1
Introduction

Making observations through the senses of the environment around us is a natural activity of living species. The information acquired is diverse, consisting for example of sound signals and images. The information is processed and used to make a particular model of the environment that is applicable to the situation at hand. This act of model building based on observations is embedded in our human nature and plays an important role in daily decision making.

Model building through observations also plays a very important role in many branches of science. Despite the importance of making observations through our senses, scientific observations are often made via measurement instruments or sensors. The measurement data that these sensors acquire often need to be processed to judge or validate the experiment, or to obtain more information on conducting the experiment. Data are often used to build a mathematical model that describes the dynamical properties of the experiment. System-identification methods are systematic methods that can be used to build mathematical models from measured data. One important use of such mathematical models is in predicting model quantities by filtering acquired measurements.

A milestone in the history of filtering and system identification is the method of least squares developed just before 1800 by Johann Carl Friedrich Gauss (1777–1855). The use of least squares in filtering and identification is a recurring theme in this book. What follows is a brief sketch of the historical context that characterized the early development of the least-squares method. It is based on an overview given by Bühler (1981).

At the time Gauss first developed the least-squares method, he did not consider it very important. The first publication on the least-squares

method was published by Adrien-Marie Legendre (1752–1833) in 1806, when Gauss had already clearly and frequently used the method much earlier. Gauss motivated and derived the method of least squares substantially in the papers *Theoria combinationis observationum erroribus minimis obnoxiae* I and II of 1821 and 1823. Part I is devoted to the theory and Part II contains applications, mostly to problems from astronomy. In Part I he developed a probability theory for accidental errors (*Zufallsfehler*). Here Gauss defined a (probability distribution) function $\phi(x)$ for the error in the observation x. On the basis of this function, the product $\phi(x)\mathrm{d}x$ is the probability that the error falls within the interval between x and $x+\mathrm{d}x$. The function $\phi(x)$ had to satisfy the normalization condition

$$\int_{-\infty}^{\infty} \phi(x)\mathrm{d}x = 1.$$

The decisive requirement postulated by Gauss is that the integral

$$\int_{-\infty}^{\infty} x^2\phi(x)\mathrm{d}x$$

attains a minimum. The selection of the square of the error as the most suitable weight is why this method is called the method of least squares. This selection was doubted by Pierre-Simon Laplace (1749–1827), who had earlier tried to use the absolute value of the error. Computationally the choice of the square is superior to Laplace's original method.

After the development of the basic theory of the least-squares method, Gauss had to find a suitable function $\phi(x)$. At this point Gauss introduced, after some heuristics, the Gaussian distribution

$$\phi(x) = \frac{1}{\pi}\mathrm{e}^{-x^2}$$

as a "natural" way in which errors of observation occur. Gauss never mentioned in his papers statistical distribution functions different from the Gaussian one. He was caught in his own success; the applications to which he applied his theory did not stimulate him to look for other distribution functions. The least-squares method was, at the beginning of the nineteenth century, his indispensable theoretical tool in experimental research; and he saw it as the most important witness to the connection between mathematics and Nature.

Still today, the ramifications of the least-squares method in mathematical modeling are tremendous and any book on this topic has to narrow

itself down to a restrictive class of problems. In this introductory textbook on system identification we focus mainly on the identification of linear state-space models from measured data sequences of inputs and outputs of the engineering system that we want to model. Though this focused approach may at first seem to rule out major contributions in the field of system identification, the contrary is the case. It will be shown in the book that the state-space approach chosen is capable of treating many existing identification methods for estimating the parameters in a difference equation as special cases. Examples are given for the widely used ARX and ARMAX models (Ljung, 1999).

The central goal of the book is to help the reader discover how the linear least-squares method can solve, or help in solving, different variants of the linear state-space model-identification problem. The linear least-squares method can be formulated as a deterministic parameter-optimization problem of the form

$$\min_x \mu^{\mathrm{T}}\mu \quad \text{subject to } y = Fx + \mu, \tag{1.1}$$

with the vector $y \in \mathbb{R}^N$ and the matrix $F \in \mathbb{R}^{N\times n}$ given and with $x \in \mathbb{R}^n$ the vector of unknown parameters to be determined. The solution of this optimization problem is the subject of a large number of textbooks. Although its analytic solution can be given in a proof of only a few lines, these textbooks analyze the least-squares solution from different perspectives. Examples are the statistical interpretation of the solution under various assumptions on the entries of the matrix F and the perturbation vector μ, or the numerical solution in a computationally efficient manner by exploiting structure in the matrix F. For an advanced study of the least-squares problem and its applications in many signal-processing problems, we refer to the book of Kailath *et al.* (2000).

The main course of this book is preceded by three introductory chapters. In Chapter 2 a refreshment survey of matrix linear algebra is given. Chapter 3 gives a brief overview of signal transforms and linear system theory for deterministic signals and systems. Chapter 4 treats random variables and random signals. Understanding the system-identification methods discussed in this book depends on a profound mastering of the background material presented in these three chapters.

Often, the starting point of identifying a dynamical model is the determination of a predictor. Therefore, in Chapter 5, we first study the

prediction of the state of a linear state-space model. The state-prediction or state-observation problem requires, in addition to the inputs and outputs, knowledge of the dynamic model (in state-space form) and the mean and covariance matrix of the stochastic perturbations. The goal is to reconstruct the state sequence from this knowledge. The optimality of the state-reconstruction problem can be defined in a least-squares sense. In Chapter 5, it is shown that the optimal predictor or Kalman-filter problem can be formulated and solved as a (weighted) linear least-squares problem. This formulation and solution of the Kalman-filter problem was first proposed by Paige (1985). The main advantage of this formulation is that a (recursive) solution can simply be derived from elementary linear-algebra concepts, such as Gaussian elimination for solving an overdetermined set of equations. We will briefly discuss the application of Kalman filtering for estimating unknown inputs of a dynamical system.

Chapter 6 discusses the estimation of input–output descriptions of linear state-space models in the frequency domain. The estimation of such descriptions, like the frequency response function (FRF) is based on the (discrete) Fourier transform of time sequences. The study in this chapter includes the effect that the practical constraint of the finite duration of the experiment has on the accuracy of the FRF estimate. A brief exposition on the use of the fast Fourier transform (FFT) in deriving fast algorithmic implementations is given. The availability of fast algorithms is one of the main advantages of frequency-domain methods when dealing with large amounts of data. In major parts of industry, such as the automobile and aircraft industry, it is therefore still the main tool for retrieving information about dynamic systems.

Chapter 7 discusses the estimation of the entries of the system matrices of a state-space model, under the assumptions that the output observations are corrupted by additive white noise and the state vector of the model has a fixed and known dimension. This problem gives rise to the so-called output-error methods (Ljung, 1999). This elementary estimation problem reveals a number of issues that are at the heart of a wide variety of identification approaches. The key problem to start with is that of how to express the entries of the system matrices as functions of an unknown parameter vector. The choice of this parameter vector is referred to in this textbook as the parameterization problem. Various alternatives for parameterizing multivariable state-space models are proposed. Once a parameterization has been chosen, the output-error problem can also be formulated as the following least-squares

problem:

$$\min_{x_1,x_2} \mu^{\mathrm{T}}\mu \quad \text{subject to } y = F(x_1)x_2 + \mu, \tag{1.2}$$

where the unknown parameter vectors x_1 and x_2 need to be determined. This type of least-squares problem is much harder to tackle than its linear variant (1.1), because the matrix F depends on the unknown parameters x_1. It is usually solved iteratively and therefore requires starting values of the parameter vectors x_1 and x_2. Furthermore, in general there is no guarantee that such an iterative numerical procedure converges to the global optimum of the cost function $\mu^{\mathrm{T}}\mu$. In this chapter special attention is paid to the numerical implementation of iterative procedures for output-error optimization. After having obtained an estimate of the unknown parameter vector, the problem of assessing the accuracy of the obtained estimates is addressed via the evaluation of the covariance matrices of the estimates under the assumption that the estimates are unbiased. We end this chapter by discussing how to avoid biased solutions when the additive noise is no longer white.

Chapter 8 presents the classical prediction-error method (PEM) (Ljung, 1999) for the identification of a predictor model (Kalman filter) with a fixed and known state dimension from measured input and output data. The problem boils down to estimating the parameters of a predictor model given by the innovation representation of the Kalman filter. The problems and solutions presented in Chapter 7 for the output-error case are adapted for these predictor models. In addition to the presentation of the prediction-error method for general multivariable state-space models, special attention is given to single-input, single-output (SISO) systems. This is done, first, to show that well-known model structures such as the ARMAX model can be treated as a particular canonical parameterization of a state-space model. Second, it enables a qualitative analysis of the bias when identifying a model that has a state dimension or a noise model different from the system that generated the data.

Chapter 9 treats the recently developed class of subspace identification methods. These methods are capable of providing accurate estimates of multivariable state-space models under general noise perturbations by just solving a linear least-squares problem of the form (1.1). The interest in subspace methods, both in academia and in industry, stems partly from the fact that no model parameterization is necessary to estimate a model and its order. This is achieved by relating key subspaces defined from matrices of the model to structured matrices

constructed from the available observations. The central role the subspace plays explains the name given to these methods. A distinction is made among different types of subspace methods, depending on how they use the concept of instrumental variables to cope consistently with various noise scenarios. Although, in practice, it has been demonstrated that subspace methods immediately provide accurate models, they do not optimize the prediction-error criterion as the prediction-error method does. To achieve this statistical optimum, we could use the estimates obtained with subspace methods as starting values for the prediction-error method. This concept has been proposed by Ljung (MathWorks, 2000a), for example.

Chapter 10 establishes the link between model estimation algorithms and their use in a real-life identification experiment. To set up, analyze, and improve an identification experiment, a cyclic procedure such as that outlined by Ljung (1999) is discussed. The cyclic procedure aims at a systematic treatment of many choices that need to be made in system identification. These choices include the selection of the experimental circumstances (for example sampling frequency, experiment duration, and type of input signal), the treatment of the recorded time sequences (detrending, removing outliers, and filtering) and the selection of a model structure (model order and delay) for the parameter-estimation algorithms. Here we include a brief discussion on how the subspace methods of Chapter 9 and the parametric methods of Chapters 7 and 8 can work together in assisting the system-identification practitioner to make choices regarding the model structure. It is this merging of subspace and prediction-error methods that makes the overall identification cycle feasible for multivariable systems. When using the prediction-error method in isolation, finding the appropriate model structure would require the testing of an extremely large amount of possibilities. This is infeasible in practice, since often not just one model needs to be identified, but a series of models for different experimental conditions.

At the end of each chapter dedicated exercises are included to let the reader experiment with the development and application of new algorithms. To facilitate the use of the methods described, the authors have developed a Matlab toolbox containing the identification methods described, together with a comprehensive software guide (Verhaegen *et al.*, 2003).

Filtering and system identification are excellent examples of multidisciplinary science, not only because of their versatility of application in

many different fields, but also because they bring together fundamental knowledge from a wide number of (mathematical) disciplines. The authors are convinced that the current outline of the textbook should be considered as just an introduction to the fascinating field of system identification. System identification is a branch of science that illustrates very well the saying that the proof of the pudding is in the eating. Study and master the material in this textbook, but, most importantly, use it!

2
Linear algebra

After studying this chapter you will be able to

- apply basic operations to vectors and matrices;
- define a vector space;
- define a subspace of a vector space;
- compute the rank of a matrix;
- list the four fundamental subspaces defined by a linear transformation;
- compute the inverse, determinant, eigenvalues, and eigenvectors of a square matrix.
- describe what positive-definite matrices are;
- compute some important matrix decompositions, such as the eigenvalue decomposition, the singular-value decomposition and the QR factorization;
- solve linear equations using techniques from linear algebra;
- describe the deterministic least-squares problem; and
- solve the deterministic least-squares problem in numerically sound ways.

2.1 Introduction

In this chapter we review some basic topics from linear algebra. The material presented is frequently used in the subsequent chapters.

Since the 1960s linear algebra has gained a prominent role in engineering as a contributing factor to the success of technological breakthroughs.

Linear algebra provides tools for numerically solving system-theoretic problems, such as filtering and control problems. The widespread use of linear algebra tools in engineering has in its turn stimulated the development of the field of linear algebra, especially the numerical analysis of algorithms. A boost to the prominent role of linear algebra in engineering has certainly been provided by the introduction and widespread use of computer-aided-design packages such as Matlab (MathWorks, 2000b) and SciLab (Gomez, 1999). The user-friendliness of these packages allow us to program solutions for complex system-theoretic problems in just a few lines of code. Thus the prototyping of new algorithms is greatly speeded-up. However, on the other hand, there is also need for a word of caution: The coding in Matlab may give the user the impression that one successful Matlab run is equivalent to a full proof of a new theory. In order to avoid the cultivation of such a "proven-by-Matlab" attitude, the refreshment survey in this chapter and the use of linear algebra in later chapters concern primarily the derivation of the algorithms rather than their use. The use of Matlab routines for the class of filtering and identification problems analyzed in this book is described in detail in the comprehensive software guide (Verhaegen *et al.*, 2003).

We start this chapter with a review of two basic elements of linear algebra: vectors and matrices. Vectors are described in Section 2.2, matrices in Section 2.3. For a special class of matrices, square matrices, several important concepts exist, and these are described in Section 2.4. Section 2.5 describes some matrix decompositions that have proven to be useful in the context of filtering and estimation. Finally, in Sections 2.6 and 2.7 we focus on least-squares problems in which an overdetermined set of linear equations needs to be solved. These problems are of particular interest, since a lot of filtering, estimation, and even control problems can be written as linear (weighted) least-squares problems.

2.2 Vectors

A *vector* is an array of real or complex numbers. Throughout this book we use $\mathbb{R}$ to denote the set of real numbers and $\mathbb{C}$ to denote the set of complex numbers. Vectors come in two flavors, *column vectors* and *row vectors*. The column vector that consists of the elements $x_1, x_2, \ldots, x_n$

with $x_i \in \mathbb{C}$ will be denoted by x, that is,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

In this book a vector denoted by a lower-case character will always be a column vector. Row vectors are denoted by x^{T}, that is,

$$x^{\mathrm{T}} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}.$$

The row vector x^{T} is also called the *transpose* of the column vector x. The number of elements in a vector is called the *dimension* of the vector. A vector having n elements is referred to as an n-dimensional vector. We use the notation $x \in \mathbb{C}^n$ to denote an n-dimensional vector that has complex-valued elements. Obviously, an n-dimensional vector with real-valued elements is denoted by $x \in \mathbb{R}^n$. In this book, most vectors will be real-valued; therefore, in the remaining part of this chapter we will restrict ourselves to real-valued vectors. However, most results can readily be extended to complex-valued vectors.

The multiplication of a vector $x \in \mathbb{R}^n$ by a scalar $\alpha \in \mathbb{R}$ is defined as

$$\alpha x = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}.$$

The sum of two vectors $x, y \in \mathbb{R}^n$ is defined as

$$x + y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

The standard *inner product* of two vectors $x, y \in \mathbb{R}^n$ is equal to

$$x^{\mathrm{T}} y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

The *2-norm* of a vector x, denoted by $||x||_2$, is the square root of the inner product of this vector with itself, that is,

$$||x||_2 = \sqrt{x^{\mathrm{T}} x}.$$