006, Newcastle. *Software Demonstration Session.*

R. Pintelon, and J. Schoukens (2001). *System Identification. A Frequency Domain Approach.* IEEE Press, Piscataway, NJ.

R. Pintelon and J. Schoukens (2002), "Measurement and Modeling of Linear Systems in the Presence of Non-Linear Distortions," *Mechanical Systems and Signal Processing* 16(5), pp. 785-801.

J. Schoukens, R. Pintelon, T. Dobrowiecki and Y. Rolain (2005), "Identification of linear systems with nonlinear distortions," *Automatica*, Volume 41, Issue 3, March 2005, pages 491-504.

G. Simon, J. Schoukens, and Y. Rolain (2000), "Automatic Model Selection for Linear Time-Invariant Systems," Proceedings of the12th IFAC Symposium on System Identification, SYSID 2000, Santa Barbara, CA, USA, 21-23 June 2000, Vol. I., pp. 379-384.
Extended electronic version: Simon, G., J. Schoukens, and Y. Rolain, "Automatic Model Selection for Linear Time-Invariant Systems – Practical Issues"
`http://www.mit.bme.hu/~simon/...`
`...publications/automodel.ps.zip`