FREQUENCY DOMAIN SYSTEM IDENTIFICATION

J. Schoukens and **R. Pintelon**
Department ELEC, Vrije Universiteit Brussel, Belgium

Keywords: system identification, frequency domain, time domain

Contents

Summary

In this chapter, a discussion of frequency domain identification methods is given. Here, we give first a general but very concise introduction to identification. Next we zoom in on the identification of dynamic systems, where most attention is spent to the user aspects of the identification problem. Eventually we deal with the choice between time domain and frequency domain identification methods, guiding the user to reasonable solutions for his problem. Some of the suggested methods will be explained in *Identification of Linear Systems in Time Domain* that deals with time domain identification methods.

In the article level contributions, we first deal with nonparametric frequency response measurements. The second contribution gives a more detailed discussion of parametric frequency domain identification methods. The third article is focused on modal analysis. Typical for these vibration problems is the very large number of outputs. The huge amounts of data require dedicated algorithms that balance between accuracy and memory/computing needs. In the last article, the relations between time domain and frequency domain identification is discussed.

1. Introduction

The goal of identification is to build a model of reality, starting from a finite set of observations. At the first glance this might seem to be a simple task, however, in practice it becomes involved because a perfect model does not exist, and the observations are disturbed by noise. Moreover, nature as a whole is too complex to be modeled; we concentrate our efforts on just one part of reality at a time, called the system, the rest of nature being referred to as the environment of the system.

Interactions between the system and its environment are described by input and output ports. So the task becomes to find a good model that describes the information that is present in the measurements of the signals at these input and output ports. In each identification task the same questions pop up: what experiments should be done? What models should be used? How to match the model to the data? How to check if all information from the experimental data is used? These are the basic questions that should be answered.

The complexity of the identification problem is illustrated by a very simple example. A group of students measures a resistor using the setup that is shown in Figure 1.They

passed a constant but unknown current through the resistor. The voltage u_0 across the resistor and the current i_0 through it were measured using a voltmeter and an ampere meter. The input impedance of the voltmeter is very large compared with the unknown resistor so that all the measured current is assumed to pass through the resistor. A set of voltage and current measurements, respectively, $u(k)$, $i(k)$ with $k = 1, 2, ..., N$ is made, resulting in a set of measured resistance values $R(k) = u(k)/i(k)$. The measurement results are shown in Figure 2.

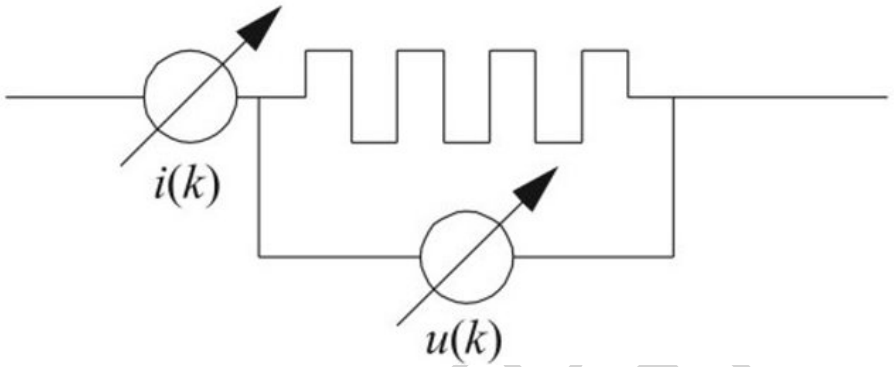

Figure 1. Measurement of the resistance of a resistor

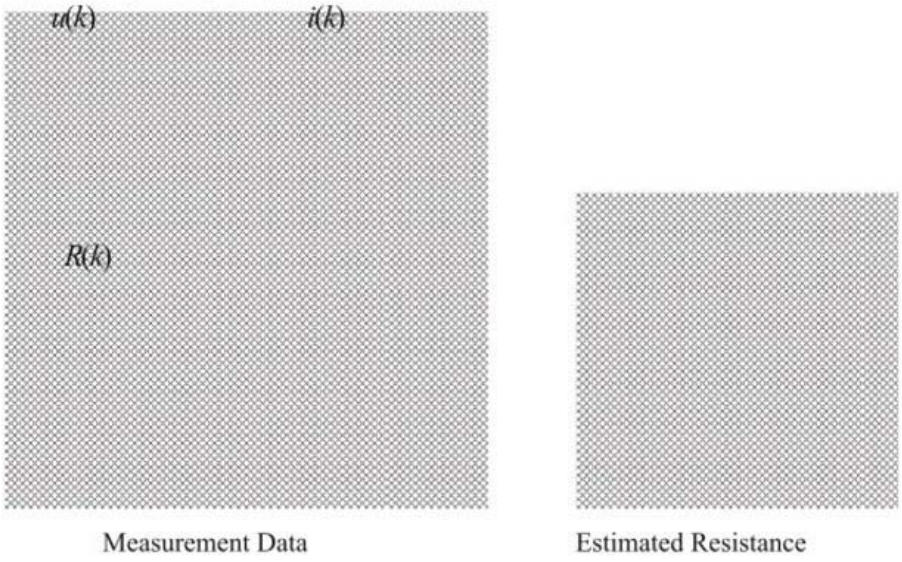

Figure 2. Left side: Measurement results $u(k), i(k)$ and $R(k) = u(k)/i(k)$. Right side: The estimated value of the resistor using 3 different estimators.

For the model, Ohm's law is used: $u = Ri$. There is no unique solution to match this model to the measurements. Here we compare 3 alternatives:

$$\hat{R}_{\mathrm{SA}}(N) = \frac{1}{N}\sum_{k=1}^{N}\frac{u(k)}{i(k)} \text{ minimizing } V_{\mathrm{SA}}(R,N) = \sum_{k=1}^{N}\left(R(k) - R\right)^2 \qquad (1)$$

$$\hat{R}_{\mathrm{LS}}(N) = \frac{\frac{1}{N}\sum_{k=1}^{N} u(k)i(k)}{\frac{1}{N}\sum_{k=1}^{N} i^2(k)} \text{ minimizing } V_{\mathrm{LS}}(R,N) = \sum_{k=1}^{N}\left(u(k) - Ri(k)\right)^2 \qquad (2)$$

$$\hat{R}_{\mathrm{EV}}(N) = \frac{\frac{1}{N}\sum_{k=1}^{N} u(k)}{\frac{1}{N}\sum_{k=1}^{N} i(k)} \text{ minimizing}$$

$$V_{\mathrm{EV}}\left(R, i_{\mathrm{p}}, u_{\mathrm{p}}, N\right) = \sum_{k=1}^{N}\left(u(k) - u_{\mathrm{p}}\right)^2 + \sum_{k=1}^{N}\left(i(k) - i_{\mathrm{p}}\right)^2, \text{ subject to } u_{\mathrm{p}} = Ri_{\mathrm{p}}. \qquad (3)$$

The notations SA, LS, EV will become clear later on in this chapter. The index N indicates that the estimate is based on N observations. Note that each estimator in this case can be interpreted as the minimizer of a cost function that is used as a measure of the goodness of the fit.

The choice of the cost function determines the estimator and its properties. The three estimators result in the same estimate on noiseless data (replace $u(k), y(k)$ by $i_0, u_0 = R_0 i_0$. However, on noisy data their behavior is different, as is illustrated by a simulation. The raw measurements and the estimated resistance values are given in Figure 2. From this figure a number of interesting observations can be made:

- All estimators have large variations for small values of N, and converge to an asymptotic value for large values of N. This corresponds to the intuitively expected behavior: if a large number of data points are processed we should be able to eliminate the noise influence due to the averaging effect.
- The asymptotic values of the estimators depend on the kind of averaging technique that is used. This shows that there is a serious problem: at least 2 out of the 3 methods converge to a wrong value. It is not even certain that any one of the estimators is doing well. This is quite disturbing and disappointing: even an infinite amount of measurements does not guarantee that the exact value is found.

These observations prove very clearly that a good theory is needed to explain and understand the behavior of candidate estimators. This will allow us to make a sound selection out of many possibilities and to indicate in advance, before running expensive experiments, if the selected method is prone to serious shortcomings.

2. A Brief Introduction to Identification

In this section we consider the general identification problem. First we discuss the basic steps that can be recognized in every identification procedure. Next we develop some tools to describe the stochastic behavior of estimates. This allows us to understand

better what can be expected from a 'good' estimator. Finally we present a statistical approach to the identification problem.

2.2 Basic Steps in the Identification Process

Each identification session consists of a sequence of basic steps:

- Collect information about the system;
- Select a model structure to represent the system;
- Choose the model parameters to fit the model as closely as possible to the measurements: selection of a "goodness of fit" criterion;
- Validate the selected model.

Each of these points is discussed in more detail in the following.

2.1.1 Collect Information about the System

If we want to build a model for a system we should get information about it. This can be done by just watching the natural fluctuations (e.g. vibration analysis of a bridge that is excited by normal traffic), but most often it is more efficient to set up dedicated experiments that actively excite the system (e.g. controlled excitation of a mechanical structure using a shaker). In the latter case the user has to select an excitation that optimizes his/her own effort (e.g. minimum cost, minimum time or minimum power consumption for a given measurement accuracy) within the operational constraints (e.g. the excitation should remain below a maximum allowable level). The quality of the final result can heavily depend on the choices that are made.

2.1.2 Select a Model Structure to represent the System

A specific model should be selected to represent the system. Again a wide variety of possibilities exist such as

• *parametric versus non-parametric models*

In a parametric model, the system is described using a limited number of characteristic quantities called the parameters of the model, while in a non-parametric model the system is characterized by measurements of a system function at a large number of points.

Examples of parametric models are the transfer function of a filter described by its poles and zeros, the equations of motion of the piston in a reciprocating engine , etc. An example of a non-parametric model is the description of a filter by its impulse response at a large number of points.

Usually it is simpler to create a non-parametric model than a parametric one because the modeler needs less knowledge about the system itself in the former case. However, physical insight and concentration of information are more significant for parametric models than for non-parametric ones.

• *white box models versus black box models*

In the construction of a model, physical laws, whose availability and applicability depend on the insight and skills of the experimenter, can be used (Kirchhoff's laws, Newton's laws, etc.). Specialized knowledge relating to different scientific fields may be brought into this phase of the identification process. The modeling of a loudspeaker, for example, requires extensive understanding of mechanical, electrical and acoustical phenomena. The result may be a physical model, based on comprehensive knowledge of the internal functioning of the system. Such a model is called a *white box model*.

Another approach is to extract a *black box model* from the data. Instead of developing a model based upon physical insight and knowledge, a mathematical model is proposed which allows sufficient description of any observed input and output measurements. This reduces the modeling effort significantly.

For example, instead of modeling the loudspeaker using physical laws, an input-output relation, taking the form of a high-order transfer function, could be proposed. The choice between the different methods depends on the aim of the study: the white box approach is better for gaining insight into the working principles of a system, but a black box model may be sufficient if the model will only be used for prediction of the output.

• *linear models versus non-linear models*

In real life almost every system is non-linear, but often linear approximation are made to simplify the problems. The validity of linearization depends strongly on the intended use of the model. For example, a non-linear model is needed to describe the distortion of an amplifier, but a linear model will be sufficient to represent its transfer characteristics if the linear behavior is dominant and is the only interest.

• *linear-in-the-parameters versus non-linear-in-the-parameters*

A model is called linear-in-the-parameters if there exists a linear relation between these parameters and the error that is minimized. This does not imply that the system itself is linear. For example $\varepsilon = y - \left(a_1 u + a_2 u^2\right)$ is linear in the parameters a_1 and a_2 but describes a non-linear system. On the other hand

$$\varepsilon(j\omega) = Y(j\omega) - \frac{a_0 + a_1 j\omega}{b_0 + b_1 j\omega} U(j\omega) \qquad (4)$$

describes a linear system but the model is non-linear in b_1 and b_2. Linearity in the parameters is a very important aspect of models since it has a strong impact on the complexity of the estimators if a (weighted) least squares cost function is used. In that case the problem can be solved analytically for models that are linear in the parameters so that an iterative optimization problem is avoided. This is illustrated in Section 2.3.

2.1.3 Match the selected Model to the Measurements

The selected model (e.g. a parametric transfer function model of given order) should be matched as closely as possible with the available information about the system. Mostly, this is done by minimizing a criterion that measures a goodness of the fit, like the cost functions in (1) to (3). The choice of this criterion is extremely important since it determines the stochastic properties of the final estimator. As seen from the resistance example, many choices are possible and each of them can lead to a different estimator with its own properties. Usually, the cost function defines a distance between the experimental data and the model. The cost function can be chosen on an *ad hoc* basis using intuition and insight, but there exists also a more systematic approach based on stochastic arguments as will be explained in Section 2.3.

2.1.4 Validate the selected Model

Finally, the validity of the selected model should be tested: does this model describe the available data properly or are there still indications that some of the data are not well modeled, indicating residual model errors? In practice the best model (= the smallest errors) is not always preferred. Often a simpler model that describes the system within user-specified error bounds is preferred. Tools exist that guide the user through this process by separating the residual errors into different classes, for example unmodeled linear dynamics and non-linear distortions. From this information further improvements of the model can be proposed, if necessary.

During the validation tests it is always important to keep the application in mind. The model should be tested under the same conditions as it will be used later. Extrapolation should be avoided as much as possible. The application also determines what properties are critical.

2.1.5 Conclusion

This brief overview of the identification process shows that it is a complex task with a number of interacting issues. It is important to pay attention to all aspects of this procedure, from experiment design to model validation, in order to get the best results. The reader should be aware of the fact that besides this list of actions other aspects are also important. A short inspection of the measurement setup can reveal important shortcomings that can jeopardize a lot of information. Good understanding of the intended applications helps to setup good experiments and is very important to make proper simplifications during the model building process. Many times, choices are made that are not based on complicated theories but are dictated by the practical circumstances.

In these cases a good theoretical understanding of the applied methods will help the user to be aware of the sensitive aspects of his techniques. This will enable him/her to put all his/her effort on the most critical decisions. Moreover, he/she will become aware of the weak points of the final model.

2.2 Description of the Stochastic Behavior of Estimators: What can be expected

from a good Estimator?

Since the estimates are obtained as a function of a finite number of noisy measurements, they are stochastic variables as well. Their probability density functions (pdf) are needed in order to characterize them completely. However, in practice it is usually very hard to derive them, and the behavior of the estimates is described only by a few numbers, such as their mean value (as a description of the location) and the covariance matrix (to describe the dispersion). Both aspects are discussed below.

2.2.1 Location Properties: Unbiased and Consistent Estimates

For simplicity, we choose the mean value over other possibilities like the median. It seems very natural to require that the mean equals the true value, but it turns out to be impractical. What are the true parameters of a system? We can only speak about true parameters if an exact model exists. This is a purely imaginary situation; in practice we always have model errors. For theoretical reasons it still makes sense to consider the concept of "true parameters", but it is clear at this point that we have to generalize to more realistic situations. One possible generalization is to consider the estimator evaluated in the noiseless situation as the "best" approximation. These parameters are then used as reference values to compare the results obtained from noisy measurements. The goal is then to remove the influence of the disturbing noise so that the estimator converges to this reference value.

Definition 2.1 (unbiasedness): An estimator $\hat{\theta}$ of the parameters θ_0 is unbiased if $E\left\{\hat{\theta}\right\} = \theta_0$, for all true parameters θ_0. Otherwise it is a biased estimator.

If the expected value only equals the true value for an infinite number of measurements, then the estimator is called asymptotically unbiased. In practice it turns out that (asymptotic) unbiasedness is a hard requirement to deal with. Often, it is very difficult or even impossible to find the expected value analytically; sometimes it does not even exist. Consequently, a more conveneint tool (e.g. consistency) is needed.

Definition 2.2. (consistency): An estimator $\hat{\theta}(N)$ of the parameters θ_0 is weakly consistent, if it converges in probability to θ_0 : $\operatorname{plim}_{N\to\infty} \hat{\theta}(N) = \theta_0$, and strongly consistent if it converges with probability one (almost surely) to θ_0 : $\text{a.s.}\lim_{N\to\infty} \hat{\theta}(N) = \theta_0$.

The precise explanation of these probability limits is beyond the scope of this chapter. Loosely speaking, it means that the pdf of $\hat{\theta}(N)$ contracts around the true value θ_0, or $\lim_{N\to\infty} \text{Prob}\left(\left|\hat{\theta}(N) - \theta_0\right| > \delta > 0\right) = 0$. The major advantage of the consistency concept is purely mathematical: it is much easier to prove consistency than unbiasedness using probabilistic theories starting from the cost function interpretation.

2.2.2 Dispersion Properties: Efficient Estimators

We use the covariance matrix to describe the dispersion of an estimator, i.e. to ascertain how much the actual estimator is scattered around its limiting value? Again this choice, among other possibilities (like, for example, percentiles), is highly motivated from a mathematical point of view: within the stochastic framework used it will be quite easy to calculate the covariance matrix. As users we are highly interested in estimators with minimal errors. However, since we can collect only a finite number of noisy measurements it is clear that there are limits on the accuracy and precision we can reach. This is precisely quantified in the Cramér-Rao inequality. This inequality provides a lower bound on the covariance matrix of a(n) (un)biased estimator starting from the likelihood function. First we introduce the likelihood function; next we present the Cramér-Rao lower bound.

Consider the measurements $z \in R^N$ obtained from a system described by a hypothetical, exact model that is parameterized in θ. These measurements are disturbed by noise and are hence stochastic variables that are characterized by a probability density function $f\left(z\middle|\theta_0\right)$ that depends on the exact model parameters θ_0 with $\int_{z \in R^N} f\left(z\middle|\theta_0\right) dz = 1$. Next we can interpret this relation conversely, viz., how likely is it that a specific set of measurements $z = z_m$ are generated by a system with parameters θ? In other words, we consider now a given set of measurements and view the model parameters as the free variables:

$$L\left(z_m\middle|\theta\right) = f\left(z = z_m\middle|\theta\right) \tag{5}$$

with θ the free variables. $L\left(z_m\middle|\theta\right)$ is called the likelihood function. In many calculations the log likelihood function $l\left(z\middle|\theta\right) = \ln\left(L\left(z\middle|\theta\right)\right)$ is used. In (5) we used z_m to indicate explicitly that we use the numerical values of the measurements that were obtained from the experiments. From here on we just use z as a symbol because it will be clear from the context what interpretation should be given to z. The reader should be aware that $L\left(z\middle|\theta\right)$ is not a probability density function with respect to θ since $\int_\theta L\left(z\middle|\theta\right) d\theta \neq 1$. Notice the subtle difference in terminology, i.e. probability is replaced by likelihood.

The Cramér-Rao lower bound gives a lower limit on the covariance matrix of parameters. This limit is universal and independent of the selected estimator: no estimator that violates this bound can be found. It is given by

$$\begin{aligned} CR\left(\theta_0\right) &= \left(I_{n_\theta} + \frac{\partial b_\theta}{\partial \theta}\right)^T Fi^{-1}\left(\theta_0\right)\left(I_{n_\theta} + \frac{\partial b_\theta}{\partial \theta}\right) \\ Fi\left(\theta_0\right) &= E\left\{\left(\frac{\partial l\left(z\middle|\theta\right)}{\partial \theta}\right)^T \left(\frac{\partial l\left(z\middle|\theta\right)}{\partial \theta}\right)\right\} = -E\left\{\frac{\partial^2 l\left(z\middle|\theta\right)}{\partial \theta^2}\right\} \end{aligned} \tag{6}$$

The derivatives are calculated in $\theta = \theta_0$, and $b_\theta = E\left\{\hat{\theta}\right\} - \theta_0$ is the bias on the estimator.

$Fi(\theta)$ is called the Fisher information matrix: it is a measure for the information in an experiment: the larger the matrix the more information there is. In (6) it is assumed that the first and second derivatives of the log likelihood function exist with respect to θ.

Adding additional parameters to a model increases the minimum attainable uncertainty on it; the Cramér-Rao lower bound will increase. Of course, these parameters may be needed to remove systematic errors so that a balance between stochastic errors and systematic errors is achieved.

This problem is studied in the model selection step of the identification procedure.

The Cramér-Rao bound is used to verify the efficiency of an estimator.

- *(efficiency):* An unbiased estimator is called efficient if its covariance matrix is smaller than that of any other unbiased estimator.

An unbiased estimator that reaches the Cramér-Rao lower bound is also an efficient estimator. For biased estimators, the generalized expression should be used.

2.3. A Statistical Approach to the Estimation Problem

From the resistance example, it turns out that an intuitive approach to a parameter estimation problem can cause serious errors without even being noticed. To avoid severe mistakes, a theoretical framework is needed. Here a statistical development of the parameter estimation theory is made.

Often, four related estimators are studied: the least squares (LS) estimator, weighted least squares (WLS) estimator, maximum likelihood (ML) estimator and, finally, the Bayes estimator.

It should be clear that as mentioned before, it is still possible to use other estimators, like the least absolute values. However, a comprehensive overview of all possible techniques is beyond the scope of this chapter. For these reasons we focus on the LS, WLS and ML estimator because we will use these later on in this section.

To use the Bayes estimator, the *a priori* pdf of the unknown parameters and the pdf of the noise on the measurements are required. Although it seems, at first, quite strange that the parameters have a pdf, we use this concept regularly in daily life, where we combine observations with our past experience that indicates to us that some results or more probable than others.
The ML estimator only requires knowledge of the pdf of the noise on the measurements, and the WLS estimator can be applied optimally if the covariance matrix of the noise is known. Even if this information is lacking, the LS method is usable. Some of these estimators will be explained in more detail in the following sections.

-
-
-

Bibliography

General introduction to identification theory

Beck, J. V. and K. J. Arnold (1977). Parameter Estimation in Engineering and Science. John Wiley and Sons, New York. [A very general introduction to identification].

Goodwin, G. C. and R. L. Payne (1977). Dynamic System Identification. Experimental Design and Data Analysis. Academic Press, New York. [Special attention is paid to experiment design].

Kendall, M. and A. Stuart (1979). Inference and Relationship, vol. 2 of The Advanced Theory of Statistics (4th ed.). Charles Griffin, London. [Basic work on statistics].

Norton, J. P. (1986). An Introduction to Identification. Academic Press, London. [A basic introduction to system identification].

Sorenson, H. W. (1980). Parameter Estimation: Principles and Problems. Marcel Dekker, New York. [A general introduction, not devoted to a specific field. Gives insight in the general identification theory].

Walter, E. and L. Pronzato (1997). Identification of Parametric Models from Experimental Data. Springer, Paris. [An introduction to system identification with special attention for identifiability, experiment design, and the generation of uncertainty bounds].

Books dealing with system identification

Hannan E.J. and Manfred Deistler (1988), The Statistical Theory of Linear Systems, Wiley. [A thorough statistical approach to system identification].

Johansson R., (1993). System Modeling and Identification, Prentice Hall. [A general introduction to the field of system identification].

Ljung, L. (1999). System Identification: Theory for the User (second edition). Prentice Hall, Upper Saddle River. [State of the art on time domain system identification].

Ljung, L. and T. Söderström (1983). Theory and Practice of Recursive Identification. MIT Press, Cambridge, MA. [An introduction to recursive (online) identification].

Pintelon R. and J. Schoukens (2001). System Identification: a Frequency Domain Approach. IEEE Press. [State of the art of frequency domain system identification].

Söderström, T. and P. Stoica (1989). System Identification. Prentice-Hall, Englewood Cliffs. [State of the art of timde domain system identification].

Biographical Sketches

Johan Schoukens was born in Belgium in 1957. He received the degree of engineer in 1980 and the degree of doctor in applied sciences in 1985, both from the Vrije Universiteit Brussel, Brussels, Belgium. He is presently professor at the Vrije Universiteit Brussel. The prime factors of his interest are in the field of system identification for linear and nonlinear systems, and growing tomatoes in his green house.

Rik Pintelon was born in Gent, Belgium, on December 4, 1959. He received the degree of electrical engineer (burgerlijk ingenieur) in July 1982, the degree of doctor in applied sciences in January 1988, and the qualification to teach at university level (geaggregeerde voor het hoger onderwijs) in April 1994, all from the Vrije Universiteit Brussel (VUB), Brussels, Belgium. From October 1982 till September 2000 he was a researcher of the Fund for Scientific Research - Flanders at the VUB. Since October 2000 he is professor at the VUB in the Electrical Measurement Department (ELEC). His main research interests are in the field of parameter estimation / system identification, and signal processing.