Model Error Modeling and Control Design

Lennart Ljung

Department of Electrical Engineering
Linkping University, S-581 83 Linkping, Sweden
WWW: http://www.control.isy.liu.se
Email: ljung@isy.liu.se

2000-03-07

Report no.: LiTH-ISY-R-2220
For the IFAC Symposium on System Identification, SYSID2000,
Santa Barbara, CA, July 2000

Technical reports from the Automatic Control group in Linkping are available
by anonymous ftp at the address ftp.control.isy.liu.se. This report is
contained in the pdf file 2220.pdf.
MODEL ERROR MODELING AND CONTROL DESIGN

Lennart Ljung *

* Division of Automatic Control, Linköping University,
SE-58183, Linköping, Sweden, email ljung@isy.liu.se

Abstract: Model validation and estimating the size of a possible model error is a central aspect of System Identification. In this contribution we discuss the model error concepts and model error modeling for control design. Of special interest is how to make use of periodic inputs, and how to deal with non-linear error models. The discussion is limited to SISO models and stability robustness issues.

1. INTRODUCTION

In this contribution we shall consider the estimation of linear models, of the type

\[ y(t) = G(q, \theta)u(t) + H(q, \theta)e(t) \quad (1) \]

This is of course a standard topic in the literature, e.g., (Ljung 1999a). The parameter vector \( \theta \) will be estimated using a prediction error method, leading to an estimate \( \hat{\theta}_N \) with corresponding transfer function estimate

\[ \hat{G}_N(q) = G(q, \hat{\theta}_N) \quad (2) \]

This estimate will then be used to design a regulator

\[ u(t) = K(q)(r(t) - y(t)) \quad (3) \]

so that the closed loop system

\[ y(t) = \frac{K(q)\hat{G}_N(q)}{1 + K(q)\hat{G}_N(q)} \quad (4) \]

behaves well.

The problem is that \( \hat{G}_N \) will only be an approximate description of the true system. The extensive literature on robust control design, e.g., (Zhao et al. 1996), (Skogestad and Postlethwaite 1996) deals with the question to ensure good behavior of the closed loop system, despite inaccuracies in the model \( \hat{G}_N \).

When the model has been identified from data, there is the additional question of how to estimate the uncertainty of the model. The interplay between the estimation and the design processes has received substantial attention over the past year, e.g., (Smith and Doyle 1992), (Kosut et al. 1992a), (Rangan and Poolla 1996), (Skelton 1989), (Gevers 1993), (Kosut et al. 1992b), (Smith and Dullerud 1996), (Poolla et al. 1994).

Part of this problem is to estimate, in reliable way, the model error or model uncertainty, not assuming that the true system can be described within the chosen structure (1). See, e.g., (Nimmes and Goodwin 1995), (Ljung 1999b). Then it is mostly assumed that the true system is linear, and given by a transfer function \( G_0(q) \), and that the model error is captured by estimates of the size of the frequency function

\[ \tilde{G}_N(e^{j\omega}) = \hat{G}_N(e^{j\omega}) - G_0(e^{j\omega}) \quad (5) \]

In this contribution we shall consider the more general situation of possibly non-linear model errors, and discuss the roles that an explicit model error model will play in such a set-up.

The discussion is confined to single-input-single-output systems/models and to stability robustness issues only. This is a simplification and does mean a loss of generality.
2. MODEL VALIDATION AND MODEL ERROR MODELING

To validate a model is to confront it with facts about the system. An ideal situation is that there is prior knowledge about the system which can be compared with the model’s properties. Often, however, the only “facts” that are at hand are the data themselves. It is desirable that the data used for validation are not the same as for estimation, a separate “validation data set” should be used. It then boils down to comparing what the model thinks the output should be with what the actual measured output turned out to be. That is, the residuals, the model left-overs, should be examined:

\[ \varepsilon(t) = y(t) - \hat{y}_N(t) \]  

Here \( \hat{y} \) is the simulated output

\[ \hat{y}_N(t) = \hat{G}_N(q)u(t) \]  

\( \hat{G}_N \) is the nominal model, which may be the one estimated in (2), be may also be chosen in other ways.

If a noise model \( \hat{H} \) has been estimated it is natural to consider a filtered version

\[ \varepsilon_F(t) = \hat{H}_N^{-1}(q)\varepsilon(t) \]  

and subject that to the analysis below:

2.1 Classical Correlation Analysis of Residuals

One of the most basic tests, (Draper and Smith 1981), is to compute the correlation between the regressors, in our case the past inputs, and the residuals:

\[ \hat{r}^N(\tau) = \frac{1}{N} \sum_{t=1}^{N} u(t - \tau)\varepsilon(t) \]  

It is customary to plot these estimates as a function of \( \tau \) and compare with their standard deviations to check if they have not significantly different from zero. If not, we have not traced any significant influence of \( u \) in \( \varepsilon \), and so we cannot say that the model \( \hat{G} \) has not picked up all the influence of \( u \) on \( y \). (Note the double negation: we are not saying that “\( \hat{G} \) has picked up all ...”). It is convenient to form

\[ \varphi(t) = [u(t − 1) \ldots u(t − M)]^T \]  

\[ h_N^T = \begin{bmatrix} \hat{r}^N(1) \\ \vdots \\ \hat{r}^N(M) \end{bmatrix} = \frac{1}{N} \sum_{t=1}^{N} \varphi(t)\varepsilon(t) \]  

Under the assumption that \( \varepsilon \) is white noise with variance \( \lambda \), \( h \) has a normal distribution with zero mean and variance \( \lambda/NR_N \), where

\[ R_N = \frac{1}{N} \sum_{t=1}^{N} \varphi(t)\varphi^T(t) \]  

so

\[ \xi^M_N = \frac{1}{\lambda N} \left\| \sum_{t=1}^{N} \varphi(t)\varepsilon(t) \right\|_{R_N^{-1}} \]  

will in this case have a \( \chi^2 \) distribution, and the familiar \( \chi^2 \)-test

\[ \xi^M_N < \alpha \]  

is based on this. The result is typically presented as a plot of the autocorrelation of the residuals and a plot of (9). See Figure 1.

Note that other kind of dependences can be tested quite analogously by letting \( \varphi(t) \) be other, nonlinear, functions of past inputs:

\[ \varphi(t) = f(u^t) \]  

2.2 Correlation Analysis as Model Error Modeling

Fig. 1. Traditional residual analysis: Auto- and cross-correlation functions with uncertainty regions.

The information from the cross correlation analysis between \( \varepsilon \) and \( u \), can also be interpreted as an implicit FIR model for the transfer function \( \hat{G} \) in

\[ \varepsilon(t) = \hat{G}(q)u(t) + w(t) \]  

from \( u \) to \( \varepsilon \). For control purposes, it is much more effective to present the (amplitude) frequency function of this model error model, with uncertainty bounds as in Figures 2 – 4. The data used in these figures are simulated from a second order ARMAX model. It is clear that conventional model validation corresponds to increasing the model complexity until the model error model has uncertainty bounds that include zero (as in Figure 4), since then there is no clear evidence that \( \hat{G} \) is not zero – the estimated model is then not falsified. But it is also clear that the two plots together; the model and its “sidekick”, the model error model, can be used for control design, even if the model is falsified. Look at Figure 3. According to the model error model there is significant, but rather small errors in the mid frequency range.
The model is thus falsified, but could still very well be used for control design if the information in the lower plot is taken into account.

3. MODEL ERROR MODELS AND THEIR GAINS

3.1 The General Form

A model error model is a description of how \( u \) affects \( \varepsilon \). We will typically not be interested in a detailed such description – that should rather be done as a more detailed nominal model. Primarily we focus on bounds on the gain of such a model. To go well together with linear control design, we shall work with combinations of frequency weightings and unstructured, unknown models with bounded (estimated) gain. That means that the general structure of the model from \( u \) to \( \varepsilon \) can be depicted as in Figure 5. Formally we have

\[
\varepsilon_F(t) = W_2^{-1}(q)\varepsilon(t), \quad u_F(t) = W_1(q)u(t) \tag{17}
\]

\[
\varepsilon_F(t) = \hat{g}_{mem}(u_F^{t-1}) + w(t) \tag{18}
\]

Here, \( W_1 \) and \( W_2 \) are given linear filters, superscript \( t-1 \) denotes all the signal values up to time \( t-1 \), and \( \hat{g}_{mem} \) is an uncertain model. The bottom line is that, with some amount of confidence, based on the measured data, and possibly on prior information, we should be able to say that its gain is bounded by 1:

\[
\|\hat{g}_{mem}\|_\infty < 1 \tag{19}
\]

Normalization can always be achieved using \( W_1 \) and \( W_2 \). We shall shortly discuss how to estimate such a bound.

The model \( \hat{g}_{mem} \) can come in different shapes. It could be

- Linear Time Invariant. In this case we can take either \( W_1 \) or \( W_2 \) to be unity, since all the blocks in in Figure 5 commute.
- Linear Time Varying.
- General Non-Linear.

3.2 Estimating the Gain of the Model Error Model: Distinguishing Model Error from Noise

The estimation task is to determine the filters \( W_i \) and estimate the gain of \( \hat{g}_{mem} \) in

\[
\varepsilon_F(t) = \hat{g}_{mem}(u_F^{t-1}) + w(t)
\]

i.e. estimate

\[
\sup_t \sup_{u^t} \frac{\|\hat{g}_{mem}(u^t)\|}{\|u^t\|}
\]
Now, let us say right away that it is impossible to determine this gain without prior knowledge. Even in the linear case, an arbitrarily thin resonance peak will require an arbitrarily long sequence of data, due to the uncertainty principle. After having considered a finite data record, no guarantee can be given for the “true” gain. The estimate of the gain will always reflect the data that have been measured, possibly in conjunction with prior knowledge.

A first fundamental problem is to distinguish the model error $\tilde{g}_{mem}$ from the noise contribution $w$. Loosely speaking, the noise $w$ would be that contribution to $\epsilon$ that would not change if $u$ is changed. This problem is addressed in (Smith and Dullerud 1996) by looking at trade-offs between model error size and hard amplitude bounds on $w$.

From a statistical perspective, it is more natural to try and “correlate out” $w$. There are two basic methods for that:

- Use a periodic input, and eliminate/reduce $w$ by averaging over the periods.
- Build a (parametric) model of $\tilde{g}_{mem}$ in (18) assuming $u$ and $w$ to be independent.

3.3 Using Periodic Inputs

Suppose the noise contribution is additive as in (18) and that $u$ is periodic with period $P$. Then, after a transient has died out, also $\hat{g}_N(t)$ (see (7)) will be periodic with the same period. Moreover, the part of $y$ that originates from the input will, under weak assumptions, also be periodic with period $P$. Let $\hat{y}$ denote the output over one period, obtained by averaging over all the periods, and similarly for $\hat{\epsilon}_F(t)$:

$$\hat{\epsilon}_F(t) = W_N^{-1}(q)(\hat{y}(t) - \hat{g}_N(t)), \quad t = 1, ..., P$$

From (18) we now have

$$\hat{\epsilon}_F(t) = \tilde{g}_{mem}(u_F^{-1}) + \hat{\alpha}(t)$$

where $\hat{\alpha}$ will tend to zero as $N$ tends to infinity (more and more periods are averaged over).

A simple under-bounding estimate of the gain of $\tilde{g}_{mem}$ is

$$||\tilde{g}_{mem}||^2 = \frac{\sum_{t=1}^{P} \epsilon_F^2(t)}{\sum_{t=1}^{P} u_F^2(t)}$$

(20)

A natural choice of filter is $W_1(q) = \hat{G}(q)$ (the nominal model). This makes

$$u_F(t) = \hat{g}_N(t)$$

(21)

Moreover, $W_2$ can be chosen as $W_2(q) = \hat{H}(q)$, a “noise model” for $\hat{y}(t) - \hat{g}_N(t)$,

$$\hat{y}(t) - \hat{g}_N(t) = \hat{H}(q)\nu(t)$$

(22)

with $\nu(t)$ being white noise. That will make the numerator in (20) as small as possible. Note that this makes

$$||\tilde{g}_{mem}||^2 = \frac{\sum_{t=1}^{P} [W_2^{-1}(q)\hat{y}(t) - \hat{g}_N(t)]^2}{\sum_{t=1}^{P} u_F^2(t)}$$

(23)

a measure of the proportion of “unexplained” output variation. This has a close relationship to the multiple correlation coefficient used in residual analysis, page 33 in (Draper and Smith 1981).

3.4 General Linear Black Box Model Error Models

The model error model concept gives us more freedom in investigating the residuals than the classical residual correlation test, which essentially employs a FIR model. A more general linear model

$$\epsilon(t) = \hat{G}(q)u(t) + \hat{H}(q)w(t)$$

(24)

could improve the estimate of the error model, since a model of the disturbance is used.

Instead of parametric linear models, we may apply spectral analysis to try and extract any linear influence of $u$ on $\epsilon$, (Koot 1986), (Stenman and Tjärnström 2000). In any case, there is a close relationship between the Blackman-Tukey spectral analysis estimate of this transfer function and the one obtained by a FIR-model.

Once we have a specific, linear model of the influence from $u$ to $\epsilon$, along with an uncertainty (confidence) region, the error model in Figure 5 is directly defined: Take $W_2=1$ and $W_1$ as $\hat{G}_N + \Delta$. Here $\Delta$ is the estimated variance error (at a certain confidence level), so that the amplitude curve of $W_1$ corresponds to the upper curves in (the lower plots of) Figures 2-4. If no nonlinearity has been detected in the error model, this choice of $W_1$ in Figure 5 gives a correct description of the error and uncertainty associated with the nominal model.

3.5 Non-linear Model Error Models

In the literature, most model error discussions as well as the identification-for-control approaches are dealt with in a setting where “the true system” is a high order linear model, and the models are of lower order. In practical use, it is of course more common that the model errors are ignored nonlinearities rather than unmodeled linear dynamics. From a model error perspective, this simply means that we should test non-linear models:

$$\epsilon(t) = \tilde{g}_{mem}(u_F^{-1}) + w(t)$$

(25)

In the absence of specific, suspected non-linearities, it is reasonable to test non-linear black boxes like
neural network NNFIR model, cf (Sjöberg et al. 1995):

\[ \varepsilon_F(t) = \hat{g}_{mem}(u_F(t-1), \ldots, u_F(t-m), \eta) + w(t) \]  

(26)

The number of lagged inputs can be chosen relatively small here, like \( m = 5 \) or so. To appreciate the size (gain) of any estimated non-linearity (in particular for control applications) it is natural to use the sup-norm

\[ \| \hat{g}_{mem} \|_\infty = \sup_{u_1, \ldots, u_m} \| \hat{g}_{mem}(u_1, u_2, \ldots, u_m, \eta) \|_2 \]  

(27)

Then also determine the worst case value of this gain in a properly chosen confidence region for the estimate \( \hat{\eta} \):

\[ \| \hat{g}_{mem} \| = \sup_{\hat{\eta} \in \Theta} \| \hat{g}_{mem}(u(t-1), \ldots, u(t-m), \hat{\eta}) \|_\infty \]  

(28)

4. ROBUST CONTROL DESIGN FOR UNCERTAIN SYSTEMS

4.1 Linear Model Errors

Once a model with its model error uncertainty is delivered, the question is how to design a controller that will stabilize the system robustly. By this we would mean that the chosen controller should stabilize all models in the “region” defined by the nominal model and the model error model.

In case we have used a linear model error model, this will look like a strip in the Bode, or Nyquist plot. See Figures 2-4, i.e.

\[ G \in \mathcal{G} = \{ G \mid \| G(e^{j\omega}) - G_{nom}(e^{j\omega}) \| < \Delta(\omega) \} \]  

(29)

How to achieve robust stability for such a set of models is well known: Choose a regulator \( K \), such that the complementary sensitivity function

\[ T = \frac{G_{nom}K}{1 + G_{nom}K} \]  

(30)

is less than the inverse relative model error bound:

\[ |T(e^{j\omega})| < \frac{G_{nom}(e^{j\omega})}{\Delta(\omega)}, \quad \forall \omega \]  

(31)

H\(_\infty\) techniques can be used to determine if such a \( K \) exists, for given \( G_{nom} \) and \( \Delta \). See, e.g. (Zhou et al. 1996).

4.2 Frequency Weighted Non-linear Model Error Model

The error model (18) corresponds to a closed loop block diagram as in Figure 6. This can be

![Block diagram of the feedback loop with model error](image)

rearranged to be seen as feedback between the non-linear part of the error model \( \hat{g}_{mem} \) and \( KW_1W_2/(1 + KG_{nom}) \) (keeping in mind that we only consider SISO models here). Suppose that we normalize \( W_1W_2 \) so that the gain of the non-linear part is subject to

\[ \| \hat{g}_{mem} \|_\infty < 1 \]  

(32)

The small gain theorem tells us that stability is assured if

\[ \frac{|W_1(e^{j\omega})W_2(e^{j\omega})K(e^{j\omega})|}{1 + |K(e^{j\omega})G_{nom}(e^{j\omega})|} < 1 \quad \forall \omega \]  

(33)

Comparing with (31) we realize that we just can consider the set of possible system descriptions to be linear and given by

\[ G \in \mathcal{G} = \{ G \mid \| G(e^{j\omega}) - G_{nom}(e^{j\omega}) \| < \ldots \} \]

(34)

By stabilizing any linear model in this set, i.e., achieving (31) for \( \Delta = W_1W_2 \), we have also made the linear control design robust against non-linear model errors of the type (18). (Considering only stability robustness.)

5. ESTIMATING NON-LINEARITY WEIGHTED ERROR MODELS

When dealing with non-linear model error models for control design, the option of frequency weighting is essential. As an example, suppose the true system is

\[ \hat{y}(t) = \frac{1}{(q-1)(q-\alpha)}u(t) \]  

(35)

\[ y(t) = f(\hat{y}(t)) \]  

(36)

for some static, linearly bounded, non-linearity \( f \). Suppose, for simplicity, that the nominal model is \( G_{nom}(q) = 1/(q-1)(q-\alpha) \) so the model error is this nominal system followed by the static non-linearity \( f(y) - 1 \). Suppose that \( \| f(y) - 1 \|_1 < 1 \).

The gain of this model error model is infinite, due to the infinite gain of the linear part. So, this error model does not do much good for control design.

If we now pick the weightings \( W_1(q) = 1/(1-q), W_2(q) = 1 \) the error model \( \hat{g}_{mem} \) in (18) has a
maximum gain of 1. Handling the infinite gain weighting $W_2$ in (33) is no major problem, since the denominator also has infinite gain at zero.

Note that it is only the product $W_1W_2$ that affects the robustness test, while the split into $W_1$ and $W_2$ may have a substantial influence on the norm of $\hat{y}_{mem}$. This gives an extra dimension to the model error modeling. In loose terms the weightings should be chosen so as to minimize $\|\hat{y}_{mem}\|_\infty$ while letting $W_1W_2$ have a shape that suits (33). A more comprehensive discussion of how to deal with this freedom has to be deferred to another occasion. Some hints are, however, as follows:

- Using $W_2 = \hat{H}$ (the nominal noise model) reduces the power of the output of the non-linear block as much as possible, giving the possibility of a smaller gain of the non-linear part.
- Using $W_1 = \hat{G}$ (the nominal model), will focus the error model on the relative error.

6. CONCLUSIONS

Building linear models may be deceptive. Given any data set, of arbitrary length, we can always build a linear model, that will pass standard residual analysis validation tests. The reason is that any (non-linear, time-varying) system from which quasistationary input-output sequences are collected, has a linear second order equivalent. We cannot distinguish the true system from this linear model, using just second order tests. This means that the standard linear identification setup will deliver a model with uncertainty

$$\hat{G} \pm \Delta$$

where the (variance) error $\Delta$ will tend to zero as more data are processed. Subjecting the data to the model-error-model treatment described in this contribution will give a model

$$\hat{G} \pm \Delta, \quad \Delta = W_1W_2$$

where $\Delta$ will tend to zero, only if non-linearity tests show no sign of non-linearities. Otherwise $\Delta$ will reflect a remaining uncertainty/model error. However, if we just consider stability robustness, it will be safe to treat the system as a linear one, with the given uncertainty, even if non-linearities have been detected.

7. REFERENCES


