Initialization and model reduction for Wiener model identification

Anna Hagenblad [*]

Department of Electrical Engineering
Linköping University
SE - 581 83 Linköping, Sweden

Abstract

The identification of nonlinear systems by the minimization of a prediction error criterion suffers from the problem of local minima. To get a reliable estimate we need good initial values for the parameters. In this paper we discuss the class of nonlinear Wiener models, consisting of a linear dynamic system followed by a static nonlinearity. By selecting a parameterization where the parameters enter linearly in the error, we can obtain an initial estimate of the model via linear regression. An example shows that this approach may be preferential to trying to estimate the linear system directly form input-output data, if the input is not Gaussian. We discuss some of the user's choices and how the linear regression initial estimate can be converted to a desired model structure to use in the prediction error criterion minimization. The method is also applied to experimental data.

1   Introduction

The nonlinear Wiener model is depicted in figure 1. It consists of a linear dynamic system $G(q)$
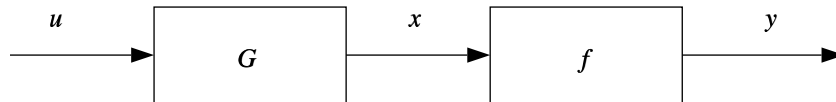

Figure 1: The Wiener model. G is a linear dynamic system and f a static nonlinear system. The input u and the output y are measurable, but not the intermediate signal x.

in series with a static nonlinearity $f(\cdot)$. Only the input u and the output y are measurable, not the intermediate signal x. Examples of Wiener models arising in practice are pH control systems, and linear systems where the measurement device has nonlinear characteristics.

The prediction error approach to identification tries to minimize the error between the measured output and the best prediction of the output. We will denote the parameters of the linear system G with θ and the parameters of the nonlinearity f with η. The prediction of the output will then depend on the parameters θ and η and we denote it with $\hat{y}(t,\theta,\eta)$. The prediction error estimate of the parameters is the one minimizing the following criterion:

$$V_N(\theta,\eta) = \frac{1}{N} \sum_{t=1}^{N} (y(t) - \hat{y}(t,\theta,\eta))^2 \tag{1}$$

---

[*]Email: `annah@isy.liu.se`

One problem with the prediction error approach is that the criterion (1) normally have several local minima. Because of the complicated structure of the predictor, the criterion has to be minimized numerically, e.g. with a Gauss-Newton numerical search (see Dennis and Schnabel, 1983). By selecting the step size appropriately, convergence may be guaranteed, but only to a local minimum. The final estimate thus depends strongly on the initial estimate.

The correct predictor to use in (1) is the expected value of $y(t)$ given the parameters and old input and output values.

$$\hat{y}(t, \theta, \eta) = \mathsf{E}(y(t)|y^{t-1}, u^{t-1}, \theta, \eta) \tag{2}$$

where u^{t-1} and y^{t-1} denotes old inputs and outputs. Depending on the characteristics of the noise and how the noise enters the system, this may be complicated to calculate. A reasonable approximation is then

$$\tilde{y}(t, \theta, \eta) = f(G(q, \theta)u(t), \eta) \tag{3}$$

where $G(q, \theta)$ denotes the linear dynamic system and $f(\cdot, \eta)$ the static nonlinearity. For the case of white measurement noise, this is also the correct predictor (2).

In order to avoid numerical problems during the minimization of (1), one of the parameters in the Wiener model, either an element of θ or of η, must be fixed. Since we only can measure the input and the output, not the intermediate signal x, a constant gain can be arbitrarily distributed between the linear and the nonlinear system.

2   Possible approaches

The rule "try simple things first" suggests that we start with trying to estimate a linear model $G(q, \theta_0)$ from the input-output data, since linear identification is a well-known area (see, e.g., Ljung, 1999). The simulated output from $G(q, \theta_0)$ can then be plotted against the measured output. If the estimate of the linear system is good enough, the nonlinear characteristics will show up clearly in the plot. We can then use some scalar function approximation technique such as splines (de Boor, 1978) to estimate the nonlinearity.

By posing some constraints on the data, this method can be shown to work. Using Bussgang's theorem (Bussgang, 1952) it is possible to show that if the input is Gaussian, this method will give a consistent estimate. This is used in Westwick and Verhaegen (1996) for subspace models and in Hunter and Korenberg (1986) using a correlation approach. In Bruls *et al.* (1997) it is used as an initial estimate when minimizing a prediction error criterion. We will call this approach the *output-error* approach.

Kalafatis *et al.* (1997) has another approach. By describing the linear system with an FIR filter,

$$x(t) = b_1 u(t-1) + \cdots + b_{n_b} u(t - n_b) \tag{4}$$

and the *inverse* of the nonlinearity with B-splines,

$$x(t) = f_1 B_1(y(t)) + \ldots f_{n_f} B_{n_f}(y(t)) \tag{5}$$

the error between equations (4) and (5) will be linear in the parameters, and a quadratic error criterion will be possible to solve with linear regression. The B_i:s here denote the B-spline basis functions. As noted above, one parameter must be fixed to avoid the trivial solution $b_i = f_i = 0$. Also other representations where the parameters enter linearly are possible, e.g. FSF filters for

the linear system and a power series for the nonlinearity. A similar approach is used by Zhu (1998b), where a rational transfer function description of the linear system makes the criterion bilinear in the parameters. In both these cases the nonlinearity is assumed to be invertible.

Since an FIR model can describe any system arbitrarily well when the number of parameters goes to infinity, and any nonlinear function can be arbitrarily well described by a piecewise linear function, we have a consistent estimate in the noise free case. If there is noise present it may not be consistent (depending on where the noise enter, and the properties of the noise), but it can still be used as an initial estimate to use in the numerical minimization of the prediction error criterion (1). This second approach will be referred to as the *internal error* approach, since it aims to minimizing the intermediate error between the output of the linear subsystem and the input of the nonlinear subsystem.

3   An example

The output-error and the internal error approaches described in the previous section were applied to a simple example system. To further point out the difficulties that may arise even in noise-free cases, no noise was added. The example system is described by the following equations:

$$G(q) = \frac{q^{-1}}{1 + 0.7q^{-1}} \tag{6}$$

$$f(x) = e^x \tag{7}$$

q^{-1} denotes the time shift operator, $q^{-1}u(t) = u(t-1)$. The input was chosen to be the following signal: $u(t) = 3t\sin(1/t)$ for $0 < t \le 1$, sampled with 1000 Hz.

To compare the methods we have plotted the simulated output from the model of the linear system versus the measured output. The plots are shown in figure 2, with the first described approach to the left, and the linear regression approach to the right. 20 FIR parameters, and
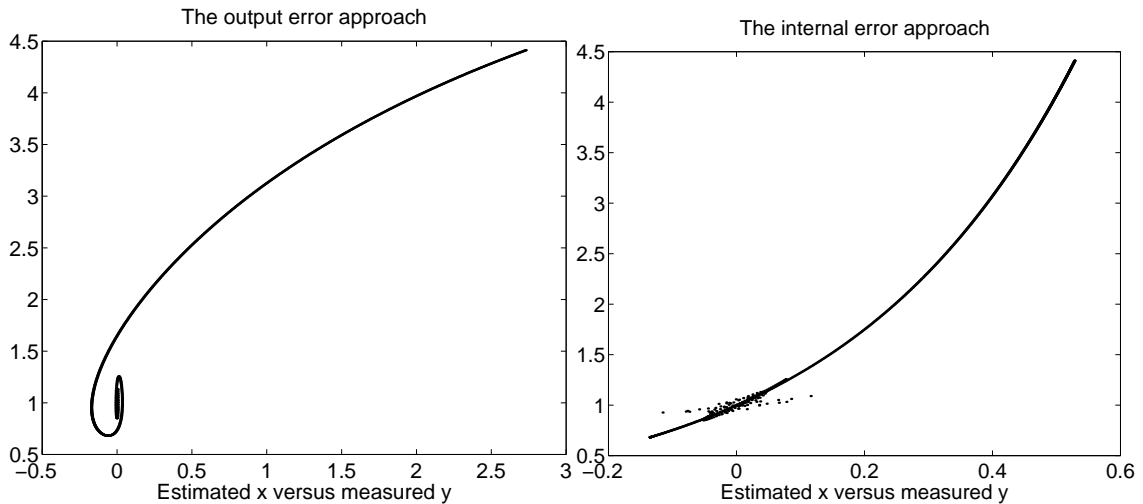

Figure 2: Example of initial estimates with non-Gaussian input. The output error approach is to the left and the internal error approach to the right.

5 B-spline parameters were used in the internal error approach. In the output-error approach a first order OE model was estimated.

An important thing to consider here is prefiltering of the data. Since e^x is positive for all x, the mean value of the output will differ from zero. This cannot be handled by the linear model. A normal precaution is to first remove the mean of the input and output. This was done in the example described here. The example above shows that the output-error approach not always works well. In the right picture, the exponential nonlinearity can clearly be seen, while the left picture shows a quite different nonlinear characteristics. The picture is of course more complex when noise is present, and also depends on where the noise enter the system: on the input, as process noise; or on the output, as measurement noise.

We may also note that if we intend to proceed with a numerical prediction error minimization, an initial estimate that appears worse may very well lead to a better global minimum. The error surface is too complicated to allow an analysis of which initial estimate will lead to the better local minimum. If we use a descent algorithm however, we can be sure that out final estimate is never worse than our initial estimate. It is thus of interest to start with as good an estimate as possible.

4   User's choices using the linear regression method

As we saw above, there may be cases where treating the system as linear to obtain an initial estimate is not the best approach. The internal error approach of Kalafatis *et al.* (1997) is then an interesting alternative. The user still have a number of choices, and we will discuss some of them in this section.

First, we have to choose the number of parameters in the linear and in the nonlinear part. For the linear part this is the number of FIR parameters. The number has to be sufficiently large to capture all important features of the impulse response. The drawback of using too many parameters is that the estimation may take longer time, and that we may need a lot of data to estimate many parameters. The second issue is more grave than the first, since a linear regression problem like this can be solved very quickly using QR factorization. If the data stems from a sampled system, which is usually the case, the sampling period should have been chosen using some knowledge about the system time constant, and this knowledge can be used also when selecting the number of FIR parameters.

Describing the nonlinear system with splines, we then have to select the number of breakpoints and their location. The first problem is similar to the selection of the number of FIR parameters for the linear system. Of course any prior knowledge should be used also here. Having selected the number of breakpoints, two possible distributions are either to spread them with equal distance between the minimum and maximum of the output, or to make sure they have equal support from data. An advantage with the latter approach may be that the characteristics of the nonlinearity may be more important where there are more output points. Also, the former approach may cause some breakpoints to have poor support, and lead to numerical problems.

5   The model reduction

Having used the internal error approach to obtain an initial estimate of the Wiener model, it is not immediate to proceed with the numerical minimization of the prediction error criterion. We will address the linear and the nonlinear system separately.

The initial estimate of the linear system is an FIR model, possibly with a large number of parameters. Often another model structure is desired, such as a rational transfer function representation (output-error) or a state-space model.

A straight-forward approach to the linear model reduction is to use the FIR model to simulate input-output data for the linear system, and then identify it from data. Since no noise is present in the simulation, this may be done with an ARX model (see Ljung, 1999).

Another possibility is to use balanced model reduction, in the state space framework. Using a canonical form we can convert the FIR model to a state-space representation with as many states as there are parameters. The state-space representation can then be transformed to a balanced realization, and the least significant singular values removed. Balanced realization model reduction is treated in, e.g., Zhou *et al.* (1995).

Since the initial estimate gives us the inverse of the nonlinearity, we must first invert it. This may be a problem, since there is no guarantee that the estimate obtained is invertible. One possibility is to require invertibility in the initial estimation. The problem will then no longer be linear regression, but it will still be a quadratic minimization problem, now with linear constraints, which also has a unique solution, and can be solved relatively fast.

During the identification process, a plot of the simulated x versus the measured y will often give leads about how the nonlinearity should be inverted. Some points may be outliers that can be discarded.

The inverted nonlinearity will be a piecewise linear function, with a lot of breakpoints. A large number of breakpoints will make the prediction error criterion minimization harder and more time-consuming. A large number of breakpoints may reduce the bias of the estimate, but increase the variance.

To get an automatic breakpoint reduction, the **newnot** procedure of de Boor (1978) may be used. It reduces the breakpoints by looking at a higher order variation of the function, and selecting breakpoints for this variation to be distributed evenly over the interval. There are also other possibilities, but it is as well important to examine the data. In the plots of the estimated x versus the measured y, the general shape of the nonlinearity is often visible. More breakpoints are needed where the nonlinearity has abrupt changes.

6 An application to experimental data

As an example, we will study the identification of a distillation column from measured data. This is the same data as used in (Bruls *et al.*, 1997). The data is there described as follows:

> The inputs are the temperatures at two different plates inside the column and they are sampled every two minutes. The output of the system is the product quality measured by gas chromatography (GC) and is available at irregular sampling intervals (but always an integral multiple of 2 minutes) of 18 and 20 minutes.

The input is here the difference between the two inputs mentioned above. The input and output data is shown in figure 3.

The output between the sampling intervals is calculated via zero order hold. The system turns out to have a very large time constant. This may partly depend on the sampling interval; if another sampling interval had been used, this may not have been so pronounced. One may also consider another approximation of the output between the sampling intervals than zero order hold. To get a good estimate with the internal error approach we have used 200 FIR parameters and 20 B-splines. To use more is hardly possible with this limited number of data. The breakpoints were distributed with equal distance between the maximum and minimum value of the output y.
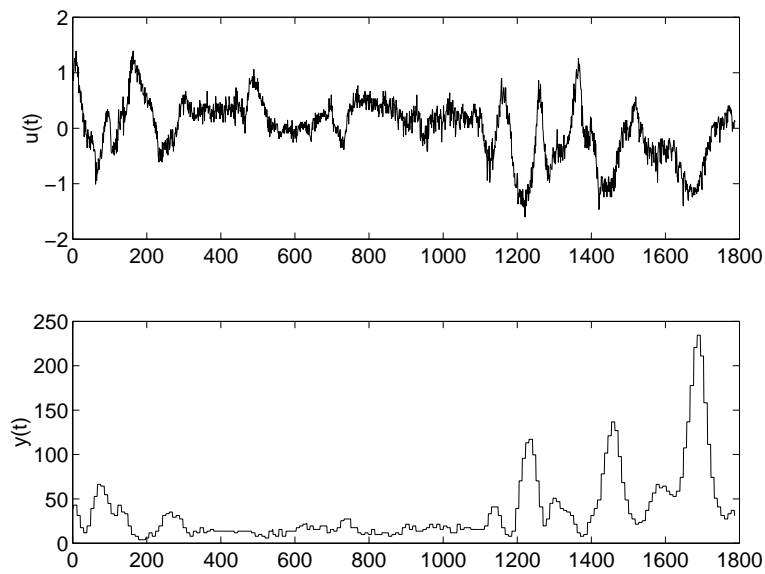
Figure 3: The input and output data from the distillation column.

The data set at identification is often divided into estimation and validation data. We have chosen not to do this here, since there are relatively few data, and also because the second part of the data set excites parts of the nonlinearity not seen in the first part. Thus, all data are used for both estimation and validation.

The order of the linear transfer function in the output-error approach was selected to 2. 6 breakpoints were used to estimate the nonlinearity.

The initial estimate is visualized in figure 4 by plotting the simulated x versus the measured y. An exponential-like nonlinearity can clearly be seen.
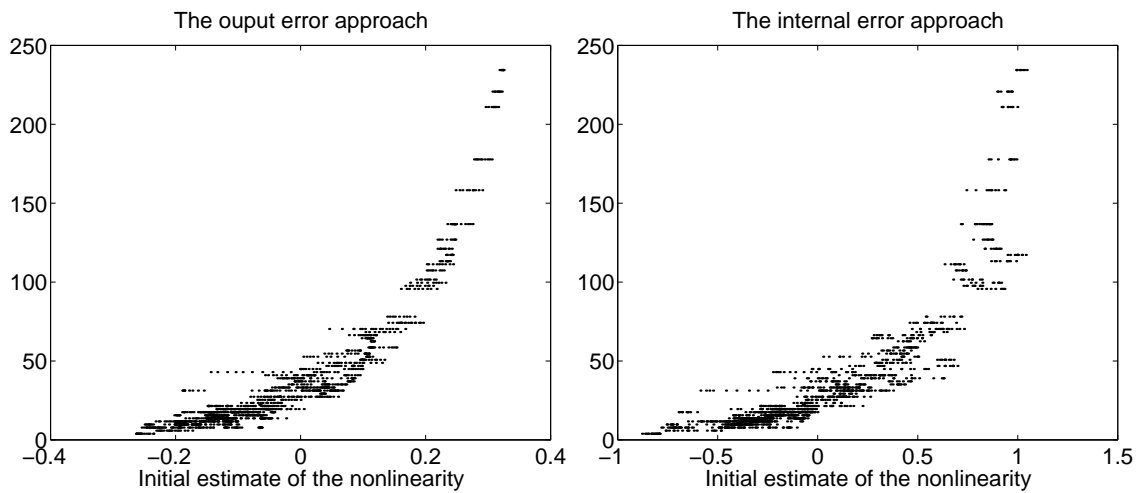

Figure 4: Plot of the estimated nonlinearities after initialization. The estimate from the output-error approach is shown to the left, and the estimate from the internal error approach to the right.
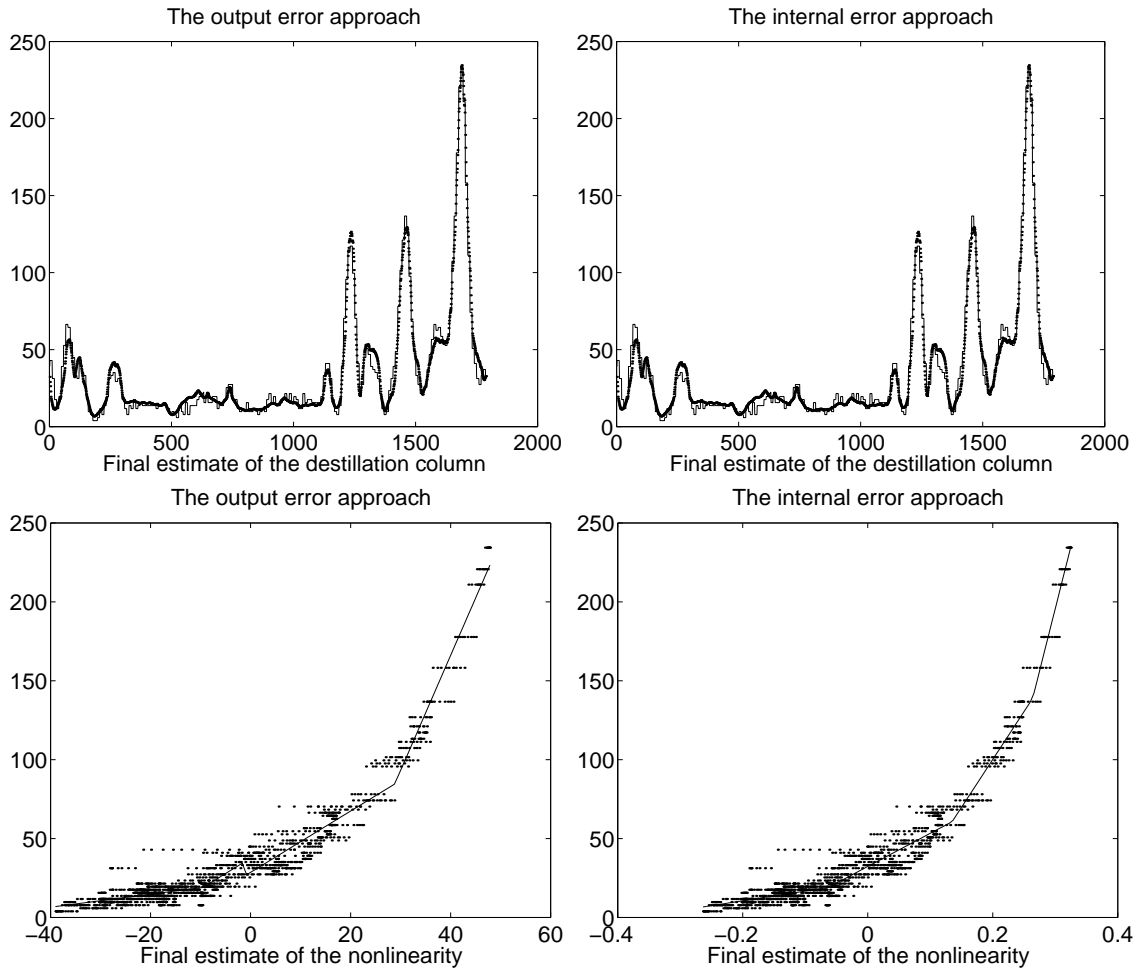
Figure 5: Plot of the final estimates. The upper plots show the estimated and measured output; the estimated values are marked with dots while the measured values are connected with a straight line. The lower plots shows the estimated x versus the measured y as dots; the straight line represents the estimated nonlinearity. The output-error approach is to the left and the internal error approach to the right.

After converting the linear regression model as described in section 5, Gauss-Newton search was used to minimize the prediction error criterion. The same orders as for the output-error approach was used; a second order linear transfer function and a piecewise linear function with 6 breakpoints. The final estimates are shown in figure 5. The two upper plots show the estimated and measured output. The two lower plots show the simulated x versus the measured y, as before.

As can be seen in the plots, there are no large differences between the estimates obtained using the two different approaches for the initial estimate. The value of the error criterion (1) was 48.9 for the output error initialization and 44.5 for the internal error initialization.

7   Conclusions

We have described some approaches to the identification of Wiener models, with focus on the prediction error method, and especially expressed the importance of finding a good initial estimate. Two different methods have been discussed; one where the whole system is treated as a linear system, and one where a particular parameterization makes the parameters enter linearly in the error criterion. The former gives a consistent estimate if the input and noise is Gaussian, but examples show that this is not necessarily the case for other inputs signals, even if no noise is present. The latter method is then interesting and have given good results in the simulations shown. Also issues on how to proceed to be able to numerically minimize the prediction error criterion, and how to perform the model reduction needed, are addressed. The method is also applied to experimental data with some success.

8   Acknowledgment

References

Bruls, J., C. T. Chou, B. R. J. Haverkamp, and M. Verhaegen (1997). "Linear and non-linear system identification using separable least-squares," Submitted to European Journal of Control.

Bussgang, J. J. (1952). "Crosscorrelation functions of amplitude-distorted Gaussian signals," Tech. Rep. 216, MIT Research Laboratory of Electronics.

de Boor, C. (1978). *A Practical Guide to Splines*, vol. 27 of *Applied mathematical sciences*, Springer.

Dennis, J. E., Jr and R. B. Schnabel (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ.

Hunter, I. W. and M. J. Korenberg (1986). "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models," *Biological Cybernetics*, , no. 55, pp. 135–144.

Kalafatis, A. D., L. Wang, and W. R. Cluett (1997). "Identification of Wiener-type nonlinear systems in a noisy environment," *Int J Control*, **66**, no. 6, pp. 923–941.

Ljung, L. (1999). *System Identification, Theory for the User*, Prentice Hall, 2nd edn.

Westwick, D. and M. Verhaegen (1996). "Identifying MIMO Wiener systems using subspace model identification methods," *Signal Processing*, , no. 52, pp. 235–258.

Zhou, K., J. C. Doyle, and K. Glover (1995). *Robust and Optimal Control*, Prentice Hall.

Zhu, Y. (1998a). "Identification of Hammerstein models for control," in *Proceedings of the 37th IEEE Conference on Decision and Control*, Tampa, Florida, USA, pp. 219–220.

Zhu, Y. (1998b). "Parametric Wiener model identification for control," Personal Communication at the 37th IEEE CDC, Tampa, Florida (see also (Zhu, 1998a)). To appear at IFAC World Congress 1999, Beijing.