1

Identification

What you will learn: This chapter introduces the reader to the basic methods that are used in system identification. It shows what the user can expect from the identification framework to solve her/his modelling problems. For that purpose the following topics are addressed:

- An estimator is a random variable that can be characterized by its mean value and covariance matrix (see Exercises 1.a, 7).

- The stochastic characteristics of an estimator depend on the number of raw data, the experiment design, the choice of the cost function, ... (see Exercises 1.b, 8, 9).

- The estimates are asymptotically normally distributed when a growing amount of raw data is processed (see Exercise 2).

- Noise disturbances on the regressor (e.g., the input data) can create a systematic error on the estimate, and the choice of the regressor variable has an impact on the final result. Specific methods are needed to deal with that problem (see Exercises 3, 4, 12, 13).

- The choice of the cost function influences the properties of the estimator (see Exercises 5.a, 5.b, 8, 9).

- The "optimal" choice of the cost function depends on the disturbing noise probability density function (see Exercise 9).

- Least squares problems can be explicitly solved if the model is linear-in-the-parameters (see Exercise 6).

- The numerical properties of the algorithms are strongly affected by the choice of the model parameters (see Exercise 6).

- The model complexity (number of unknown model parameters) should be carefully selected. Systematic tools are available to help the user to make this choice (see Exercise 10, 11).

1.1 INTRODUCTION

The aim of system identification is to extract a mathematical model $M(\theta)$ from a set of measurements Z. Measurement data are disturbed by measurement errors and process noise, described as disturbing noise n_z on the data:

$$Z = Z_0 + n_z. \tag{1-1}$$

Since the selected model class M does not, in general, include the true system S_0, model errors appear:

$$S_0 \in M_0 \text{ and } M_0 = M + M_\varepsilon, \tag{1-2}$$

with M_ε the model errors. The goal of the identification process is to select M and to tune the model parameters θ such that the "distance" between the model and the data becomes as small as possible. This distance is measured by the cost function that is minimized. The selection of these three items (data, model, cost function) sets the whole picture; all the rest are technicalities that do not affect the quality of the estimates. Of course this is an oversimplification. The numerical methods used to minimize the cost function, numerical conditioning problems, model parameterizations, and so on are all examples of very important choices that should be properly addressed in order to get reliable parameter estimates. Failing to make a proper selection can even drive the whole identification process to useless results. A good understanding of each of these steps is necessary to find out where a specific identification run is failing: Is it due to numerical problems, convergence problems, identifiability problems, or a poor design of the experiment?

In this chapter we will study the following issues:

- What is the impact of noise on the estimates (stochastic and systematic errors)?
- What are the important characteristics of the estimates?
- How to select the cost function?
- How does the choice of the cost function affect the results?
- How to select the complexity of the model? What is the impact on the estimates?

1.2 ILLUSTRATION OF SOME IMPORTANT ASPECTS OF SYSTEM IDENTIFICATION

In this book almost all the estimators that will be studied and used are based on the minimization of a cost function. It is a measure for the quality of the model and it is calculated starting from the errors: the differences between the actual measurements and their modeled values. We will use mostly a least squares cost function that is the sum of the squared errors.

In this section we present a simple example to illustrate some important aspects of system identification. Specifically, the impact of the noise on the final estimates is illustrated. It will be shown that zero mean measurement noise can result in systematic errors on the estimates (the mean of the parameter errors is not equal to zero!). Also the uncertainty is studied. Depending on the choice of the cost function, a larger or smaller noise sensitivity will be observed. All these aspects are studied using a very simple example: the measurement of the value of a resistance starting from a series of voltage and current measurements.

1.2.1 Least squares estimation: A basic approach to system identification

Exercise 1.a (Least squares estimation of the value of a resistor) Goal: Estimate the resistance value R_0 starting from a series of repeated current and voltage measurements:

$$u_0(t) = R_0 i_0(t), \quad t = 1, 2, \ldots, N \tag{1-3}$$

with u_0, i_0 the exact values of the voltage and the current.

Generate an experiment with $N = 10, 100, 1000$, and $10{,}000$ measurements. The current i_0 is uniformly distributed in $[-i_{max}, i_{max}]$ with $i_{max} = 0.01$ A (use the MATLAB® routine `rand(N,1)`), $R_0 = 1000$. The current is measured without errors; the voltage is disturbed by independent, zero mean, normally distributed noise n_u with $N(0, \sigma_u^2 = 1)$.

$$\begin{aligned} i(t) &= i_0(t) \\ u(t) &= u_0(t) + n_u(t) \end{aligned}, \quad t = 1, 2, \ldots, N \tag{1-4}$$

To measure the distance between the model and the data, we select in this exercise a least squares cost function: $V(R) = \frac{1}{N}\sum_{t=1}^{N} (u(t) - Ri(t))^2$. Notice that many other possible choices can be made.

The least squares estimate $\hat{R}$ is defined as the minimizer of the cost function $V(R)$:

$$\hat{R} = \arg\min_R V(R) \tag{1-5}$$

- Show that the minimizer of (1-5) is given by

$$\hat{R} = \sum_{t=1}^{N} u(t)i(t) / \sum_{t=1}^{N} i(t)^2. \tag{1-6}$$

- Generate 100 data sets with a length $N = 10, 100, 1000$, and $10{,}000$, and calculate the estimated value $\hat{R}$ for each N.
- Plot the 100 estimates, together with the exact value for each N, and compare the results.

Observations (see Figure 1-1) From the figure it is seen that the estimates are scattered around the exact value. The scattering decreases for an increasing number N. It can be shown that under very general conditions, the standard deviation of the estimates decreases as $1/\sqrt{N}$. This is further elaborated in the next exercise.

Exercise 1.b (Analysis of the standard deviation) In this exercise, it is verified how the standard deviation varies as a function of N. Consider the resistance

$$u_0(t) = R_0 i_0(t), \; t = 1, 2, \ldots, N. \tag{1-7}$$

with a constant current $i_0 = 0.01$ A, and $R_0 = 1000\ \Omega$. Generate 1000 experiments with $N = 10, 100, 1000$, and $10{,}000$ measurements. The current is measured without errors, the voltage is disturbed by independent, zero mean Gaussian distributed noise $n_u \sim N(0, \sigma_u^2 = 1)$ (use the MATLAB® routine `randn(N,1)`):

$$\begin{aligned} i(t) &= i_0 \\ u(t) &= u_0(t) + n_u(t) \end{aligned}, \quad t = 1, 2, \ldots, N \tag{1-8}$$

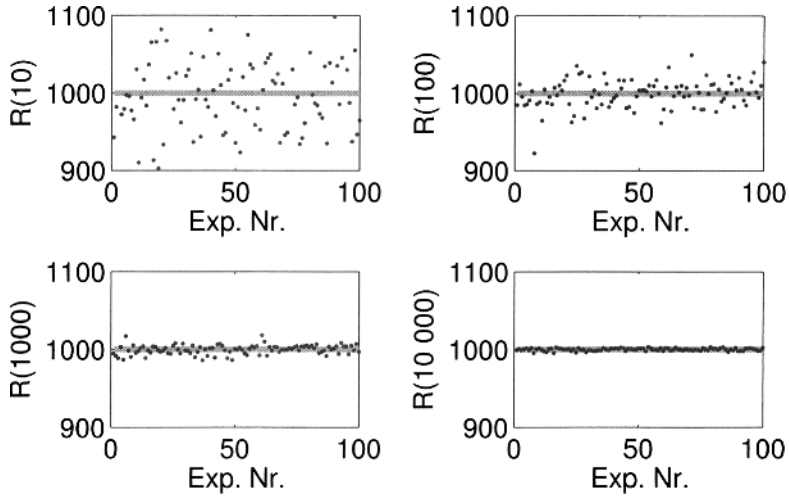

Figure 1-1 Estimated resistance values $\hat{R}(N)$ for N = 10, 100, 1000, and 10,000 for 100 repeated experiments. Gray line: exact value; dots: estimated value.

- Calculate for the four values of N the standard deviation of $\hat{R}$ using the MATLAB® routine `std(x)`. Make a loglogplot of the standard deviation versus N.
- Compare it with the theoretical value of the standard deviation that is given in this simplified case (constant current) by

$$\sigma_R = \frac{1}{\sqrt{N}} \frac{\sigma_u}{i_0}. \tag{1-9}$$

Observations (see Figure 1-2) From the figure it is seen that the standard deviation decreases as $1/\sqrt{N}$. Collecting more data makes it possible to reduce the uncertainty. To get a reduction with a factor 10 in uncertainty, an increase of the measurement time with a factor

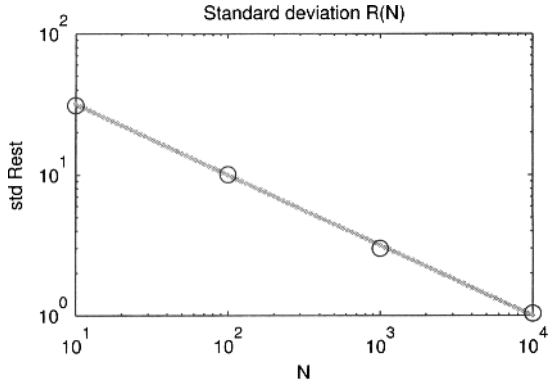

Figure 1-2 Experimental (black circles) and theoretical (gray dots) standard deviation on $\hat{R}$ as a function of N. The error drops with $\sqrt{10}$ if the number of data N grows with a factor 10.

100 is needed. This shows that the measurement time grows quadratically with the required noise reduction, and hence it still pays off to spend enough time on a careful setup of the experiment in order to reduce the level of the disturbing noise σ_u on the raw data.

Remark: For the general situation with a varying current, the expression for the standard deviation σ_R for a given current sequence $i_0(t)$ is

$$\sigma_R = \frac{\sigma_u}{\sqrt{\sum_{t=1}^{N} i_0^2(t)}} \tag{1-10}$$

Exercise 2 (Study of the asymptotic distribution of an estimate) The goal of this exercise is to show that the distribution of an estimate is asymptotic for $N \to \infty$ normally distributed, and this is independent of the distribution of the disturbing noise (within some regularity conditions, like finite variance, and a restricted "correlation" length of the noise).

Consider the previous exercise for $N = 1, 2, 4, 8$, and 10^5 repetitions. Use a constant current $i_0 = 0.01$ A, measured without errors. For the voltage we consider two situations. In the first experiment, the voltage is disturbed by independent, zero mean Gaussian distributed noise $N(0, \sigma_u^2 = 0.2^2)$. In the second experiment the voltage noise is uniformly distributed in $[-\sqrt{3}\sigma_u, \sqrt{3}\sigma_u]$.

- Verify that the standard deviation of the uniformly distributed noise source also equals σ_u.
- Calculate the least squares solution [see equation (1-6)] for $N = 1, 2, 4, 8$ and repeat this 10^5 times for both noise distributions. Plot the estimated pdf for the eight different situations. The pdf can be estimated by making a proper normalization of the histogram of the estimates (use the MATLAB® routine `hist(x)`). The fraction of data in each bin should be divided by the width of the bin.
- Calculate the mean value and the standard deviation over all realizations (repetitions) for each situation, and compare the results.

□

Observations (see Figure 1-3) From the figure it is seen that the distribution of the estimates depends on the distribution of the noise. For example, for $N = 1$, the pdf for the Gaussian disturbed noise case is significantly different from that corresponding to the uniformly disturbed experiment. These differences disappear for a growing number of data per experiment (N increases), and for $N = 8$ it is impossible to identify a different shape visually. The uniform distribution converges to the Gaussian distribution for growing values of N. This is a general valid result.

In this case, the mean value and the variance is the same for both disturbing noise distributions, and this for each value of N. This is again a general result for models that are linear in the measurements (e.g., $y_0 = au_0$ is linear in u_0, while $y_0 = au_0^3$ is nonlinear in the measurements). The covariance matrix of the estimates depends only on the second-order properties of the disturbing noise. This conclusion cannot be generalized to models that are nonlinear in the measurements. In the latter case, the estimates will still be Gaussian distributed, but the mean value and variance will also depend on the distribution of the disturbing noise.

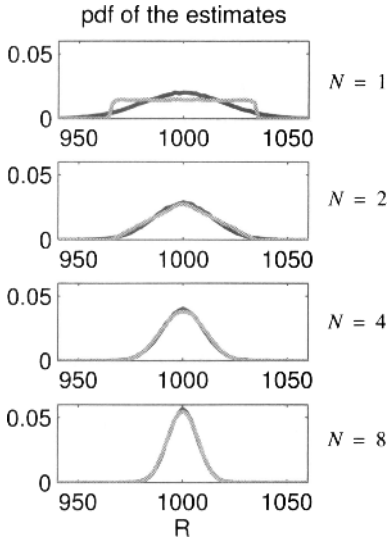

Figure 1-3 Evolution of the pdf of $\hat{R}$ as a function of N, for $N = 1, 2, 4, 8$. Black: Gaussian disturbing noise; Gray: uniform disturbing noise.

1.2.2 Systematic errors in least squares estimation

In the previous section it was shown that disturbing noise on the voltage resulted in noisy estimates of the resistor value, the estimated value of the resistor varies from one experiment to the other. We characterized this behavior by studying the standard deviation of the estimator. The mean value of these disturbances was zero: the estimator converged to the exact value for a growing number of experiments. The goal of this exercise is to show that this behavior of an estimator cannot be taken for granted. Compared with the Exercises 1.a–2, we add in the next two exercises also disturbing noise on the current. The impact of the current noise will be completely different from that of the voltage noise, besides the variations from one experiment to the other, a systematic error will arise. This is called a bias.

Exercise 3 (Impact of noise on the regressor (input) measurements) Consider Exercise 2 for $N = 100$, and 10^5 repetitions. The current i_0 is uniformly distributed between $[-10, 10]$ mA. It is measured this time with white disturbing noise added to it: $i(t) = i_0 + n_i(t)$, with a normal distribution $N(0, \sigma_i^2)$. The voltage measurement is also disturbed with normally distributed noise: $N(0, \sigma_u^2 = 1)$.

- Repeat the simulations of the previous exercise once without and once with noise on the current. Vary the current noise standard deviation in 3 successive simulations: $\sigma_i = 0, 0.5, 1$ mA.
- Calculate the least squares solution [see eq. (1-6)] for $N = 100$ and repeat this 10^5 times for all situations and plot the pdf for each of them.
- Calculate the mean value and the standard deviation over all realizations (repetitions) for each situation, and compare the results.

□

Observations (see Figure 1-4) From the figure it is seen that the distribution of the estimates depends strongly on the presence of the noise on the current measurement. Not only is the standard deviation affected, but also a visible bias grows with the variance of the current noise. This result is closely linked to the fact that the current is used as regressor or independent variable that makes the voltage a dependent variable: We used a model where the current is the input, and the voltage is the output. Whenever the measurement of the input variable is disturbed by noise, bias problems will appear unless special designed methods are used. These will be studied in Section 1.6.

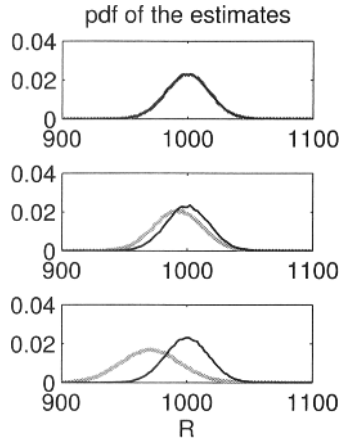

Figure 1-4 Evolution of the pdf of $\hat{R}$ as a function of the noise level at the current. Black: Only noise on the voltage $\sigma_u = 1$ V. Gray: Noise on the voltage $\sigma_u = 1$ V and the current. $\sigma_i = 0, 0.5, 1$ mA (top, middle, bottom).

Exercise 4 (Importance of the choice of the independent variable or input) In Exercise 3 it became clear that noise on the input or independent variable creates a bias. The importance of this choice is explicitly illustrated by repeating Exercise 2, where the disturbing noise is only added to the voltage: $i(t) = i_0(t)$, $u(t) = u_0 + n_u(t)$ with $n_u(t) \sim N(0, \sigma_u^2 = 1)$. In this exercise the same data are processed two times:

- Process the data using the current as independent variable, corresponding to the function $u(t) = Ri(t)$ and an estimate of R:

$$\hat{R} = \sum_{t=1}^{N} u(t)i(t) / \sum_{t=1}^{N} i(t)^2. \tag{1-11}$$

- Process the data using the voltage as independent variable, corresponding to $i(t) = Gu(t)$, with G the conductance:

$$\hat{G} = \sum_{t=1}^{N} u(t)i(k) / \sum_{t=1}^{N} u(t)^2 \text{ and } \hat{R} = 1/\hat{G}. \tag{1-12}$$

- Repeat each experiment 10^5 times for $N = 100$, then calculate and plot the pdf of the estimated resistance for both cases.

□

Discussion (see Figure 1-5) Whenever the measurement of the variable that appears squared in the denominator of (1-11) or (1-12) is disturbed by noise, a bias will become visible. This shows that the signal with the highest SNR should be used as independent variable or input in order to reduce the systematic errors. The bias will be proportional to the inverse SNR (noise power/signal power).

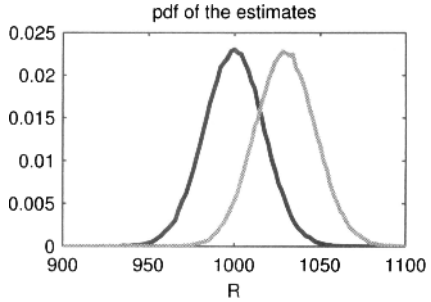

Figure 1-5 Study of the impact of the selection of the independent variable for the estimation of the resistance. Only the voltage is disturbed with noise. The pdf of the estimated resistance is shown for the independent variable being the current (black) or the voltage (gray).

1.2.3 Weighted least squares: optimal combination of measurements of different quality

The goal of this section is to combine measurements with different quality. A first possibility would be to throw away the poorest data, but even these poor data contain information. It is better to make an optimal combination of all measurements taking into account their individual quality. This will result in better estimates with a lower standard deviation. The price to be paid for this improvement is the need for additional knowledge about the behavior of the disturbing noise. While the least squares (LS) estimator does not require information at all about the disturbing noise distribution, we have to know the standard deviation (or in general, the covariance matrix) of the disturbing noise in order to be able to use the improved weighted least squares (WLS) estimator, illustrated in this exercise.

Exercise 5.a (combining measurements with a varying SNR: Weighted least squares estimation) Estimate the resistance value starting from a series of repeated current and voltage measurements:

$$u_0(t) = R_0 i_0(t), \quad t = 1, 2, \ldots, N \tag{1-13}$$

with u_0, i_0 the exact values of the voltage and the current. Two different voltmeters are used, resulting in two data sets, the first one with a low noise level, the second one with a high noise level.

- Generate an experiment with N measurements, i_0 uniformly distributed in $[-0.01, 0.01]$ A, $R_0 = 1000\ \Omega$. The current is measured without errors, the voltage measured by the 2 voltmeters is disturbed by independent, zero mean, normally distributed noise n_u with $N(0, \sigma_u^2 = 1)$ for the first, good voltmeter and $N(0, \sigma_u^2 = 16)$ for the second, bad one.

$$\begin{aligned} i(t) &= i_0(t) \\ u(t) &= u_0(t) + n_u(t) \end{aligned}, \quad t = 1, 2, \ldots, N \tag{1-14}$$

- Calculate the weighted least squares solution as the minimizer of

$$V_{\text{WLS}} = \frac{1}{N}\sum_{t=1}^{2N} \frac{(u(t) - Ri(t))^2}{w(t)} \tag{1-15}$$

using (1-16), given below:

$$\hat{R} = \frac{\sum_{t=1}^{2N} \frac{u(t)i(t)}{w(t)}}{\sum_{t=1}^{2N} \frac{i(t)^2}{w(t)}}, \tag{1-16}$$

with $w(t)$ the weighting of the tth measurement: $w(t) = \sigma_{u1}^2$ for the measurements of the first voltmeter and $w(t) = \sigma_{u2}^2$ for the measurements of the second one.

- Repeat this exercise 10^5 times for $N = 100$. Estimate the resistance also with the least squares method of Exercise 1.a. Make an histogram of both results.

Observations (see Figure 1-6) From the figure it is seen that the estimates are scattered

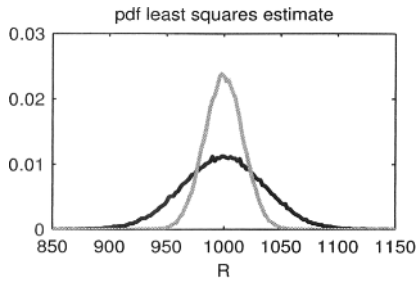

Figure 1-6 Estimated resistance values for $N = 100$, combining measurements of a good and a bad voltmeter. Black: pdf of the least squares; gray: pdf of the weighted least squares estimates.

around the exact value. However, the standard deviation of the weighted least squares is smaller than that of the least squares estimate. It can be shown that the inverse of the covariance matrix of the measurements is the optimal weighting for least squares methods.

Exercise 5.b (Weighted least squares estimation: A study of the variance) In this exercise we verify by simulations the theoretical expressions that can be used to calculate the variance of a least squares estimator and a weighted least squares estimator. It is assumed that there is only noise on the voltage. The exact, measured current is used as regressor (input). The theoretical variance of the linear least squares estimator (no weighting applied) for the resistance estimate is given by

$$\hat{\sigma}_{\text{LS}}^2 = \frac{\sum_{t=1}^{2N} \sigma_u^2(t) i^2(t)}{\left(\sum_{t=1}^{2N} i(t)^2\right)^2}, \tag{1-17}$$

and the variance of the weighted least squares estimator using the variance on the output (the voltage) as weighting is

$$\hat{\sigma}^2_{\mathrm{WLS}} = \frac{1}{\sum_{t=1}^{2N} \frac{i(t)^2}{\sigma_u^2(t)}}. \tag{1-18}$$

- Consider Exercise 5.a, calculate the theoretical value for the standard deviation, and compare this with the results obtained from the simulations.

Observations A typical result of this exercise is:

theoretical standard deviation LS, 35.6; experimental standard deviation, 35.8

theoretical standard deviation WLS, 16.8; experimental standard deviation, 16.8

Remark: The expressions (1-17) and (1-18) for the theoretical values of the variance are valid for a given input sequence. If the averaged behavior over all (random) inputs is needed, an additional expectation with respect to the input current should be calculated.

1.2.4 Models that are linear-in-the-parameters

The least squares estimates of the resistor that have been studied thus far were based on the minimization of the weighted cost function

$$V(R) = \frac{1}{N}\sum_{t=1}^{N} \frac{(u(t) - Ri(t))^2}{w(t)}, \tag{1-19}$$

with u, i the measured voltage (output) and current (input), respectively.

In general, the difference between a measured output $y(t)$ and a modeled output $\hat{y}(t) = g(t, u_0, \hat{\theta})$ is minimized for a given input signal $u_0(t)$. All model parameters are grouped in $\theta \in \mathrm{R}^{n_\theta}$. This can be formulated under a matrix notation for models that are linear-in-the-parameters. Define the signal vectors $\hat{y}$, u_0, $g \in \mathrm{R}^N$, for example:

$$y^T = \{y(1), \ldots, y(N)\} \tag{1-20}$$

and a positive weighting matrix $W \in \mathrm{R}^{N \times N}$. Then the weighted least squares cost function becomes

$$V_{\mathrm{WLS}} = (y - g(u_0, \theta))W^{-1}(y - g(u_0, \theta))^T. \tag{1-21}$$

For a diagonal matrix $W_{ii} = w(t)$, $W_{ij} = 0$ elsewhere, and equation (1-21) reduces to

$$V_{\mathrm{WLS}} = \frac{1}{N}\sum_{t=1}^{N} \frac{(y(t) - g(t, u_0, \theta))^2}{w(t)}. \tag{1-22}$$

The estimate $\hat{\theta}$ is found as the minimizer of this cost function:

$$\hat{\theta} = \arg\min_{\theta} V_{\mathrm{WLS}}(\theta). \tag{1-23}$$

In general it is impossible to solve this minimization problem analytically. However, if the model is linear-in-the-parameters, then it is possible to formulate the solution explicitly, and it is also possible to calculate it in a stable numerical way with one instruction in MATLAB®. A model is called linear-in-the-parameters if the output is a linear combination of the model parameters:

$$y = K(u_0)\theta \text{ with } K \in \mathrm{R}^{N x n_\theta}. \tag{1-24}$$

Note that K can be a nonlinear function of the input. The explicit solution of the linear (weighted) least squares problem becomes

$$\hat{\theta}_{\mathrm{WLS}} = (K^T W K)^{-1} K^T W y \text{ and } \hat{\theta}_{\mathrm{LS}} = (K^T K)^{-1} K^T y. \tag{1-25}$$

Solutions that are numerically stable for expression (1-25) exclude the explicit calculation of the product $K^T W^{-1} K$ or $K^T K$, thus improving the numerical conditioning. This can be done with the MATLAB® solution given by

$$\begin{aligned} \hat{\theta}_{\mathrm{WLS}} &= (W^{1/2}K)\backslash(W^{1/2}y) \text{ with } W = W^{1/2}W^{1/2} \\ \hat{\theta}_{\mathrm{LS}} &= K\backslash y. \end{aligned} \tag{1-26}$$

Exercise 6 (Least squares estimation of models that are linear in the parameters) Consider the model `y0 = tan(u0*0.9*pi/2)`, evaluated for the inputs `u0 = linspace(0,1,N)`. Use the model

$$\hat{y}(t) = \sum_{i=0}^{n} \theta_i u_0^i(t) \tag{1-27}$$

to describe these data. Note that this is a nonlinear model that is linear-in-the-parameters θ_i.

- Generate a data set $y = y_0$. Put $N = 100$, and vary $n = 1$ to 20.
- Calculate the least squares solution ($W = I^{NxN}$) for the different values of n, using the stable MATLAB® solution (1-26) and the direct implementation (1-25).
- Compare the solutions, and calculate the condition number of K and $K^T K$. This can be done with MATLAB® instruction `cond( )`
- Compare the modeled output with the exact output and calculate the rms value of the error.

□

Observations (see Figure 1-7) From this figure, it can be seen that the condition number of the numerical unstable method (1-25) grows two times faster on a logarithmic scale than that of the stable method (1-26). The number of digits required to make the equations is given by the exponent of the condition number. From order 10 or larger, more than 15 digits are needed which is more than the calculation precision of MATLAB®. As a result, the obtained models are no longer reliable, even if there was no disturbing noise in the experiment. This shows that during system identification procedures, it is always necessary to verify the numerical conditions of the calculations. The condition number of the stable numerical im-

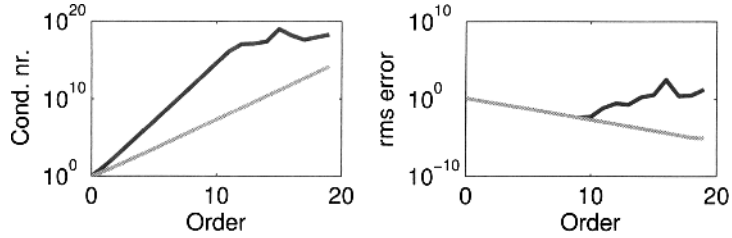

Figure 1-7 Identification of a polynomial model that is linear-in-the-parameters using a method that is numerical stable $\hat{\theta}_{WLS} = K\backslash y$ (gray lines) or numerical unstable $\hat{\theta}_{WLS} = (K^T K)^{-1} K y$ (black lines). Left: Condition number as a function of the model order. Right: The rms error as a function of the model order.

plementation grows slower, making it possible to solve higher order polynomial approximations.

Remark: If very high order polynomial approximations are needed, other more robust polynomial representations can be applied using orthogonal polynomials. The nature of these polynomials will depend upon the applied input signal.

1.2.5 Interpretation of the covariance matrix & Impact on experiment design

In this section, a one- and a two-parameter model will be considered. It is shown that: (1) The variance of a set of parameters is not enough to make conclusions on the model uncertainty; the full covariance matrix is needed. (2) The covariance matrix (and the correlation between the parameters) is strongly influenced by the design of the experiment.

Exercise 7 (Characterizing a 2-dimensional parameter estimate) Generate a set of measurements:

$$y(t) = a u_0(t) + n_t. \tag{1-28}$$

In the first experiment, $u_0(t)$ is generated by `linspace(-3,3,N)`, distributing N points equally between -3 and 3. In the second experiment, $u_0(t)$ is generated by `linspace(2,5,N)`.

- Choose $a = 0.1$, $N = 1000$, and $n_t \sim N(0, \sigma_n^2)$ with $\sigma_n^2 = 1$.
- Use as a model $y = a u_0 + b$, and estimate the parameters (a, b) using the method of Exercise 6.
- Repeat this experiment 10^5 times.
- Estimate the LS-parameters for both experiments, calculate the covariance matrix, and plot $\hat{a}(i)$ as a function of $\hat{b}(i)$.
- Plot also the estimated lines for the first 50 experiments.

Observations (Figure 1-8) In Figure 1-8 top, the parameters are plotted against each other. For the second experiment ($u \sim$ uniform in [2,5]), the parameters are strongly correlated, as can be seen from the linear relation between the estimated values $\hat{a}(i)$ and $\hat{b}(i)$. This is not so for the first experiment ($u = [-3, 3]$), the black cloud has its main axis parallel to the horizontal and vertical axis which is the typical behavior of an uncorrelated variable. This can also be seen in the covariance matrices:

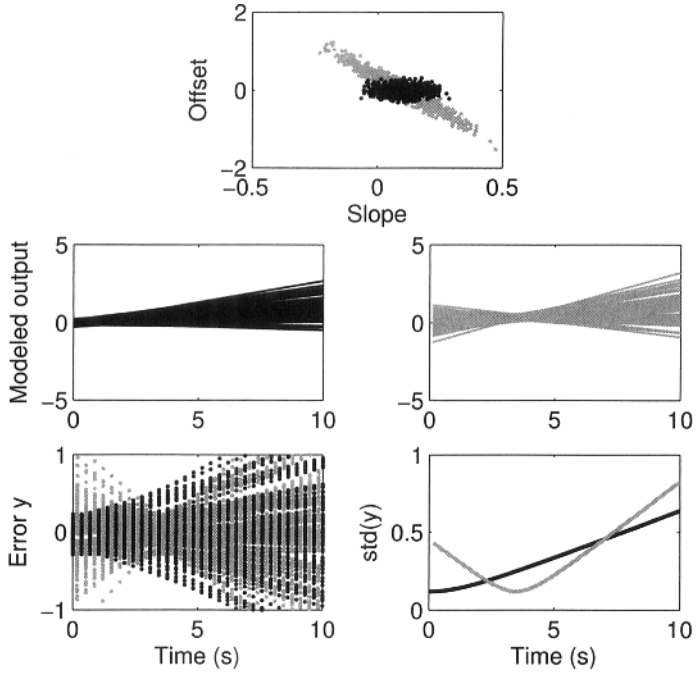

Figure 1-8 Black: Experiment in time interval $[-3, 3]$. Gray: Experiment in time interval $[2, 5]$. Top: Scatter plot (slope, offset). Middle: Modeled output. Bottom: Error on modeled output (left) and its standard deviation (right).

$$C_{exp1} = \begin{bmatrix} 3.2\times10^{-3} & 0.85\times10^{-4} \\ 0.85\times10^{-4} & 10.5\times10^{-3} \end{bmatrix}, \text{ and } C_{exp2} = \begin{bmatrix} 1.31\times10^{-2} & -4.6\times10^{-2} \\ -4.6\times10^{-2} & 16.9\times10^{-2} \end{bmatrix}, \tag{1-29}$$

or even better from the correlation matrices

$$R_{exp1} = \begin{bmatrix} 1 & 0.02 \\ 0.02 & 1 \end{bmatrix}, \text{ and } R_{exp2} = \begin{bmatrix} 1 & -0.97 \\ -0.97 & 1 \end{bmatrix}. \tag{1-30}$$

The correlation in the first matrix is almost zero, while for the second experiment it is almost one, indicating that a strong linear relation between the offset and slope estimate exists. This means that both variables vary considerably (a large standard deviation), but they vary together (a large correlation) so that the effect on the modelled output is small in the input range that the experiment was made (see Figure 1-8, middle and bottom). In that range, the variations of $\hat{a}$ are mostly canceled by those of $\hat{b}$. Outside this range, the standard deviation of the modeled output will be larger compared to that obtained with the first experiment because there the offset-slope compensation is no longer valid. This shows that the covariances play an important role in the model uncertainty.

1.2.6 What have your learned In Section 1.2? Further reading

In this section we studied the properties of linear (weighted) least squares estimators. Because these models are linear-in-the-parameters, it is possible to calculate the weighted least squares solution explicitly.

It is important to use numerical stable algorithms to calculate this explicit solution because otherwise the numerical noise can jeopardize the theoretical properties. The bias and covariance matrix of the estimates can be explicitly calculated provided that the covariance matrix of the disturbing noise is known.

The parameter uncertainty is directly influenced by the choice of the input signals. Experiment design methods provide systematic tools to get the best results within user specified constraints, for example for a given input power the determinant of the covariance matrix of the parameter estimates should be minimized.

The selection of the weighting matrix in the weighted least squares method influences the covariance matrix of the parameters. The smallest covariance matrix is obtained by choosing the weighting matrix as the inverse of the disturbing noise covariance matrix.

The parameter estimates are asymptotically Gaussian distributed under very weak conditions of the disturbing noise.

The books of Sorenson (1980) and van den Bos (2007) provide a general introduction to system identification, spending a lot of attention to weighted least squares estimation. Also the first chapter of the book of Pintelon and Schoukens (2001) introduces the reader to the general ideas of identification theory. More information on the numerical issues can be found in Golub and Van Loan (1996). Experiment design is covered in the books of Federov (1972), Goodwin and Payne (1977), and Zarrop (1979). Recently a new interest in this topic emerged, for example in the work of Hjalmarsson (2009), Gevers *et al.* (2009), and Bombois *et al.* (2006). The new design methods aim for an integrated design that optimizes the model for its final purpose, like, for example, the design of a controller.

1.3 MAXIMUM LIKELIHOOD ESTIMATION FOR GAUSSIAN AND LAPLACE DISTRIBUTED NOISE

In Sections 1.2 and 1.3, Gaussian distributed noise was added as disturbances to the measurements. It is shown in theory that least squares estimators, where the cost function is a quadratic function of the errors, perform optimal under these conditions. The smallest uncertainty on the estimators is found if a proper weighting is selected. This picture changes completely if the disturbing noise does not have a Gaussian distribution. In the identification theory it is shown that for each noise distribution, there corresponds an optimal choice of the cost function. A systematic approach to find these estimators is through the maximum likelihood theory, which is not within scope of this book, but some of its results will be illustrated on the resistance example. The disturbances will be selected once to have a normal distribution, and once to have a Laplace distribution. The optimal cost functions corresponding to these distributions are a least squares and a least absolute value cost function.

Exercise 8 (Dependence of the optimal cost function on the distribution of the disturbing noise) Consider a set of repeated measurements:

$$u_0(t) = R_0 i_0(t), \quad t = 1, 2, \ldots, N \tag{1-31}$$

with u_0, i_0 the exact values of the voltage and the current. Two different voltmeters are used, resulting in two data sets, the first one is disturbed by Gaussian (normal) distributed noise, the second one is disturbed with Laplace noise.

Generate an experiment with $N = 100$ measurements, with i_0 uniformly distributed in $[0, i_{max} = 0.01\ \text{A}]$, and $R_0 = 1000\ \Omega$. The current is measured without errors. The voltage measured with the first voltmeter is disturbed by independent, zero mean, normally distributed noise $n_u \sim N(0, \sigma_u^2 = 1)$, the second voltmeter is disturbed by Laplace distributed noise with zero mean, and $\sigma_u^2 = 1$.

$$\begin{aligned} i(t) &= i_0(t) \\ u(t) &= u_0(t) + n_u(t) \end{aligned}, \quad t = 1, 2, \ldots, N. \tag{1-32}$$

For the Gaussian noise the maximum likelihood solution reduces to a least squares (LS) estimate as in (1-6); for the Laplace distribution the maximum likelihood estimator is found as the minimizer of

$$V_{\text{LAV}}(R) = \frac{1}{N}\sum_{t=1}^{N} |u(t) - Ri(t)| \text{ and } \hat{R}_{\text{LAV}} = \arg\min_R V_{\text{LAV}}(R), \tag{1-33}$$

called the least absolute values (LAV) estimate.

- Repeat this exercise 10,000 times for $N = 100$.
- Apply both estimators also to the other data set.
- Calculate the mean value, the standard deviation, and plot for each case the histogram.

Help 1: Laplace distributed noise with zero mean and standard deviation *stdu* can be generated from uniformly distributed noise [0, 1] using the following MATLAB® implementation:

```
x = rand(NData,1);          % generate uniform distributed noise
nLap = zeros(size(x));      % vector used to store Laplace noise
nLap(x<=0.5) = log(2*x(x<=0.5))/sqrt(2)*stdU;
nLap(x>0.5) = - log(2*(1-x(x>0.5)))/sqrt(2)*stdU;
```

Help 2: to minimize $V_{\text{LAV}}(R)$, a simple scan can be made over R belonging to [800:0.1:1200]

Observations (see Figure 1-9) From Figure 1-9, it is seen that the estimates are scattered around the exact value. For the Gaussian case, the LS squares estimate is less scattered

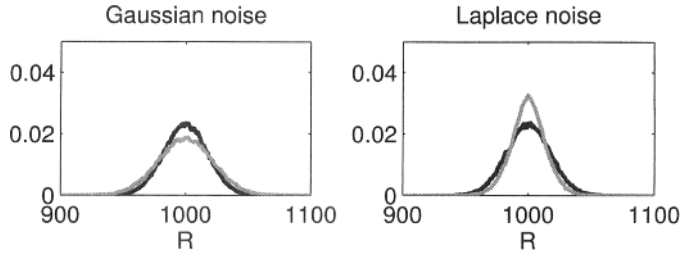

Figure 1-9 PDF of the Gaussian $\hat{R}_{\text{LS}}$ and Laplace $\hat{R}_{\text{LAV}}$ Maximum Likelihood estimators, applied to a Gaussian (left) and Laplace (right) noise disturbance. Black line: $\hat{R}_{\text{LS}}$; gray line: $\hat{R}_{\text{LAV}}$.

than the LAV estimate. For the Laplace case the situation is reversed. The estimated mean values $\hat{\mu}$ and standard deviations $\hat{\sigma}$ are given in Table 1-1. This shows that the maximum

TABLE 1-1 Mean and standard deviation of the Gaussian and Laplace maximum likelihood estimators, applied to a Gaussian and Laplace noise disturbance

	$\hat{\mu}_{LS}$	$\hat{\sigma}_{LS}$	$\hat{\mu}_{LAV}$	$\hat{\sigma}_{LAV}$
Gaussian noise	1000.040	17.5	999.94	22.0
Laplace noise	1000.002	17.3	999.97	13.7

likelihood estimator is optimal for the distribution that it is designed for. If the noise distribution is not known a priori, but the user can guarantee that the variance of the noise is finite, then it can be shown that the least squares estimate is optimal in the sense that it minimizes the worst possible situation among all noise distributions with a finite variance.

Further reading: The books of Sorenson (1980) and van den Bos (2007) give an introduction to maximum likelihood estimation, including illustrations on non-Gaussian distributions. The properties of the maximum likelihood estimator (consistency, efficiency, asymptotic normal distribution) are studied in detail.

1.4 IDENTIFICATION FOR SKEW DISTRIBUTIONS WITH OUTLIERS

In Section 1.3, it was shown that the optimal choice of the cost function depends on the distribution of the disturbing noise. The maximum likelihood theory offers a theoretical framework for the generation of the optimal cost function. In practice a simple rule of thumb can help to select a good cost function. Verify if the disturbing noise has large outliers: large errors appear to be more likely than expected from a Gaussian noise distribution.

In Exercise 9, the LS and the LAV estimates are applied to a χ^2 distribution with 1 degree of freedom: this is nothing other than a squared Gaussian distributed variable. Compared to the corresponding Gaussian distribution, the extremely large values appear too frequently (due to the squared value). Neither of both estimates (LS, LAV) is the MLE for this situation. But from the rule of thumb we expect that the LAV will perform better than the LS estimator. It will turn out that a necessary condition to get good results is to apply a proper calibration procedure for each method, otherwise a bias will become unavoidable.

Exercise 9 (Identification in the presence of outliers) Consider a set of repeated measurements:

$$u_0(t) = R_0 i_0(t), \quad t = 1, 2, \ldots, N \tag{1-34}$$

with u_0, i_0 the exact values of the voltage and the current. The voltage measurement is disturbed by noise, generated from a χ^2-distribution with 1 degree of freedom (= squared Gaussian noise).

Generate an experiment with N measurements, i_0 uniformly distributed in $[0, i_{max} = 0.01\ \text{A}]$ (use the MATLAB® routine `rand`), $R_0 = 1000\ \Omega$. The current is measured without errors. The measured voltage $u(t)$ is disturbed by χ^2 distribution distributed noise n_u with

$n_u = n^2$, with n generated as $N(0, \sigma_u^2 = 1)$.

Note that the mean value of $(\mathrm{E}\{n_u\})$ is 1, and median(n_u) is 0.455.

$$\begin{aligned} i(t) &= i_0(t) \\ u(t) &= u_0(t) + n_u(t) \end{aligned}, \quad t = 1, 2, \ldots, N \tag{1-35}$$

- In order to reduce the systematic errors, calibrate the data first. To do so, the mean value or the median of the noise should be extracted from the measurements. First, perform a measurement with zero current, so that $u(t) = n_u(t)$. The calibration is done using this record.
- Repeat the exercise 10,000 times and estimate, each time, the LS and the LAV estimate for both data sets.
- Estimate the pdf of the estimates, and calculate their mean value and standard deviation.

Observations (see Figure 1-10) From the figure it is seen that the estimates are no

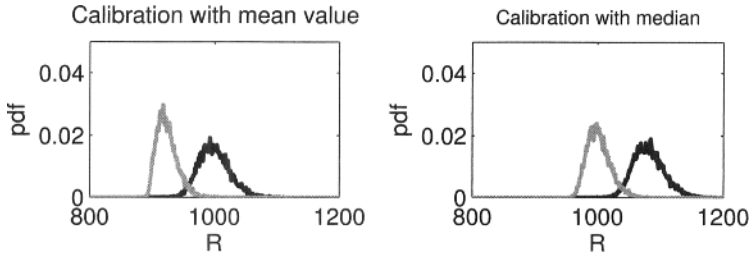

Figure 1-10 PDF of the Gaussian $\hat{R}_{LS}$ and Laplace $\hat{R}_{LAV}$ applied to χ^2 disturbed data. Left: Calibration with the mean value. Right: Calibration with the median value. Black line: $\hat{R}_{LS}$. Gray line: $\hat{R}_{LAV}$.

longer scattered around the exact value $R = 1000\ \Omega$. Only the combinations (LS estimate, mean value calibration) and the (LAV estimate, median value calibration) work well. The other combinations show a significant bias.

The mean and standard deviations are given in Table 1-2. Observe that the standard deviation of the LAV estimate is smaller than that of the LS estimate. LAV estimates are less sensitive to outliers! Note that the mean value of the LAV estimator, combined with the median calibration has still a small systematic error of 1.85, which is larger than the uncertainty on the mean value: 18.62/sqrt(10,000) = 0.18. If instead of using the mean, the median value is selected to average the 10,000 estimates, the bias disappears completely.

TABLE 1-2 Mean and standard deviation of the Gaussian and Laplace maximum likelihood estimators, applied to a Gaussian and Laplace noise disturbance

	$\hat{\mu}_{LS}$	$\hat{\sigma}_{LS}$	$\hat{\mu}_{LAV}$	$\hat{\sigma}_{LAV}$
Calibr.: mean value	999.84	24.30	924.29	16.26
Calibr.: median	1081.86	24.43	1001.85	18.62

Conclusion: The LS estimate should be combined with a calibration based on the mean, and the mean should be used to average the results. It is sensitive to outliers.
The LAV estimate should be combined with a calibration based on the median, the median should be used to average the results, and it is less sensitive to the presence of outliers.

1.5 SELECTION OF THE MODEL COMPLEXITY

1.5.1 Influence of the number of parameters on the uncertainty of the estimates

In this exercise it will be shown that, once the model includes all important contributions, the uncertainty grows if the number of model parameters is still increased.

Exercise 10 (Influence of the number of parameters on the model uncertainty)
In order to measure the flow of tap water, the height $y_0(t) = a_0 t$ of the water level in a measuring jug is recorded as a function of time t. However, the starting point of the measurements is uncertain. Hence two models are compared:

$$y(t) = at \quad \text{and} \quad y(t) = at + b. \tag{1-36}$$

The first model estimates only the flow, assuming that the experiment started at time zero, while the second one also estimates the start of the experiment.

Generate a set of measurements:

$$y(t) = a_0 t + n(t), \quad \text{with} \quad t = [0{:}N]/N. \tag{1-37}$$

- Choose $a_0 = 1$, $N = 1000$, and $n_t \sim N(0, \sigma_n^2)$ with $\sigma_n^2 = 1$.
- Repeat this experiment 10^5 times.
- Estimate the LS parameters of both models, and compare $\hat{a}$ for the one parameter model $y(t) = at$ and two-parameter model $y(t) = at + b$, by estimating the pdf of $\hat{a}$.
- Calculate the mean value and the standard deviation of the slope.
- Plot also the error $y_0(t) - \hat{y}(t)$ for the first 50 experiments, for $t \in [0, 2]$, with $\hat{y}(t)$ the modeled output.

□

TABLE 1-3 Mean and standard deviation of $\hat{a}$ in the one- and two-parameter model

	One-Parameter Model	Two-Parameter Model
mean	0.996	0.987
std. dev.	0.057	0.113

Observations The results are shown below in Table 1-3 and Figure 1-11. From the table it is seen that the uncertainty of the one-parameter estimate is significantly smaller than that of the two-parameter model. The mean values of both estimates are unbiased; the error equals the exact value within the uncertainty after averaging 100 experiments. Also in Figure 1-11

the same observations can be made. Note that due to the prior knowledge of the one-parameter model (at time zero the height is zero), a significantly smaller uncertainty on $\hat{a}$ is found for small values of t, and the one-parameter model is less scattered than the two-parameter model. If prior knowledge is incorrect, systematic errors would be made on the flow estimate; if it is correct, better estimates are found. An analysis of the residuals can guide the user to find out which of both cases s/he is faced with.

1.5.2 Model selection

The goal of this section is to show how to select an optimal model for a given data set. Too simple a model will fail to capture all important aspects of the output, and this will result in errors that are too large in most cases. Too complex models use too many parameters. As was illustrated in the previous section, such models also result in a poor behavior of the modeled output because the model becomes too sensitive to the noise. Hence, we need a tool that helps us to select the optimal complexity that balances the model errors against the sensitivity to the noise disturbances. It is clear that this choice will depend on the quality of the data. All these aspects are illustrated in the next exercise where we propose the Akaike information criterion (AIC) as a tool for model selection.

Consider a single input single output linear dynamic system, excited with an input $u_0(t)$ and output $y_0(t) = g_0(t)*u(t)$. The system has an impulse response $g_0(t)$ that is infinitely long (infinite impulse response or IIR system). For a stable system, $g_0(t)$ decays exponentially to zero, so that the IIR system can be approximated by a system with a finite length impulse response $g(t)$, $t = 0, 1, \ldots, I$ (finite impulse response or FIR system). For $t > I$, the remaining contribution can be considered to be negligible. The choice of I will depend not only on $g(t)$, but also on the signal-to-noise-ratio (SNR) of the measurements as well as on the length of the available data record.

$$\hat{y}(t) = \hat{g}(t)*u_0(t) = \sum_{k=0}^{I} \hat{g}(k)u_0(t-k), \quad \text{with } u_0(k) = 0 \text{ for } k<0. \tag{1-38}$$

In (1-38) it is assumed that the system is initially at rest. If this is not the case, transient errors will appear, but these disappear in this model for $t > I$ (why?).

The model parameters θ are, in this case, the values of the impulse response. θ is estimated from the measured data $u_0(t)$, $y(t)$, $t = 0, 1, \ldots, N$, with $y(t)$ the output measurement that is disturbed with i.i.d. (independent and identically distributed) noise $v(t)$ with zero mean and variance σ_v^2:

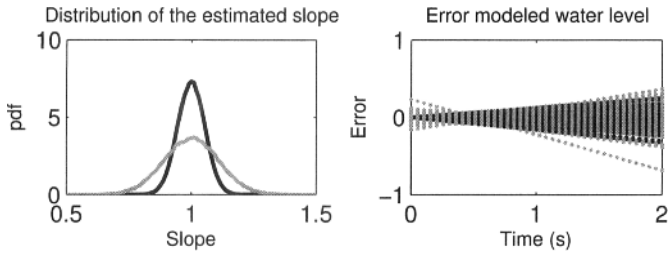

Figure 1-11 Impact of the number of parameters on the uncertainty of the slope estimate and the variability of the model. Black: One-parameter model $y = au$. Gray: Two-parameter model $y = au + b$. Left: The pdf of the estimated slope. Right: The error on the modeled output as a function of time.

$$y(t) = y_0(t) + v(t). \tag{1-39}$$

The estimates $\hat{\theta}$ are estimated by minimizing the least squares cost function:

$$V_{est}(\theta, Z^N) = \frac{1}{N}\sum_{t=0}^{N} |y(t) - \hat{y}(t)|^2, \text{ with } \hat{y}(t) = \hat{g}(t)*u_0(t) = u_0(t)*\hat{g}(t). \tag{1-40}$$

Note that this model is linear-in-the-parameters, and solution (1-26) can be used.

In order to find the "best" model, a balance is made between the model errors and the noise errors using a modified cost function that accounts for the complexity of the model. Here we propose to use of, amongst others, the AIC criterion:

$$V_{AIC} = V_M(\theta)\left(1 + 2\frac{\dim\theta}{N}\right). \tag{1-41}$$

Exercise 11 (Model selection using the AIC criterion) Consider the discrete time system $g_0(t)$ given by its transfer function:

$$G_0(z) = \sum_{k=0}^{n_b} b_k z^{-k} / \sum_{k=0}^{n_a} a_k z^{-k}, \tag{1-42}$$

Generate the filter coefficients a_k, b_k using the MATLAB® instruction

```
[b,a] = cheby1(3,0.5,[2*0.15  2*0.3])
```  (1-43)

This is a band pass system with a ripple of 0.5 dB in the pass band. Generate two data sets D_{est} and D_{val}, the former with length N_e being used to identify the model, the latter with length N_v to validate the estimated model. Note that the initial conditions for both sets are zero! Use the MATLAB® instruction

```
y0 = filter(b,a,u0),  y = y0+ny
```  (1-44)

with u_0 zero mean normally distributed noise with $\sigma_{u_0} = 1$, and $n_y \sim N(0, \sigma^2 = 0.5^2)$ for a first experiment, and $n_y \sim N(0, \sigma^2 = 0.05^2)$ for a second experiment. Put $N_{est} = 1000$, and $N_{val} = 10{,}000$ in both experiments.

- Use the linear least squares procedure (1-26) to estimate the model parameters of an approximating FIR model, for varying orders from 0 to 100.
- Calculate for each of the models the simulated output $\hat{y} = \text{filter}(\hat{g}, 1, u_0)$, and calculate the cost function (1-40) on D_{est} and on D_{val}.
- Calculate V_{AIC}.
- Calculate $V_0 = \frac{1}{N_{val}}\sum_{t=0}^{N_{val}} |y_0(t) - \hat{y}(t)|^2$ on the undisturbed output of the validation set.

- Plot V_{est}, V_{AIC}, V_{val} as a function of the model order. Normalize the value of the cost function by $\sigma_{n_y}^2$ to make an easier comparison of the behavior for different noise levels.

- Plot $\sqrt{V_0/\sigma_{n_y}^2}$ as a function of the model order.

□

Observations The results are shown in Figure 1-12, and the following observations can be made:

(i) Increasing the model order results in a monotonically decreasing cost function V_{est}. This result was to be expected because a simpler model is always included by the more complex model, and the linear LS always retrieves the absolute minimum of the cost function, so that no local minima of the cost function as a function of the model order exist. Hence, increasing the complexity of the model should reduce the value of the cost function (it is a monotonic not increasing function of the model complexity).

(ii) On the validation data we observe first a decrease and then an increase of V_{val}. In the beginning, the additional model complexity is mainly used to reduce the model errors, a steep descent of the cost function is observed. From a given order on, the reduction of the model errors is smaller than the increased noise sensitivity due to the larger number of parameters, resulting in a deterioration of the capability of the model to simulate the validation output. As a result the validation cost function V_{val} starts to increase.

(iii) V_{AIC} gives a good indication, starting from the estimation data only, where V_{val} will be minimum. This reduces the need for long validation records, and it allows us to use as much data as possible for the estimation step.

(iv) The optimal model order increases for a decreasing disturbing noise variance. Since the plant is an IIR system with an infinite long impulse response, it is clear that in the absence of disturbing noise $\sigma_n = 0$, the optimal order would become infinite. In practice this value is never reached due to the presence of calculation errors that act also as a disturbance.

(v) A fair idea about the quality of the models is given by V_0. The normalized rms value $\sqrt{V_0/\sigma_y^2}$ is plotted on the right side of Figure 1-12. This figure shows that a wrong selection of the model can result in much larger simulation errors. The good news is that the selection of the best model order is not so critical, the minimum is quite flat and all model orders in the

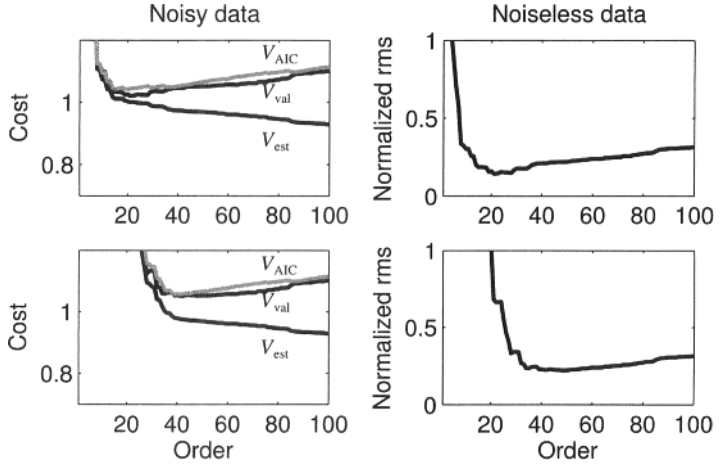

Figure 1-12 : Comparison of the normalized cost function V_{est}, the AIC criterion V_{AIC}, and the validation V_{val} for $\sigma_n = 0.5$ (top) and $\sigma_n = 0.05$ (bottom).

neighborhood of the minimum result in good estimates. Note that in real-life experiments, V_0 is not available.

Remark: In practice the validation set is chosen always (much) smaller than the test data. The data should be primaly used to estimate a good model. In this exercise we selected an extremely large data set to eliminate the noise variation in the validation to better visualize the quality of the AIC model selection. The AIC method makes it possible to select the model on the test data without using a validation set. In practice it is however still advisable to test the final model on a validation test to verify if the model can explain also fresh data that were not used to build it.

1.5.3 What have we learned In Section 1.5? Further reading

In this section we learned that the choice of the model complexity is an important issue. Too simple models lead to bias errors, while models with too many parameters suffer from an increased variability. Model selection tools balance the bias and variance errors. A first possibility is to verify the identified model on a data set that was not used during the parameter estimation step, this is called a validation set. Instead of saving a part of the available data for this test it is also possible to predict the behavior of an identified model on a new data set. We have learned that the Akaike information criterion AIC is able to predict the value of the cost function on the validation set. This allows the user to use all the data in the estimation step, leading to a smaller variance. In the literature a lot of results are published on this topic. Besides the original paper of Akaike (1974), we refer the reader to the classical textbooks on system identification, — for example, Ljung (1999), Söderström and Stoica (1989), and Johansson (1993) — to learn more about this topic.

1.6 NOISE ON INPUT AND OUTPUT MEASUREMENTS: THE IV METHOD AND THE EIV METHOD

In Section 1.2.2 it was shown that the presence of disturbing noise on the input measurements creates a systematic error. In this set of exercises, more advanced identification methods are illustrated that can deal with this situation. Two methods are studied: The first is called the instrumental variables method (IV), the second is the errors-in-variables (EIV) method. The major advantage of the IV method is its simplicity. No additional information is required from the user. The disadvantage is that this method does not always perform well. Both situations are illustrated in the exercises. The EIV performs well in many cases, but in general additional information from the user is required. The covariance matrix of the input–output noise should be known. All methods are illustrated again on the resistance example with measured current and voltage $i(t), u(t)$, $t = 1, 2, \ldots, N$. Both measurements are disturbed by mutually uncorrelated Gaussian noise, $n_i(t), n_u(t)$:

$$\begin{aligned} i(t) &= i_0(t) + n_i(t), \\ u(t) &= u_0(t) + n_u(t). \end{aligned} \tag{1-45}$$

The least squares estimate is given by

$$\hat{R}_{\mathrm{LS}} = \frac{\sum_{t=1}^{N} u(t)i(t)}{\sum_{t=1}^{N} i(t)^2}, \tag{1-46}$$

the instrumental variables estimator (IV) is

$$\hat{R}_{\mathrm{IV}} = \frac{\sum_{t=1}^{N} u(t)i(t+s)}{\sum_{t=1}^{N} i(t)i(t+s)}, \tag{1-47}$$

with s a user selectable shift parameter. Note that the IV estimator equals the LS estimator for $s = 0$.

The EIV estimator is given by

$$\hat{R}_{\mathrm{EIV}} = \frac{\frac{\sum u(t)^2}{\sigma_u^2} - \frac{\sum i(t)^2}{\sigma_i^2} + \sqrt{\left(\frac{\sum u(t)^2}{\sigma_u^2} - \frac{\sum i(t)^2}{\sigma_i^2}\right)^2 + 4\frac{(\sum u(t)i(t))^2}{\sigma_u^2\sigma_i^2}}}{2\frac{\sum u(t)i(t)}{\sigma_u^2}}, \tag{1-48}$$

with σ_u^2, σ_i^2 the variances of the voltage and current noise respectively, the covariance is assumed to be zero in this expression: $\sigma_{ui}^2 = 0$.

Exercise 12 (Noise on input and output: The instrumental variables method applied on the resistor estimate) Generate the current $i_0(k)$ from a Gaussian white noise source e_1 filtered by a first order Butterworth filter with cutoff frequency f_{Gen}

$$i_0 = \mathrm{filter}(b_{\mathrm{Gen}}, a_{Gen}, e_1), \tag{1-49}$$

with $[b_{\mathrm{Gen}}, a_{Gen}] = \texttt{butter}(1, 2*f_{\mathrm{Gen}})$. Next this filtered sequence is scaled to get a signal i_0 with standard deviation σ_{i_0}.

Generate the measured current and voltage (1-45), where $n_u(k)$ is white Gaussian noise: $N(0, \sigma_{n_u}^2)$. The current noise $n_i(k)$ is obtained from a Gaussian white noise source filtered by a second-order Butterworth filter with cut-off frequency f_{Noise}:

$$n_i = \texttt{filter}(b_{\mathrm{Noise}}, a_{\mathrm{Noise}}, e_2), \tag{1-50}$$

with $[b_{\mathrm{Noise}}, a_{\mathrm{Noise}}] = \texttt{butter}(2, 2*f_{\mathrm{Noise}})$, and e_2 white Gaussian noise. Its variance is scaled to $\sigma_{n_i}^2$.

- Experiment 1: Generate three sets of 1000 experiments with $N = 5000$ measurements each on a resistor $R_0 = 1000\ \Omega$, and the following parameter settings:

$$\begin{aligned} f_{\mathrm{Gen}} &= 0.05, \quad f_{\mathrm{Noise}} = [0.4995, 0.475, 0.3], \\ \sigma_{i_0} &= 0.1, \quad \sigma_{n_i} = 0.1, \quad \sigma_{n_u} = 1. \end{aligned} \tag{1-51}$$

- Process these measurements with the LS estimator, as well as with the IV-estimator with the shift parameter $s = 1$.
- Experiment 2: Generate 1000 experiments with $N = 5000$ measurements each, and the following parameter settings:

$$f_{\text{Gen}} = 0.05, \quad f_{\text{Noise}} = 0.3, \quad \sigma_{i_0} = 0.1, \quad \sigma_{n_i} = 0.1, \quad \sigma_{n_u} = 1. \tag{1-52}$$

- Process these measurements with the LS estimator, and with the IV estimator with the shift parameter $s = 1, 2, 5$.

Plot for both experiments:

- the pdf of $\hat{R}_{\text{LS}}$ and $\hat{R}_{\text{IV}}$,
- the autocorrelation function of i_0 and n_i (*hint*: use the MATLAB® instruction `xcorr`),
- the FRF of the generator and the noise filter.

□

Observations The results are shown below in Figure 1-13 and Figure 1-14. In Figure 1-13, the results are for a fixed generator filter and a varying noise filter. The shift parameter for the IV is kept constant at 1. From this figure it is clearly seen that the LS results are

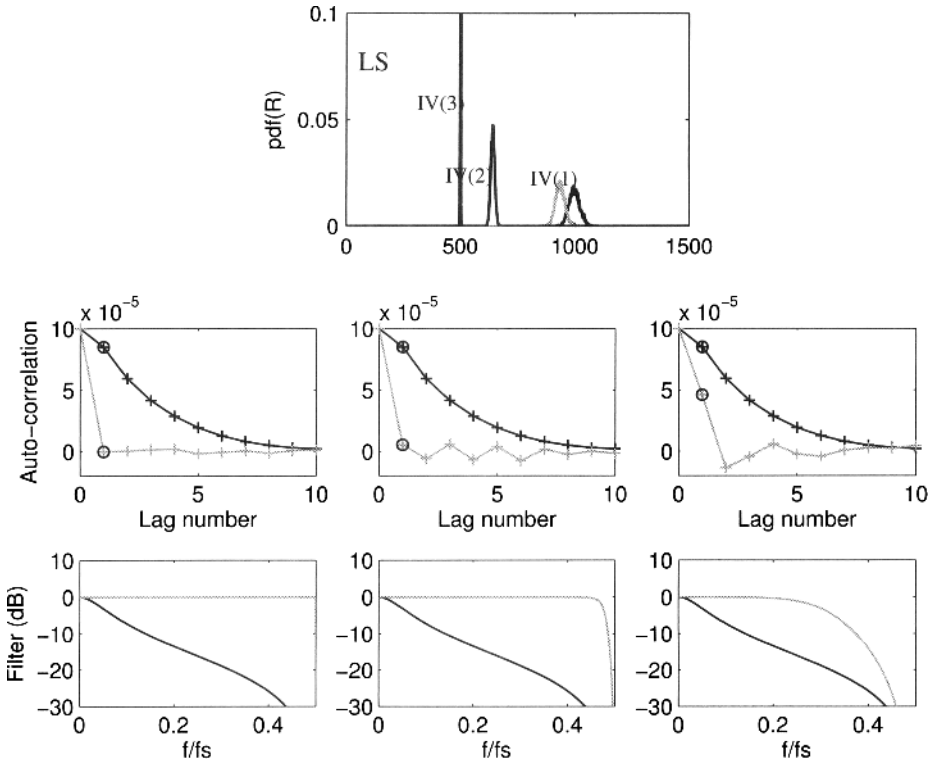

Figure 1-13 Study of the LS and IV estimate for a varying noise filter bandwidth and fixed shift $s = 1$. Top: The LS (black line) and IV estimate (black or gray line). IV(1), IV(2), and IV(3) correspond to the first, second, and third filter. Middle: The auto correlation of i_0 (black) and n_i (gray) for the different noise filters. Bottom: The filter characteristics of i_0 (black) and the noise n_i (gray).

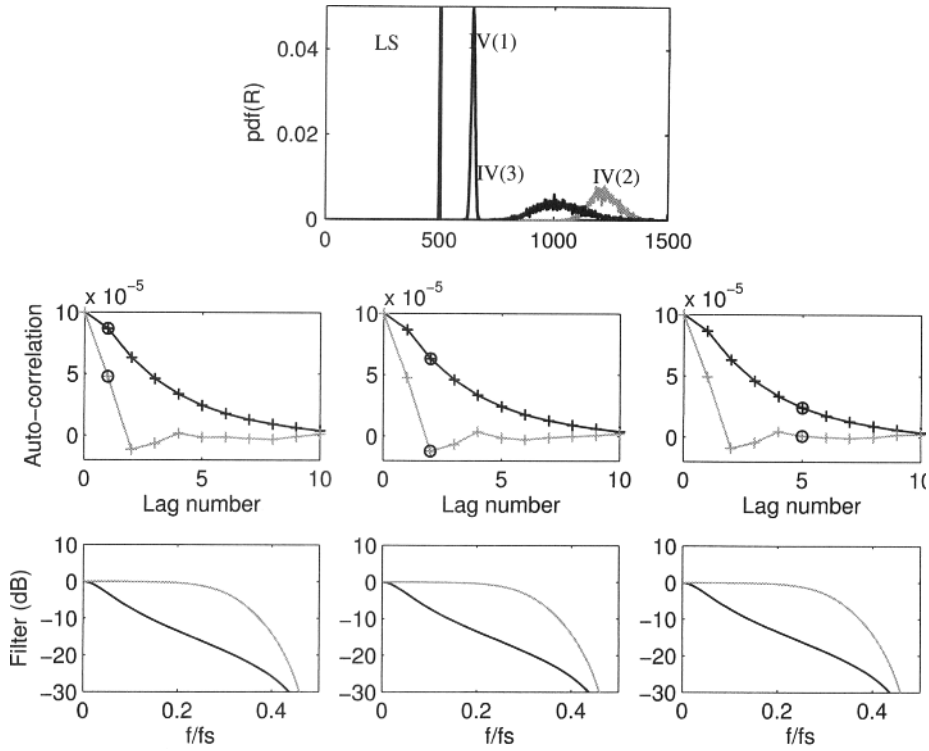

Figure 1-14 Study of the LS and IV estimate for a fixed noise filter bandwidth and a varying shift $s = 1, 2, 5$. Top: The LS (black) and IV (black and gray) estimate. IV(1), IV(2), and IV(3) correspond to a shift of 1,2, and 5 tabs. Middle: The auto correlation of i_0 (black) and n_i (gray). Bottom: The filter characteristics of i_0 (black) and the noise n_i (gray).

strongly biased. This is due to the noise on the input, the relative bias is in the order of $\sigma_{n_i}^2/\sigma_{i_0}^2$. For the IV results, the situation is more complicated. For the white noise situation, no bias is visible. However, once the output noise is filtered, a bias becomes visible. The relative bias is proportional to the ratio of the autocorrelation functions of the noise and the current $R_{n_i n_i}(s)/R_{i_0 i_0}(s)$.

The same observations can also be made in Figure 1-14. In this figure, the shift parameter is changed while the filters are kept constant. It can be seen that the bias becomes smaller with increasing shift s, because $R_{n_i n_i}(s)/R_{i_0 i_0}(s)$ is getting smaller. At the same time the dispersion is growing, mainly because $R_{i_0 i_0}(s)$ is getting smaller. Observe also that the sign of the bias depends on the sign of $R_{n_i n_i}(s)$. The IV method works well if the bandwidth of the generator signal is much smaller than that of the noise disturbances.

Exercise 13 (Noise on input and output: the errors-in-variables method) In this exercise the EIV method is used as an alternatives for IV method to reduce/eliminate the bias of the least squares estimate. This time no constraint is put on the power spectra (bandwidth) of the excitation and the disturbing noise, but instead the variance of the input and output disturbing noise should be given in advance (prior information). This is illustrated again on the resistance example with measured current and voltage $i(t)$, $u(t)$, $t = 1, 2, \ldots, N$.

The least squares estimate is given in (1-46), the EIV-estimator is given in (1-48), where the sum runs over $t = 1, \ldots, N$. It is shown to be the minimizer of the following cost function:

$$V_{\text{EIV}} = \frac{1}{N}\sum_{t=1}^{N}\left\{\frac{(u(t)-u_0(t))^2}{\sigma_{n_u}^2}+\frac{(i(t)-i_0(t))^2}{\sigma_{n_i}^2}\right\}, \tag{1-53}$$

with respect to u_0, i_0, R_0 under the constraint that $u_0(t) = R_0 i_0(t)$.

- Setup: Generate the current $i_0(t)$ from a white zero mean Gaussian noise source $N(0, \sigma_{i_0}^2)$.

  Generate the measured current and voltage as

$$\begin{aligned} i(t) &= i_0(t) + n_i(t), \\ u(t) &= u_0(t) + n_u(t); \end{aligned} \tag{1-54}$$

  $n_u(t)$ and $n_i(t)$ are white Gaussian noise sources with zero mean and variance $\sigma_{n_u}^2$ and $\sigma_{n_i}^2$, respectively.
- Generate a set of 1000 experiments with $N = 5000$ measurements each, and the following parameter settings:

$$R_0 = 1000, \quad \sigma_{i_0} = 0.01, \quad \sigma_{n_i} = 0.001, \quad \sigma_{n_u} = 1. \tag{1-55}$$

Calculate the LS and EIV estimate. Plot the histogram with $\hat{R}_{\text{LS}}$ and $\hat{R}_{\text{EIV}}$.

Observations The results are shown below in Figure 1-15, From this figure it is clearly

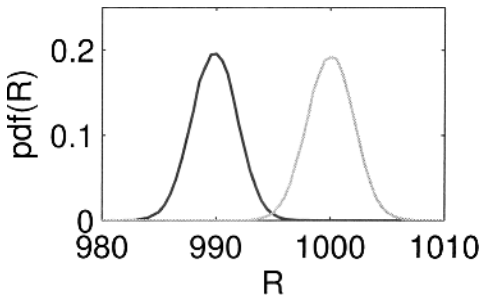

Figure 1-15 Comparison of the pdf of the LS (black) and the EIV estimate (gray), calculated with prior known variances.

seen that the LS estimates are strongly biased (mean value is 990.15). This is due to the noise on the input, the relative bias is in the order of $\sigma_{n_i}^2/\sigma_{i_0}^2$. No systematic error can be observed in the EIV results (mean value is 999.96). The IV estimate would fail completely in this situation (why?).

Discussion We learned that noise on the input (regressor) results in a bias on the estimates that depends on the inverse SNR of the input (regressor). Both methods, the IV and the EIV method, have pros and cons. While the IV method requires a lot of insight of the user (are the tight experimental conditions met?), the EIV method needs knowledge of the covariance matrix of the input and output noise. In the literature, alternative methods are proposed to deal with this problem. A first possibility is to use repeated experiments. From the variations from one experiment to the other, it is possible to estimate the noise covariance matrix from the data. This is later illustrated in this book; see, for example, Exercise 47 (Schoukens *et al.*, 1997). Alternative methods estimate at the same time a parametric plant and noise

model (Söderström, 2007). This allows to relax the experimental conditions (no repeated experiments are required), at a cost of more difficult optimization problem to be solved (estimate the plant- and noise model parameters).