Revisiting Total Model Errors and Model Validation

Lennart Ljung

J Syst Sci Complex (20XX) XX: 1–7

Revisiting Total Model Errors and Model Validation

LJUNG Lennart

DOI:
Received: x x 20xx / Revised: x x 20xx

Abstract The paper contains a discussion of earlier work on Total Model Errors and Model Validation. It is maintained that the recent "change of paradigm" to kernel based system identification has also affected the basis for (and interest in) giving bounds for the "total model error".

Keywords System identification, model errors, bias, variance, regularization, kernel methods.

1 Introduction

The purpose of this note is twofold:

First, the note is my congratulation to Lei Guo on his 60th birthday with deep appreciation and respect for his fundamental contributions to system identification and adaptive control.

Second, the paper will give my comments to a re-reading of some articles we wrote together in the 1990's on model validation and bounds on the total error. How do they fit in with the development of the area in the last quarter of century?

I have known Lei Guo since we met at the IFAC congress in Sydney 1993. We were both working on convergence issues of recursive algorithms and their implications for adaptive control. Lei had published the books [1, 3] and we had intense discussions on these problems. Lei worked further on these issues, which led to the fundamental contributions [4, 5]. He visited me in Linköping on two occasions in 1993-94 and also in 1995 and we wrote a few papers on recursive algorithms [6–8]. We also worked on model error issues, [17–19], which was a hot topic at the time. This is what I like to add a few comments on, in the light of the later development.

2 Total Model Errors

We limit the discussion to linear systems/models. The problem is as follows: Given a true system with transfer function $G_0(q)$, from discrete time input-output data $u(t), y(t), t = 1, \ldots .N$ we build a model $\widehat{G}_N(q)$. Give a bound on the total model error $\widetilde{G}_N = G_0 - \widehat{G}_N$. Let it be said right away that the problem as such is impossible: The linear system/model is characterized

LJUNG Lennart
Linköping University, Sweden. Email: lennart.ljung@liu.se

$^\diamond$ *This paper was recommended for publication by Editor .*

by an infinite impulse response $g_k^0, k = 1, \ldots, \infty; y(t) = \sum_{k=1}^{\infty} g_k^0 u(t-k)$. After having seen N data we have no information at all about the tail $G_{(tail)} = \{g_k^0, k = N+1, \ldots, \infty\}$. Any statement about the total model error thus must hinge upon an assumption about $G_{(tail)}$.

3 Bounds for the Total Model Error

Many approaches exist about dealing with the problem of giving useful total error bounds.

3.1 Using estimates of bias and variance

A traditional approach is to see total error as the sum of the variance of the estimate and the bias error of the estimate, e.g. [12], [9] , that is the MSE (mean square error). While the variance can be evaluated with standard probabilistic estimation techniques, e.g. [14] the bias error may require innovative approaches, e.g. [23].

3.2 Model error models

Another approach is to focus on the model residuals $\varepsilon(t) = y(t) - \widehat{y}_N(t)$ where $\widehat{y}_N(t)$ is the model output $\widehat{G}_N(q)u(t)$. For a perfect model $\varepsilon(t)$ should be independent of the output. To analyse model deficiencies, a *model error model* $G_{\text{MEM}}(q)$ can be built as $\varepsilon(t) = G_{\text{MEM}}(q)u(t)$. The total model error is then characterized by evaluating the "size" of $G_{\text{MEM}}(q)$, This idea is pursued e.g. in [13].

3.3 Stochastic embedding

Graham Goodwin and coworkers introduced the concept of *stochastic embedding*, e.g. in [2]. The idea is that the unknown model error is seen as a random variable with unknown distribution (often Gaussian). The problem of characterizing the total model errors is thus embedded in a stochastic framework where this distribution is estimated at the same time as a nominal model. There are links between stochastic embedding and the kernel methods of Section 4, which are discussed in [16].

3.4 Total model errors and model validation

It is in this problem environment Lei and I undertook the research that led to the publications [17–19]. In short, the idea is to characterize the total error $\widetilde{G}(q)$ of a model $\widehat{G}(q)$ as follows:

Let the filtered residuals be $\varepsilon(t) = L(q)[y(t) - \widehat{G}(q)u(t)]$ and define the traditional model validation test quantity which measured the correlation between residuals and past inputs:

$$\xi_N^M = \frac{1}{N} \left| \sum_{t=1}^{N} \varphi(t)\varepsilon(t) \right|_{R_M^{-1}}^2 \tag{1}$$

where $\varphi(t)$ is the vector of the M most recent inputs and R_M its covariance matrix.

Then the total model error obeys

$$\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\widetilde{G}(e^{i\omega})\right|^2|L(e^{i\omega})|^2|U_N(\omega)|^2 d\omega\right]^{1/2} \tag{2}$$

$$\leq [\frac{1}{N}\xi_N^M]^{1/2} + x_N + 2C_u\sum_{k=M}^{\infty}|\rho_k| \tag{3}$$

Here ρ_k is the impulse response of $L(q)\widetilde{G}(q)$, $|U_N|^2$ is the periodogram of the input, which is assumed to be tapered. x_N is the correlation between φ and the filtered additive noise in the system, C_u is a bound on the input.

The right hand side is basically known and computable. The two essential terms are

ξ_N^M which is the model validation residual quantity which is computed as in (1)

and the impulse response tail, which is unknown which is a basic flaw when total model errors are computed, as we said at the end of Section 2.

In [18] this error expression is applied to estimates of a model with the impulse response of Figure 1.

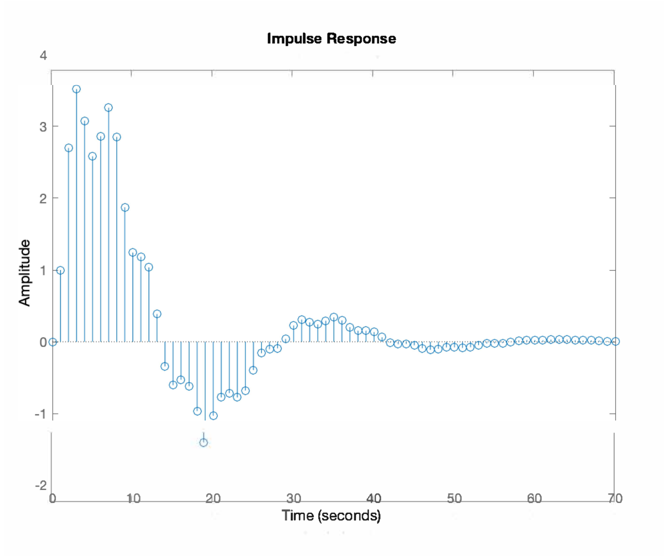

Figure 1: The impulse response of the tested system

That example and various choices of model structures and prefilters are used in [18] to discuss and illustrate the usefulness of the results for control design.

4 The Kernel Approach to System Identification (Regularization)

We will limit this discussion in this section to Finite Impulse Response (FIR) models:

$$y(t) = \sum_{k=1}^{M} g_k u(t-k) = \theta^T \varphi(t) \tag{4}$$

$$\theta^T = [g_1 \ldots, g_M] \tag{5}$$

$$\varphi(t) = [u(t-1) \ldots, u(t-M)]^T \tag{6}$$

The least squares FIR model will be

$$\widehat{\theta} = \arg\min_{\theta} \sum \|y(t) - \theta^T \varphi(t)\|^2 \tag{7}$$

An immediate idea to come to grips with the problem of unknown tails $G_{(tail)}$ is to make M large. The problem with that is that the estimates of g_k will be uncertain (have large variance), especially if the input u has poor excitation. We generate 1000 data point from the system in Figure 1 with an input that is low pass filtered white noise, and add Gaussian white noise with variance 1 to the output, and then estimate a FIR mode of order 100. The resulting model has an impulse response shown in Figure 2, left pane.

The high variance can be curbed by adding some regularization to the criterion

$$\widehat{\theta} = \arg\min_{\theta} \sum \|y(t) - \theta^T \varphi(t)\|^2 + \theta^T R \theta \tag{8}$$

for a suitable positive definite matrix R. For a particular choice of R we get a regularized FIR model, whose impulse response is shown in Figure 2, right pane.

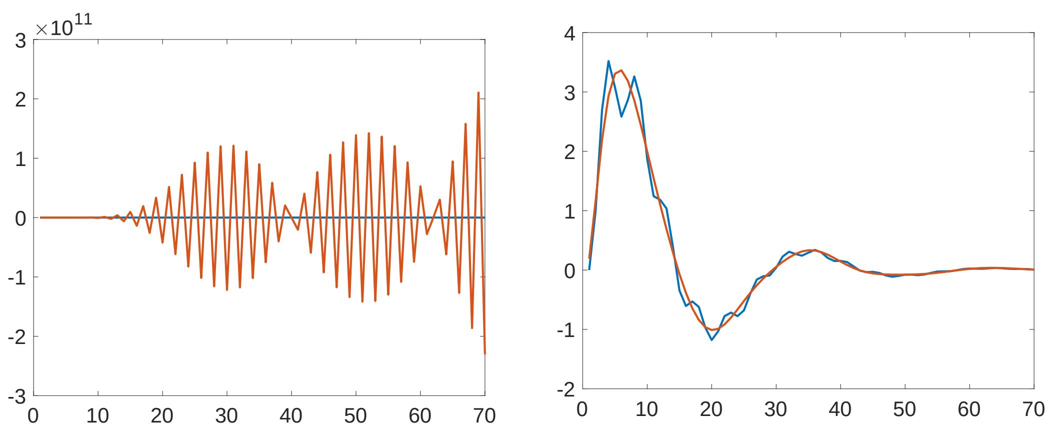

Figure 2: left pane: least squares model impulse response and true impulse response. Right pane: regularized least squares model impulse response and true impulse response.

Clearly the basic LS estimate is useless, while the regularized one gives a good estimate. This example is a very simple illustration of the basic idea of the important contribution [22].

This deals with a general identification problem, approached by the "kernel" methods that had been around for a decade in the statistics literature. The matrix R is then the kernel. The link with the kernel approach and regularization also makes contacts with statistic techniques (Ridge regression and Tikhonov regularization) which had been around for a while.

A third interpretation is also important: We recognize in the argument of (8) the posterior density of θ with a prior distribution being Gaussian with zero mean and covariance matrix R. The regularized estimate $\widehat{\theta}^R$ is thus the *MAP - maximum a posteriori estimate*, given the prior and the observations. This gives a link to Bayesian estimation theory.

The interest in the kernel approach increased further when it turned out that the methods outperformed classical system identification, [21]. One could really talk about a "change of paradigm", [15].

5 Concluding Remarks

There is nothing wrong in the work on total model errors described in Section 3, but in a way the interest has faded in "total model errors" as the discrepancy between the true system and the model. With the kernel framework of Section 4 which essentially is a Bayesian world, there is not really any "true system". All system descriptions are random variables which have a mean and a variance. There is a prior distribution encompassing what we believe about the system before we have seen the data and a posterior distribution that includes the information about the system gathered from the data measurements. The posterior distribution is a total model including both prior and data information. The variance of this posterior distribution is really a measure of the total model error.

Thus the interest in system identification has shifted to focus on forming effective and flexible prior distributions, e.g. [20].

At the same time Lei Guo's interest has shifted to fundamental limitations of adaptive control, e.g. [10], [24] and [11].

6 Finally...

I admire Lei Guo not only for his fundamental and innovative contributions to system identification and adaptive control. He is also a good friend of mine and of the Swedish control community. He has served as chairman of several of the Chinese-Swedish Control Conferences.

We are proud to have him as a foreign member of the Royal Swedish Academy of Engineering. We wish him a very happy 60th birthday and we look forward to many years of new groundbreaking insights from him.

References

[1] Chen H-F and Guo L, *Identification and Stochastic Adaptive Control*, Springer, 1991.

[2] Goodwin G C and Salgado M, A stochastic embedding approach for quantifying uncertainty in estimation of restricted complexity models, *Int. J. of Adaptive Control and Signal Processing*, 1989, **3**: 333–356.

[3] Guo L, *Time-Varying Stochastic Systems: Stability, Estimation and Control*, Jilin Science & Technology Press, Changchun, 1993. (in Chinese)

[4] Guo L, Stability of recursive stochastic tracking algorithms, *SIAM Journal of Control and Optimization*, 1994, **32**: 1195–1225.

[5] Guo L, Convergence and logarithms laws of self-tuning regulators, *Automatica*, 1995, **31**(3): 435–450.

[6] Guo L and Ljung L, Exponential stability of general tracking algorithms, *IEEE Trans. Autom. Control*, 1995.

[7] Guo L and Ljung L, Performance analysis of general tracking algorithms, *IEEE Trans. Autom. Control*, 1995.

[8] Guo L, Ljung L, and Priouret P, Performance analysis of the forgetting factor RLS algorithm, *Int. J. Adaptive Control and Signal Processing*, 1993, **7**: 525–537.

[9] Guo L and Wei C, Robust identification of systems with both bias and variance disturbances, *Chinese Science Bulletin*, 1994, **39**(20): 1673–1679.

[10] Guo L, On critical stability of discrete-time adaptive nonlinear control, *IEEE Trans. Automatic Control*, 1997, **43**(11): 1488-1499.

[11] Guo L, Exploring the maximum capability of adaptive feedback, *Int. J. Adaptive Control and Signal Processing*, 2002, **16**: 341–354.

[12] Hakvoort R G and van den Hof P M, Identification of probabilistic system uncertainty regions by explicit evaluation of bias and variance errors, *IEEE Trans. Autom. Control*, 1997, **42**(11): 1516–1528.

[13] Ljung L, Model validation and model error models, In Wittenmark B and Rantzer A, editors, *The Åström Symposium on Control*, pages 15–42. Studentlitteratur, Lund, Sweden, August 1999.

[14] Ljung L, *System Identification - Theory for the User*, Prentice-Hall, Upper Saddle River, N.J., 2nd edition, 1999.

[15] Ljung L, Chen T, and Mu B, A shift in paradigm for system identification. *International Journal of Control*, 2019, 1–8.

[16] Ljung L, Goodwin G C, and Aguero J C, Stochastic embedding revisited: A modern interpretation, In *Proc 53rd IEEE Conference on Decision and Control*, Los Angeles, Dec 2014.

[17] Ljung L and Guo L, Estimating the total model error from standard model validation tests, In *34th IEEE Conf. on Decision and Control*, New Orleans, Luisiana, 1995.

[18] Ljung L and Guo L, Classical model validation for control design purposes, *Mathematical Modelling of Systems*, 1997, **3**(1): 27–42.

[19] Ljung L and Guo L, The role of model validation for assessing the size of the unmodeled dynamics, *IEEE Trans. Autom Control*, 1997, **42**(9): 1230–1239.

[20] Pillonetto G, Chen T, Chiuso A, De Nicolao, G and Ljung L, *Regularized System Identification*, Springer Nature, 2021.

[21] Pillonetto G, Dinuzzo F, Chen T, De Nicolao G, and Ljung L, Kernel methods in system identification, machine learning and function estimation: A survey, *Automatica*, 2014, **50**: 657–682.

[22] Pillonetto G and De Nicolao G, A new kernel-based approach for linear system identification, *Automatica*, 2010, **46**(1): 81–93.

[23] Wahlberg B and Ljung L, Design variables for bias distribution in transfer function estimation, *IEEE Trans. Autom. Control*, 1986, **31**: 134–144.

[24] Xie L-L and Guo L, How much uncertainty can be dealt with by feedback? *IEEE Trans. Automatic Control*, 2000, **45**(12): 2203–2217.