Kalman Filtering: Theory and Practice

Using MATLAB

Second Edition

MOHINDER S. GREWAL
California State University at Fullerton

ANGUS P. ANDREWS
Rockwell Science Center

A Wiley-Interscience Publication
John Wiley & Sons, Inc.
NEW YORK • CHICHESTER • WEINHEIM • BRISBANE • SINGAPORE • TORONTO

Contents

Preface

The first edition of this book was published by Prentice-Hall in 1993. With this second edition, as with the first, our primary objective is to provide our readers a working familiarity with both the *theoretical* and *practical* aspects of Kalman filtering by including "real-world" problems in practice as illustrative examples. We are pleased to have this opportunity to incorporate the many helpful corrections and suggestions from our colleagues and students over the last several years for the overall improvement of the textbook. The book covers the historical background of Kalman filtering and the more practical aspects of implementation: how to represent the problem in a mathematical model, analyze the performance of the estimator as a function of model parameters, implement the mechanization equations in numerically stable algorithms, assess its computational requirements, test the validity of results, and monitor the filter performance in operation. These are important attributes of the subject that are often overlooked in theoretical treatments but are necessary for application of the theory to real-world problems.

We have converted all algorithm listings and all software to MATLAB®[1], so that users can take advantage of its excellent graphing capabilities and a programming interface that is very close to the mathematical equations used for defining Kalman filtering and its applications. See Appendix A, Section A.2, for more information on MATLAB.

The inclusion of the software is practically a matter of necessity, because Kalman filtering would not be very useful without computers to implement it. It is a better learning experience for the student to discover how the Kalman filter works by observing it in action.

The implementation of Kalman filtering on computers also illuminates some of the practical considerations of finite-wordlength arithmetic and the need for alter-

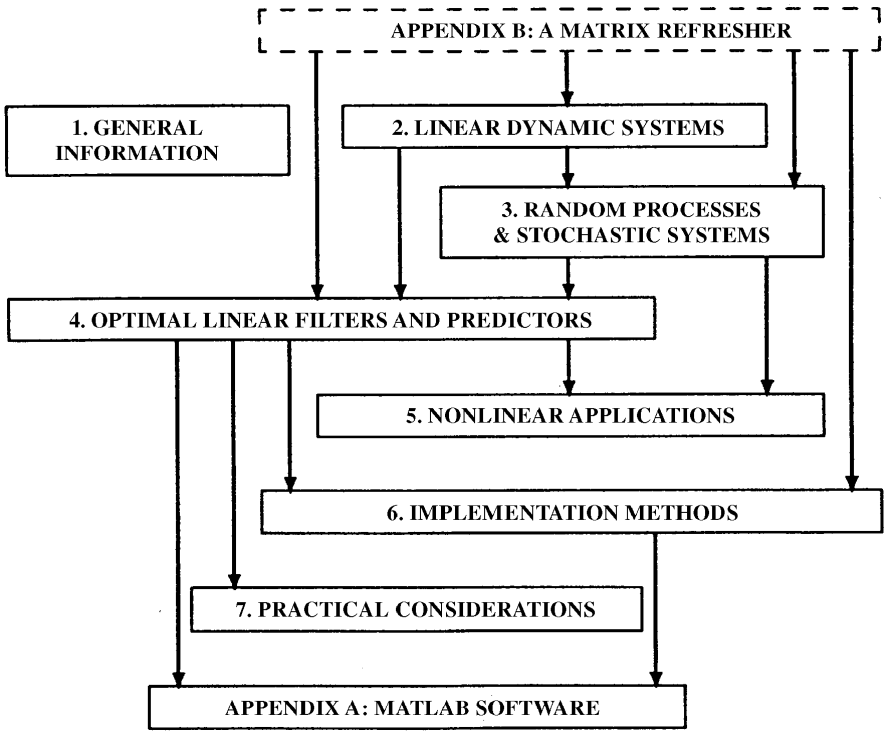
[1] MATLAB is a registered trademark of The Mathworks, Inc.

native algorithms to preserve the accuracy of the results. If the student wishes to apply what she or he learns, then it is essential that she or he experience its workings and failings—and learn to recognize the difference.

The book is organized for use as a text for an introductory course in stochastic processes at the senior level and as a first-year graduate-level course in Kalman filtering theory and application. It could also be used for self-instruction or for purposes of review by practicing engineers and scientists who are not intimately familiar with the subject. The organization of the material is illustrated by the following chapter-level dependency graph, which shows how the subject of each chapter depends upon material in other chapters. The arrows in the figure indicate the recommended order of study. Boxes above another box and connected by arrows indicate that the material represented by the upper boxes is background material for the subject in the lower box.

Chapter 1 provides an informal introduction to the general subject matter by way of its history of development and application. Chapters 2 and 3 and Appendix B cover the essential background material on linear systems, probability, stochastic processes, and modeling. These chapters could be covered in a senior-level course in electrical, computer, and systems engineering.

Chapter 4 covers linear optimal filters and predictors, with detailed examples of applications. Chapter 5 is devoted to nonlinear estimation by "extended" Kalman



filters. Applications of these techniques to the identification of unknown parameters of systems are given as examples. Chapter 6 covers the more modern implementation techniques, with algorithms provided for computer implementation.

Chapter 7 deals with more practical matters of implementation and use beyond the numerical methods of Chapter 6. These matters include memory and throughput requirements (and methods to reduce them), divergence problems (and effective remedies), and practical approaches to suboptimal filtering and measurement selection.

Chapters 4–7 cover the essential material for a first-year graduate class in Kalman filtering theory and application or as a basic course in digital estimation theory and application. A solutions manual for each chapter's problems is available.

PROF. MOHINDER S. GREWAL, PHD, PE

California State University at Fullerton

ANGUS P. ANDREWS, PHD

Rockwell Science Center, Thousand Oaks, California

Acknowledgments

The authors express their appreciation to the following individuals for their contributions during the preparation of the first edition: Robert W. Bass, E. Richard Cohen, Thomas W. De Vries, Reverend Joseph Gaffney, Thomas L. Gunckel II, Dwayne Heckman, Robert A. Hubbs, Thomas Kailath, Rudolf E. Kalman, Alan J. Laub, Robert F. Nease, John C. Pinson, John M. Richardson, Jorma Rissanen, Gerald E. Runyon, Joseph Smith and Donald F. Wiberg. We also express our appreciation to Donald Knuth and Leslie Lamport for TEX and LATEX, respectively.

In addition, the following individuals deserve special recognition for their careful review, corrections, and suggestions for improving the second edition: Dean Dang and Gordon Inverarity.

Most of all, for their dedication, support, and understanding through both editions, we dedicate this book to Sonja Grewal and Jeri Andrews.

M. S. G., A. P. A.

Kalman Filtering: Theory and Practice Using MATLAB, Second Edition,
Mohinder S. Grewal, Angus P. Andrews

ISBNs: 0-471-39254-5 (Hardback); 0-471-26638-8 (Electronic)

1

General Information

. . . the things of this world cannot be made known without mathematics.
—Roger Bacon (1220–1292), Opus Majus, transl. R. Burke, 1928

1.1 ON KALMAN FILTERING

1.1.1 First of All: What Is a Kalman Filter?

Theoretically the Kalman Filter is an estimator for what is called the *linear-quadratic problem*, which is the problem of estimating the instantaneous "state" (a concept that will be made more precise in the next chapter) of a linear dynamic system perturbed by white noise—by using measurements linearly related to the state but corrupted by white noise. The resulting estimator is statistically optimal with respect to any quadratic function of estimation error.

Practically, it is certainly one of the greater discoveries in the history of statistical estimation theory and possibly the greatest discovery in the twentieth century. It has enabled humankind to do many things that could not have been done without it, and it has become as indispensable as silicon in the makeup of many electronic systems. Its most immediate applications have been for the control of complex dynamic systems such as continuous manufacturing processes, aircraft, ships, or spacecraft. To control a dynamic system, you must first know what it is doing. For these applications, it is not always possible or desirable to measure every variable that you want to control, and the Kalman filter provides a means for inferring the missing information from indirect (and noisy) measurements. The Kalman filter is also used for predicting the likely future courses of dynamic systems that people are not likely to control, such as the flow of rivers during flood, the trajectories of celestial bodies, or the prices of traded commodities.

From a practical standpoint, these are the perspectives that this book will present:

- *It is only a tool*. It does not solve any problem all by itself, although it can make it easier for you to do it. It is not a *physical* tool, but a *mathematical* one. It is made from mathematical models, which are essentially tools for the mind. They make mental work more efficient, just as mechanical tools make physical work more efficient. As with any tool, it is important to understand its use and function before you can apply it effectively. The purpose of this book is to make you sufficiently familiar with and proficient in the use of the Kalman filter that you can apply it correctly and efficiently.
- *It is a computer program*. It has been called "ideally suited to digital computer implementation" [21], in part because it uses a *finite representation* of the estimation problem—by a *finite* number of variables. It does, however, assume that these variables are *real numbers*—with *infinite* precision. Some of the problems encountered in its use arise from the distinction between finite dimension and finite information, and the distinction between "finite" and "manageable" problem sizes. These are all issues on the practical side of Kalman filtering that must be considered along with the theory.
- *It is a complete statistical characterization of an estimation problem*. It is much more than an *estimator,* because it propagates the entire *probability distribution* of the variables it is tasked to estimate. This is a complete characterization of the current *state of knowledge* of the dynamic system, including the influence of all past measurements. These probability distributions are also useful for statistical analysis and the predictive design of sensor systems.
- *In a limited context, it is a learning method*. It uses a model of the estimation problem that distinguishes between *phenomena* (what one is able to observe), *noumena* (what is really going on), and the state of knowledge about the noumena that one can deduce from the phenomena. That state of knowledge is represented by probability distributions. To the extent that those probability distributions represent *knowledge* of the real world and the cumulative processing of knowledge is *learning,* this is a learning process. It is a fairly simple one, but quite effective in many applications.

If these answers provide the level of understanding that you were seeking, then there is no need for you to read the rest of the book. If you need to understand Kalman filters well enough to use them, then read on!

1.1.2 How It Came to Be Called a Filter

It might seem strange that the term "filter" would apply to an estimator. More commonly, a filter is a physical device for removing unwanted fractions of mixtures. (The word *felt* comes from the same medieval Latin stem, for the material was used as a filter for liquids.) Originally, a filter solved the problem of separating unwanted components of gas–liquid–solid mixtures. In the era of crystal radios and vacuum tubes, the term was applied to analog circuits that "filter" electronic signals. These

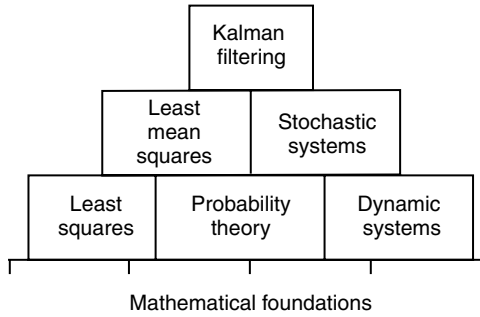


Fig. 1.1 *Foundational concepts in Kalman filtering.*

signals are mixtures of different frequency components, and these physical devices preferentially attenuate unwanted frequencies.

This concept was extended in the 1930s and 1940s to the separation of "signals" from "noise," both of which were characterized by their power spectral densities. Kolmogorov and Wiener used this statistical characterization of their probability distributions in forming an optimal estimate of the signal, given the sum of the signal and noise.

With Kalman filtering the term assumed a meaning that is well beyond the original idea of *separation* of the components of a mixture. It has also come to include the solution of an *inversion problem,* in which one knows how to represent the measurable variables as functions of the variables of principal interest. In essence, it inverts this functional relationship and estimates the independent variables as inverted functions of the dependent (measurable) variables. These variables of interest are also allowed to be dynamic, with dynamics that are only partially predictable.

1.1.3 Its Mathematical Foundations

Figure 1.1 depicts the essential subjects forming the foundations for Kalman filtering theory. Although this shows Kalman filtering as the apex of a pyramid, it is itself but part of the foundations of another discipline—"modern" control theory—and a proper subset of statistical decision theory.

We will examine only the top three layers of the pyramid in this book, and a little of the underlying mathematics[1] (matrix theory) in Appendix B.

1.1.4 What It Is Used For

The applications of Kalman filtering encompass many fields, but its use as a tool is almost exclusively for two purposes: *estimation* and *performance analysis* of estimators.

[1]It is best that one not examine the bottommost layers of these mathematical foundations too carefully, anyway. They eventually rest on human intellect, the foundations of which are not as well understood.

Role 1: Estimating the State of Dynamic Systems What is a dynamic system? Almost everything, if you are picky about it. Except for a few fundamental physical constants, there is hardly anything in the universe that is truly *constant*. The orbital parameters of the asteroid Ceres are not constant, and even the "fixed" stars and continents are moving. Nearly all physical systems are dynamic to some degree. If one wants very precise estimates of their characteristics over time, then one has to take their dynamics into consideration.

The problem is that one does not always know their dynamics very precisely either. Given this state of partial ignorance, the best one can do is express our ignorance more precisely—using probabilities. The Kalman filter allows us to estimate the state of dynamic systems with certain types of random behavior by using such statistical information. A few examples of such systems are listed in the second column of Table 1.1.

Role 2: The Analysis of Estimation Systems. The third column of Table 1.1 lists some possible sensor types that might be used in estimating the state of the corresponding dynamic systems. The objective of design analysis is to determine how best to use these sensor types for a given set of design criteria. These criteria are typically related to estimation accuracy and system cost.

The Kalman filter uses a complete description of the probability distribution of its estimation errors in determining the optimal filtering gains, and this probability distribution may be used in assessing its performance as a function of the "design parameters" of an estimation system, such as

- the types of sensors to be used,
- the locations and orientations of the various sensor types with respect to the system to be estimated,

TABLE 1.1 Examples of Estimation Problems

Application	Dynamic System	Sensor Types
Process control	Chemical plant	Pressure
		Temperature
		Flow rate
		Gas analyzer
Flood prediction	River system	Water level
		Rain gauge
		Weather radar
Tracking	Spacecraft	Radar
		Imaging system
Navigation	Ship	Sextant
		Log
		Gyroscope
		Accelerometer
		Global Positioning System (GPS) receiver

- the allowable noise characteristics of the sensors,
- the prefiltering methods for smoothing sensor noise,
- the data sampling rates for the various sensor types, and
- the level of model simplification to reduce implementation requirements.

The analytical capability of the Kalman filter formalism also allows a system designer to assign an "error budget" to subsystems of an estimation system and to trade off the budget allocations to optimize cost or other measures of performance while achieving a required level of estimation accuracy.

1.2 ON ESTIMATION METHODS

We consider here just a few of the sources of intellectual material presented in the remaining chapters and principally those contributors[2] whose lifelines are shown in Figure 1.2. These cover only 500 years, and the study and development of mathematical concepts goes back beyond history. Readers interested in more detailed histories of the subject are referred to the survey articles by Kailath [25, 176], Lainiotis [192], Mendel and Geiseking [203], and Sorenson [47, 224] and the personal accounts of Battin [135] and Schmidt [216].

1.2.1 Beginnings of Estimation Theory

The first method for forming an *optimal* estimate from noisy data is the *method of least squares*. Its discovery is generally attributed to Carl Friedrich Gauss (1777–1855) in 1795. The inevitability of measurement errors had been recognized since the time of Galileo Galilei (1564–1642) , but this was the first formal method for dealing with them. Although it is more commonly used for linear estimation problems, Gauss first used it for a nonlinear estimation problem in mathematical astronomy, which was part of a dramatic moment in the history of astronomy. The following narrative was gleaned from many sources, with the majority of the material from the account by Baker and Makemson [97]:

> On January 1, 1801, the first day of the nineteenth century, the Italian astronomer Giuseppe Piazzi was checking an entry in a star catalog. Unbeknown to Piazzi, the entry had been added erroneously by the printer. While searching for the "missing" star, Piazzi discovered, instead, a new planet. It was *Ceres*—the largest of the minor planets and the first to be discovered—but Piazzi did not know that yet. He was able to track and measure its apparent motion against the "fixed" star background during 41 nights of viewing from Palermo before his work was interrupted. When he returned to his work, however, he was unable to find Ceres again.

[2]The only contributor after R. E. Kalman on this list is Gerald J. Bierman, an early and persistent advocate of *numerically stable* estimation methods. Other recent contributors are acknowledged in Chapter 6.

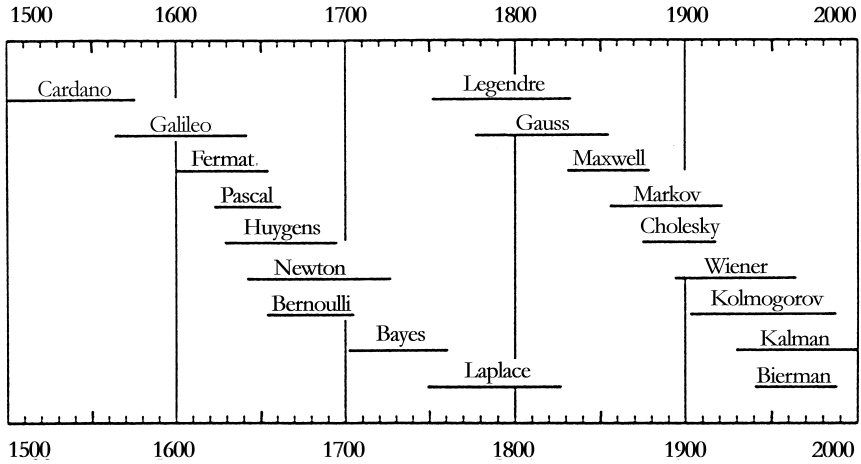


Fig. 1.2 *Lifelines of referenced historical figures and R. E. Kalman.*

On January 24, Piazzi had written of his discovery to Johann Bode. Bode is best known for *Bode's law*, which states that the distances of the planets from the sun, in astronomical units, are given by the sequence

$$d_n = \tfrac{1}{10}(4 + 3 \times 2^n) \quad \text{for } n = -\infty, 0, 1, 2, ?, 4, 5, \ldots. \tag{1.1}$$

Actually, it was not Bode, but Johann Tietz who first proposed this formula, in 1772. At that time there were only six known planets. In 1781, Friedrich Herschel discovered Uranus, which fit nicely into this formula for $n = 6$. No planet had been discovered for $n = 3$. Spurred on by Bode, an association of European astronomers had been searching for the "missing" eighth planet for nearly 30 years. Piazzi was not part of this association, but he did inform Bode of his unintended discovery.

Piazzi's letter did not reach Bode until March 20. (Electronic mail was discovered much later.) Bode suspected that Piazzi's discovery might be the missing planet, but there was insufficient data for determining its orbital elements by the methods then available. It is a problem in nonlinear equations that Newton, himself, had declared as being among the most difficult in mathematical astronomy. Nobody had solved it and, as a result, Ceres was lost in space again.

Piazzi's discoveries were not published until the autumn of 1801. The possible discovery—and subsequent loss—of a new planet, coinciding with the beginning of a new century, was exciting news. It contradicted a philosophical justification for there being only seven planets—the number known before Ceres and a number defended by the respected philosopher Georg Hegel, among others. Hegel had recently published a book in which he chastised the astronomers for wasting their time in searching for an eighth planet when there was a sound philosophical justification for there being only seven. The new planet became a subject of conversation in intellectual circles nearly everywhere. Fortunately, the problem caught the attention of a 24-year-old mathematician at Göttingen named Carl Friedrich Gauss.

Gauss had toyed with the orbit determination problem a few weeks earlier but had set it aside for other interests. He now devoted most of his time to the problem, produced an estimate of the orbit of Ceres in December, and sent his results to Piazzi. The new planet, which had been sighted on the first day of the year, was found again—by its discoverer—on the last day of the year.

Gauss did not publish his orbit determination methods until 1809.[3] In this publication, he also described the method of least squares that he had discovered in 1795, at the age of 18, and had used it in refining his estimates of the orbit of Ceres.

Although Ceres played a significant role in the history of discovery and it still reappears regularly in the nighttime sky, it has faded into obscurity as an object of intellectual interest. The method of least squares, on the other hand, has been an object of continuing interest and benefit to generations of scientists and technologists ever since its introduction. It has had a profound effect on the history of science. It was the first optimal estimation method, and it provided an important connection between the experimental and theoretical sciences: It gave experimentalists a practical method for estimating the unknown parameters of theoretical models.

1.2.2 Method of Least Squares

The following example of a least-squares problem is the one most often seen, although the *method* of least squares may be applied to a much greater range of problems.

EXAMPLE 1.1: Least-Squares Solution for Overdetermined Linear Systems Gauss discovered that if he wrote a system of equations in matrix form, as

$$\begin{bmatrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1n} \\ h_{21} & h_{22} & h_{23} & \cdots & h_{2n} \\ h_{31} & h_{32} & h_{33} & \cdots & h_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{l1} & h_{l2} & h_{l3} & \cdots & h_{ln} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_m \end{bmatrix} \tag{1.2}$$

or

$$Hx = z, \tag{1.3}$$

[3]In the meantime, the method of least squares had been discovered independently and published by Andrien-Marie Legendre (1752–1833) in France and Robert Adrian (1775–1855) in the United States [176]. [It had also been discovered and used before Gauss was born by the German-Swiss physicist Johann Heinrich Lambert (1728–1777).] Such Jungian synchronicity (i.e., the phenomenon of multiple, near-simultaneous discovery) was to be repeated for other breakthroughs in estimation theory, as well—for the Wiener filter and the Kalman filter.

then he could consider the problem of solving for that value of an *estimate* $\hat{x}$ (pronounced "x-hat") that minimizes the "estimated measurement error" $H\hat{x} - z$. He could characterize that estimation error in terms of its Euclidean vector norm $|H\hat{x} - z|$, or, equivalently, its square:

$$\varepsilon^2(\hat{x}) = |H\hat{x} - z|^2 \tag{1.4}$$

$$= \sum_{i=1}^{m} \left[\sum_{j=1}^{n} h_{ij}\hat{x}_j - z_i \right]^2, \tag{1.5}$$

which is a continuously differentiable function of the n unknowns $\hat{x}_1, \hat{x}_2, \hat{x}_3, \ldots, \hat{x}_n$. This function $\varepsilon^2(\hat{x}) \to \infty$ as any component $\hat{x}_k \to \pm\infty$. Consequently, it will achieve its minimum value where all its derivatives with respect to the $\hat{x}_k$ are zero. There are n such equations of the form

$$0 = \frac{\partial \varepsilon^2}{\partial \hat{x}_k} \tag{1.6}$$

$$= 2\sum_{i=1}^{m} h_{ik} \left[\sum_{j=1}^{n} h_{ij}\hat{x}_j - z_i \right] \tag{1.7}$$

for $k = 1, 2, 3, \ldots, n$. Note that in this last equation the expression

$$\sum_{j=1}^{n} h_{ij}\hat{x}_j - z_i = \{H\hat{x} - z\}_i, \tag{1.8}$$

the ith row of $H\hat{x} - z$, and the outermost summation is equivalent to the dot product of the kth column of H with $H\hat{x} - z$. Therefore Equation 1.7 can be written as

$$0 = 2H^{\mathrm{T}}[H\hat{x} - z] \tag{1.9}$$

$$= 2H^{\mathrm{T}}H\hat{x} - 2H^{\mathrm{T}}z \tag{1.10}$$

or

$$H^{\mathrm{T}}H\hat{x} = H^{\mathrm{T}}z,$$

where the matrix transpose H^{T} is defined as

$$H^{\mathrm{T}} = \begin{bmatrix} h_{11} & h_{21} & h_{31} & \cdots & h_{m1} \\ h_{12} & h_{22} & h_{32} & \cdots & h_{m2} \\ h_{13} & h_{23} & h_{33} & \cdots & h_{m3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{1n} & h_{2n} & h_{3n} & \cdots & h_{mn} \end{bmatrix} \tag{1.11}$$

The normal equation of the linear least squares problem. The equation

$$H^{\mathrm{T}}H\hat{x} = H^{\mathrm{T}}z \tag{1.12}$$

is called the *normal equation* or the *normal form* of the equation for the linear least-squares problem. It has precisely as many equivalent scalar equations as unknowns.

The Gramian of the linear least squares problem. The normal equation has the solution

$$\hat{x} = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}z,$$

provided that the matrix

$$\mathscr{G} = H^{\mathrm{T}}H \tag{1.13}$$

is *nonsingular* (i.e., invertible). The matrix product $\mathscr{G} = H^{\mathrm{T}}H$ in this equation is called the *Gramian matrix*.[4] The determinant of the Gramian matrix characterizes whether or not the column vectors of H are linearly independent. If its determinant is zero, the column vectors of H are linearly dependent, and $\hat{x}$ cannot be determined uniquely. If its determinant is nonzero, then the solution $\hat{x}$ is uniquely determined.

Least-squares solution. In the case that the Gramian matrix *is* invertible (i.e., nonsingular), the solution $\hat{x}$ is called the least-squares solution of the overdetermined linear inversion problem. It is an estimate that makes no assumptions about the nature of the unknown measurement errors, although Gauss alluded to that possibility in his description of the method. The formal treatment of uncertainty in estimation would come later.

This form of the Gramian matrix will be used in Chapter 2 to define the observability matrix of a linear dynamic system model in discrete time.

Least Squares in Continuous Time. The following example illustrates how the principle of least squares can be applied to fitting a vector-valued parametric model to data in continuous time. It also illustrates how the issue of *determinacy* (i.e., whether there is a *unique* solution to the problem) is characterized by the Gramian matrix in this context.

[4]Named for the Danish mathematician Jorgen Pedersen Gram (1850–1916). This matrix is also related to what is called the *unscaled Fisher information matrix*, named after the English statistician Ronald Aylmer Fisher (1890–1962). Although information matrices and Gramian matrices have different definitions and uses, they can amount to almost the same thing in this particular instance. The formal statistical definition of the term *information matrix* represents the information obtained from a sample of values from a known probability distribution. It corresponds to a scaled version of the Gramian matrix when the measurement errors in z have a joint Gaussian distribution, with the scaling related to the uncertainty of the measured data. The information matrix is a *quantitative* statistical characterization of the "information" (in some sense) that is in the data z used for estimating x. The Gramian, on the other hand, is used as an *qualitative* algebraic characterization of the uniqueness of the solution.

EXAMPLE 1.2: Least-Squares Fitting of Vector-Valued Data in Continuous Time Suppose that, for each value of time t on an interval $t_0 \le t \le t_f$, $z(t)$ is an ℓ-dimensional signal vector that is modeled as a function of an unknown n-vector x by the equation

$$z(t) = H(t)x,$$

where $H(t)$ is a known $\ell \times n$ matrix. The squared error in this relation at each time t will be

$$\begin{aligned}\varepsilon^2(t) &= |z(t) - H(t)x|^2 \\ &= x^{\mathrm{T}}[H^{\mathrm{T}}(t)H(t)]x - 2x^{\mathrm{T}}H^{\mathrm{T}}(t)z(t) + |z(t)|^2.\end{aligned}$$

The squared integrated error over the interval will then be the integral

$$\begin{aligned}\|\varepsilon\|^2 &= \int_{t_0}^{t_f} \varepsilon^2(t)\,dt \\ &= x^{\mathrm{T}}\left[\int_{t_0}^{t_f} H^{\mathrm{T}}(t)H(t)\,dt\right]x - 2x^{\mathrm{T}}\left[\int_{t_0}^{t_f} H^{\mathrm{T}}(t)z(t)\,dt\right] + \int_{t_0}^{t_f} |z(t)|^2\,dt,\end{aligned}$$

which has exactly the same array structure with respect to x as the algebraic least-squares problem. The least-squares solution for x can be found, as before, by taking the derivatives of $\|\varepsilon\|^2$ with respect to the components of x and equating them to zero. The resulting equations have the solution

$$\hat{x} = \left[\int_{t_0}^{t_f} H^{\mathrm{T}}(t)H(t)\,dt\right]^{-1}\left[\int_{t_0}^{t_f} H^{\mathrm{T}}(t)z(t)\,dt\right],$$

provided that the corresponding Gramian matrix

$$\mathscr{G} = \int_{t_0}^{t_f} H^{\mathrm{T}}(t)H(t)\,dt$$

is nonsingular.

This form of the Gramian matrix will be used in Chapter 2 to define the *observability matrix* of a linear dynamic system model in continuous time.

1.2.3 Gramian Matrix and Observability

For the examples considered above, observability does not depend upon the measurable data (z). It depends only on the nonsingularity of the Gramian matrix ($\mathscr{G}$), which depends only on the linear constraint matrix (H) between the unknowns and knowns.

Observability of a set of unknown variables is the issue of whether or not their values are *uniquely determinable* from a given set of *constraints,* expressed as equations involving functions of the unknown variables. The unknown variables are said to be *observable* if their values are uniquely determinable from the given constraints, and they are said to be *unobservable* if they are not uniquely determinable from the given constraints.

The condition of *nonsingularity* (or "*full rank*") of the Gramian matrix is an *algebraic* characterization of observability when the constraining equations are *linear* in the unknown variables. It also applies to the case that the constraining equations are not exact, due to errors in the values of the allegedly known parameters of the equations.

The Gramian matrix will be used in Chapter 2 to define observability of the states of dynamic systems in continuous time and discrete time.

1.2.4 Introduction of Probability Theory

Beginnings of Probability Theory. Probabilities represent the state of knowledge about physical phenomena by providing something more useful than "I don't know" to questions involving uncertainty. One of the mysteries in the history of science is why it took so long for mathematicians to formalize a subject of such practical importance. The Romans were selling insurance and annuities long before expectancy and risk were concepts of serious mathematical interest. Much later, the Italians were issuing insurance policies against business risks in the early Renaissance, and the first known attempts at a theory of probabilities—for games of chance—occurred in that period. The Italian Girolamo Cardano[5] (1501–1576) performed an accurate analysis of probabilities for games involving dice. He assumed that successive tosses of the dice were statistically independent events. He and the contemporary Indian writer Brahmagupta stated without proof that the accuracies of empirical statistics tend to improve with the number of trials. This would later be formalized as a *law of large numbers.*

More general treatments of probabilities were developed by Blaise Pascal (1623–1662), Pierre de Fermat (1601–1655), and Christiaan Huygens (1629–1695). Fermat's work on combinations was taken up by Jakob (or James) Bernoulli (1654–1705), who is considered by some historians to be the founder of probability theory. He gave the first rigorous proof of the law of large numbers for repeated independent trials (now called *Bernoulli trials*). Thomas Bayes (1702–1761) derived his famous rule for statistical inference sometime after Bernoulli. Abraham de Moivre (1667–1754), Pierre Simon Marquis de Laplace (1749–1827), Adrien Marie Legendre (1752–1833), and Carl Friedrich Gauss (1777–1855) continued this development into the nineteenth century.

[5]Cardano was a practicing physician in Milan who also wrote books on mathematics. His book *De Ludo Hleae,* on the mathematical analysis of games of chance (principally dice games), was published nearly a century after his death. Cardano was also the inventor of the most common type of universal joint found in automobiles, sometimes called the *Cardan joint* or *Cardan shaft.*

Between the early nineteenth century and the mid-twentieth century, the probabilities themselves began to take on more meaning as physically significant attributes. The idea that the laws of nature embrace random phenomena, and that these are treatable by probabilistic models began to emerge in the nineteenth century. The development and application of probabilistic models for the physical world expanded rapidly in that period. It even became an important part of sociology. The work of James Clerk Maxwell (1831–1879) in statistical mechanics established the probabilistic treatment of natural phenomena as a scientific (and successful) discipline.

An important figure in probability theory and the theory of random processes in the twentieth century was the Russian academician Andrei Nikolaeovich Kolmogorov (1903–1987). Starting around 1925, working with H. Ya. Khinchin and others, he reestablished the foundations of probability theory on measurement theory, which became the accepted mathematical basis of probability and random processes. Along with Norbert Wiener (1894–1964), he is credited with founding much of the theory of prediction, smoothing and filtering of Markov processes, and the general theory of ergodic processes. His was the first formal theory of optimal estimation for systems involving random processes.

1.2.5 Wiener Filter

Norbert Wiener (1894–1964) is one of the more famous prodigies of the early twentieth century. He was taught by his father until the age of 9, when he entered high school. He finished high school at the age of 11 and completed his undergraduate degree in mathematics in three years at Tufts University. He then entered graduate school at Harvard University at the age of 14 and completed his doctorate degree in the philosophy of mathematics when he was 18. He studied abroad and tried his hand at several jobs for six more years. Then, in 1919, he obtained a teaching appointment at the Massachusetts Institute of Technology (MIT). He remained on the faculty at MIT for the rest of his life.

In the popular scientific press, Wiener is probably more famous for naming and promoting *cybernetics* than for developing the Wiener filter. Some of his greatest mathematical achievements were in generalized harmonic analysis, in which he extended the Fourier transform to functions of finite *power.* Previous results were restricted to functions of finite *energy,* which is an unreasonable constraint for signals on the real line. Another of his many achievements involving the generalized Fourier transform was proving that the transform of white noise is also white noise.[6]

Wiener Filter Development. In the early years of the World War II, Wiener was involved in a military project to design an automatic controller for directing antiaircraft fire with radar information. Because the speed of the airplane is a

[6]He is also credited with the discovery that the power spectral density of a signal equals the Fourier transform of its autocorrelation function, although it was later discovered that Einstein had known it before him.

nonnegligible fraction of the speed of bullets, this system was required to "shoot into the future." That is, the controller had to predict the future course of its target using noisy radar tracking data.

Wiener derived the solution for the least-mean-squared prediction error in terms of the autocorrelation functions of the signal and the noise. The solution is in the form of an integral operator that can be synthesized with analog circuits, given certain constraints on the regularity of the autocorrelation functions or, equivalently, their Fourier transforms. His approach represents the probabilistic nature of random phenomena in terms of power spectral densities.

An analogous derivation of the optimal linear predictor for discrete-time systems was published by A. N. Kolmogorov in 1941, when Wiener was just completing his work on the continuous-time predictor.

Wiener's work was not declassified until the late 1940s, in a report titled "Extrapolation, interpolation, and smoothing of stationary time series." The title was subsequently shortened to "Time series." An early edition of the report had a yellow cover, and it came to be called "the yellow peril." It was loaded with mathematical details beyond the grasp of most engineering undergraduates, but it was absorbed and used by a generation of dedicated graduate students in electrical engineering.

1.2.6 Kalman Filter

Rudolf Emil Kalman was born on May 19, 1930, in Budapest, the son of Otto and Ursula Kalman. The family emigrated from Hungary to the United States during World War II. In 1943, when the war in the Mediterranean was essentially over, they traveled through Turkey and Africa on an exodus that eventually brought them to Youngstown, Ohio, in 1944. Rudolf attended Youngstown College there for three years before entering MIT.

Kalman received his bachelor's and master's degrees in electrical engineering at MIT in 1953 and 1954, respectively. His graduate advisor was Ernst Adolph Guillemin, and his thesis topic was the behavior of solutions of second-order difference equations [114]. When he undertook the investigation, it was suspected that second-order difference equations might be modeled by something analogous to the describing functions used for second-order differential equations. Kalman discovered that their solutions were not at all like the solutions of differential equations. In fact, they were found to exhibit chaotic behavior.

In the fall of 1955, after a year building a large analog control system for the E. I. DuPont Company, Kalman obtained an appointment as lecturer and graduate student at Columbia University. At that time, Columbia was well known for the work in control theory by John R. Ragazzini, Lotfi A. Zadeh,[7] and others. Kalman taught at Columbia until he completed the Doctor of Science degree there in 1957.

For the next year, Kalman worked at the research laboratory of the International Business Machines Corporation in Poughkeepsie and for six years after that at the

[7]Zadeh is perhaps more famous as the "father" of fuzzy systems theory and interpolative reasoning.

research center of the Glenn L. Martin company in Baltimore, the Research Institute for Advanced Studies (RIAS).

Early Research Interests. The algebraic nature of systems theory first became of interest to Kalman in 1953, when he read a paper by Ragazzini published the previous year. It was on the subject of sampled-data systems, for which the time variable is discrete valued. When Kalman realized that linear discrete-time systems could be solved by transform methods, just like linear continuous-time systems, the idea occurred to him that there is no fundamental difference between continuous and discrete linear systems. The two must be equivalent in some sense, even though the solutions of linear differential equations cannot go to zero (and stay there) in finite time and those of discrete-time systems can. That started his interest in the connections between systems theory and algebra.

In 1954 Kalman began studying the issue of *controllability*, which is the question of whether there exists an input (control) function to a dynamic system that will drive the state of that system to zero. He was encouraged and aided by the work of Robert W. Bass during this period. The issue of eventual interest to Kalman was whether there is an *algebraic* condition for controllability. That condition was eventually found as the rank of a matrix.[8] This implied a connection between algebra and systems theory.

Discovery of the Kalman Filter. In late November of 1958, not long after coming to RIAS, Kalman was returning by train to Baltimore from a visit to Princeton. At around 11 PM, the train was halted for about an hour just outside Baltimore. It was late, he was tired, and he had a headache. While he was trapped there on the train for that hour, an idea occurred to him: *Why not apply the notion of state variables*[9] *to the Wiener filtering problem?* He was too tired to think much more about it that evening, but it marked the beginning of a great exercise to do just that. He read through Loève's book on probability theory [68] and equated expectation with projection. That proved to be pivotal in the derivation of the Kalman filter. With the additional assumption of finite dimensionality, he was able to derive the Wiener filter as what we now call the Kalman filter. With the change to state-space form, the mathematical background needed for the derivation became much simpler, and the proofs were within the mathematical reach of many undergraduates.

Introduction of the Kalman Filter. Kalman presented his new results in talks at several universities and research laboratories before it appeared in print.[10] His ideas were met with some skepticism among his peers, and he chose a mechanical

[8]The *controllability matrix,* a concept defined in Chapter 2.

[9]Although function-space methods were then the preferred approach to the filtering problem, the use of state-space models for time-varying systems had already been introduced (e.g., by Laning and Battin [67] in 1956).

[10]In the meantime, some of the seminal ideas in the Kalman filter had been published by Swerling [227] in 1959 and Stratonovich [25, 226] in 1960.

engineering journal (rather than an electrical engineering journal) for publication, because "When you fear stepping on hallowed ground with entrenched interests, it is best to go sideways." [11] His second paper, on the continuous-time case, was once rejected because—as one referee put it—one step in the proof "cannot possibly be true." (It was true.) He persisted in presenting his filter, and there was more immediate acceptance elsewhere. It soon became the basis for research topics at many universities and the subject of dozens of doctoral theses in electrical engineering over the next several years.

Early Applications. Kalman found a receptive audience for his filter in the fall of 1960 in a visit to Stanley F. Schmidt at the Ames Research Center of NASA in Mountain View, California [118]. Kalman described his recent result and Schmidt recognized its potential applicability to a problem then being studied at Ames—the trajectory estimation and control problem for the Apollo project, a planned manned mission to the moon and back. Schmidt began work immediately on what was probably the first full implementation of the Kalman filter. He soon discovered what is now called "extended Kalman filtering," which has been used ever since for most real-time nonlinear applications of Kalman filtering. Enthused over his own success with the Kalman filter, he set about proselytizing others involved in similar work. In the early part of 1961, Schmidt described his results to Richard H. Battin from the MIT Instrumentation Laboratory (later renamed the Charles Stark Draper Laboratory). Battin was already using state space methods for the design and implementation of astronautical guidance systems, and he made the Kalman filter part of the Apollo onboard guidance,[12] which was designed and developed at the Instrumentation Laboratory. In the mid-1960s, through the influence of Schmidt, the Kalman filter became part of the Northrup-built navigation system for the C5A air transport, then being designed by Lockheed Aircraft Company. The Kalman filter solved the *data fusion problem* associated with combining radar data with inertial sensor data to arrive at an overall estimate of the aircraft trajectory and the *data rejection problem* associated with detecting exogenous errors in measurement data. It has been an integral part of nearly every onboard trajectory estimation and control system designed since that time.

Other Research Interests. Around 1960, Kalman showed that the related notion of observability for dynamic systems had an algebraic dual relationship with controllability. That is, by the proper exchange of system parameters, one problem could be transformed into the other, and vice versa.

Richard S. Bucy was also at RIAS in that period, and it was he who suggested to Kalman that the Wiener–Hopf equation is equivalent to the matrix Riccati equa-

[11]The two quoted segments in this paragraph are from a talk on *System Theory: Past and Present* given by Kalman at the University of California at Los Angeles (UCLA) on April 17, 1991, in a symposium organized and hosted by A. V. Balakrishnan at UCLA and sponsored jointly by UCLA and the National Aeronautics and Space Administration (NASA) Dryden Laboratory.

[12]Another fundamental improvement in Kalman filter implementation methods was made soon after by James E. Potter at the MIT Instrumentation Laboratory. This will be discussed in the next subsection.

tion—if one assumes a finite-dimensional state-space model. The general nature of this relationship between integral equations and differential equations first became apparent around that time. One of the more remarkable achievements of Kalman and Bucy in that period was proving that the Riccati equation can have a stable (steady-state) solution even if the dynamic system is unstable—provided that the system is observable and controllable.

Kalman also played a leading role in the development of realization theory, which also began to take shape around 1962. This theory addresses the problem of finding a system model to explain the observed input–output behavior of a system. This line of investigation led to a *uniqueness principle* for the mapping of exact (i.e., noiseless) data to linear system models.

In 1985, Kalman was awarded the Kyoto Prize, considered by some to be the Japanese equivalent of the Nobel Prize. On his visit to Japan to accept the Kyoto Prize, he related to the press an epigram that he had first seen in a pub in Colorado Springs in 1962, and it had made an impression on him. It said:

Little people discuss other people.
Average people discuss events.
Big people discuss ideas.

His own work, he felt, had been concerned with ideas.

In 1990, on the occasion of Kalman's sixtieth birthday, a special international symposium was convened for the purpose of honoring his pioneering achievements in what has come to be called *mathematical system theory,* and a *Festschrift* with that title was published soon after [3].

Impact of Kalman Filtering on Technology. From the standpoint of those involved in estimation and control problems, at least, this has to be considered the greatest achievement in estimation theory of the twentieth century. Many of the achievements since its introduction would not have been possible without it. It was one of the enabling technologies for the Space Age, in particular. The precise and efficient navigation of spacecraft through the solar system could not have been done without it.

The principal uses of Kalman filtering have been in "modern" control systems, in the tracking and navigation of all sorts of vehicles, and in predictive design of estimation and control systems. These technical activities were made possible by the introduction of the Kalman filter. (If you need a demonstration of its impact on technology, enter the keyword "Kalman filter" in a technical literature search. You will be overwhelmed by the sheer number of references it will generate.)

Relative Advantages of Kalman and Wiener Filtering

1. The Wiener filter implementation in analog electronics can operate at much higher effective throughput than the (digital) Kalman filter.
2. The Kalman filter is implementable in the form of an algorithm for a digital computer, which was replacing analog circuitry for estimation and control at

the time that the Kalman filter was introduced. This implementation may be slower, but it is capable of much greater accuracy than had been achievable with analog filters.

3. The Wiener filter does not require finite-dimensional stochastic process models for the signal and noise.
4. The Kalman filter does not require that the deterministic dynamics or the random processes have stationary properties, and many applications of importance include nonstationary stochastic processes.
5. The Kalman filter is compatible with the state-space formulation of optimal controllers for dynamic systems, and Kalman was able to prove useful dual properties of estimation and control for these systems.
6. For the modern controls engineering student, the Kalman filter requires less additional mathematical preparation to learn and use than the Wiener filter. As a result, the Kalman filter can be taught at the undergraduate level in engineering curricula.
7. The Kalman filter provides the necessary information for mathematically sound, statistically-based decision methods for detecting and rejecting anomalous measurements.

1.2.7 Square-Root Methods and All That

Numerical Stability Problems. The great success of Kalman filtering was not without its problems, not the least of which was marginal stability of the numerical solution of the associated Riccati equation. In some applications, small roundoff errors tended to accumulate and eventually degrade the performance of the filter. In the decades immediately following the introduction of the Kalman filter, there appeared several better numerical implementations of the original formulas. Many of these were adaptations of methods previously derived for the least squares problem.

Early ad hoc Fixes. It was discovered early on[13] that forcing symmetry on the solution of the matrix Riccati equation improved its apparent numerical stability—a phenomenon that was later given a more theoretical basis by Verhaegen and Van Dooren [232]. It was also found that the influence of roundoff errors could be ameliorated by artificially increasing the covariance of process noise in the Riccati equation. A symmetrized form of the discrete-time Riccati equation was developed by Joseph [15] and used by R. C. K. Lee at Honeywell in 1964. This "structural" reformulation of the Kalman filter equations improved robustness against roundoff errors in some applications, although later methods have performed better on some problems [125].

[13]These fixes were apparently discovered independently by several people. Schmidt [118] and his colleagues at NASA had discovered the use of forced symmetry and "pseudonoise" to counter roundoff effects and cite R. C. K. Lee at Honeywell with the independent discovery of the symmetry effect.

Square-Root Filtering. These methods can also be considered as "structural" reformulations of the Riccati equation, and they predate the Bucy–Joseph form. The first of these was the "square-root" implementation by Potter and Stern [208], first published in 1963 and successfully implemented for space navigation on the Apollo manned lunar exploration program. Potter and Stern introduced the idea of factoring the covariance matrix into *Cholesky factors*,[14] in the format

$$P = CC^{\mathrm{T}}, \tag{1.14}$$

and expressing the observational update equations in terms of the Cholesky factor C, rather than P. The result was better numerical stability of the filter implementation at the expense of added computational complexity. A generalization of the Potter and Stern method to handle vector-valued measurements was published by one of the authors [130] in 1968, but a more efficient implementation—in terms of *triangular* Cholesky factors—was published by Bennet in 1967 [138].

Square-Root and UD Filters. There was a rather rapid development of faster algorithmic methods for square-root filtering in the 1970s, following the work at NASA/JPL (then called the Jet Propulsion Laboratory, at the California Institute of Technology) in the late 1960s by Dyer and McReynolds [156] on temporal update methods for Cholesky factors. Extensions of square-root covariance and information filters were introduced in Kaminski's 1971 thesis [115] at Stanford University. The first of the triangular factoring algorithms for the observational update was due to Agee and Turner [106], in a 1972 report of rather limited circulation. These algorithms have roughly the same computational complexity as the conventional Kalman filter, but with better numerical stability. The "fast triangular" algorithm of Carlson was published in 1973 [149], followed by the "square-root-free" algorithm of Bierman in 1974 [7] and the associated temporal update method introduced by Thornton [124]. The computational complexity of the square-root filter for time-invariant systems was greatly simplified by Morf and Kailath [204] soon after that. Specialized parallel processing architectures for fast solution of the square-root filter equations were developed by Jover and Kailath [175] and others over the next decade, and much simpler derivations of these and earlier square-root implementations were discovered by Kailath [26].

Factorization Methods. The square-root methods make use of matrix decomposition[15] methods that were originally derived for the least-squares problem. These

[14]A square root S of a matrix P satisfies the equation $P = SS$ (i.e., without the transpose on the second factor). Potter and Stern's derivation used a special type of symmetric matrix called an *elementary matrix.* They factored an elementary matrix as a square of another elementary matrix. In this case, the factors were truly square roots of the factored matrix. This square-root appellation has stuck with extensions of Potter and Stern's approach, even though the factors involved are Cholesky factors, not matrix square roots.

[15]The term "decomposition" refers to the representation of a matrix (in this case, a covariance matrix) as a product of matrices having more useful computational properties, such as sparseness (for triangular factors) or good numerical stability (for orthogonal factors). The term "factorization" was used by Bierman [7] for such representations.

include the so-called *QR* decomposition of a matrix as the product of an orthogonal matrix (Q) and a "triangular"[16] matrix (R). The matrix R results from the application of orthogonal transformations of the original matrix. These orthogonal transformations tend to be well conditioned numerically. The operation of applying these transformations is called the "triangularization" of the original matrix, and triangularization methods derived by Givens [164], Householder [172], and Gentleman [163] are used to make Kalman filtering more robust against roundoff errors.

1.2.8 Beyond Kalman Filtering

Extended Kalman Filtering and the Kalman–Schmidt Filter. Although it was originally derived for a linear problem, the Kalman filter is habitually applied with impunity—and considerable success—to many nonlinear problems. These extensions generally use partial derivatives as linear approximations of nonlinear relations. Schmidt [118] introduced the idea of evaluating these partial derivatives at the *estimated* value of the state variables. This approach is generally called the *extended Kalman filter*, but it was called the *Kalman–Schmidt filter* in some early publications. This and other methods for approximate linear solutions to nonlinear problems are discussed in Chapter 5, where it is noted that these will not be adequate for all nonlinear problems. Mentioned here are some investigations that have addressed estimation problems from a more general perspective, although they are not covered in the rest of the book.

Nonlinear Filtering Using Higher Order Approximations. Approaches using higher order expansions of the filter equations (i.e., beyond the linear terms) have been derived by Stratonovich [78], Kushner [191], Bucy [147], Bass et al. [134], and others for quadratic nonlinearities and by Wiberg and Campbell [235] for terms through third order.

Nonlinear Stochastic Differential Equations. Problems involving nonlinear and random dynamic systems have been studied for some time in statistical mechanics. The propagation over time of the *probability distribution* of the state of a nonlinear dynamic system is described by a nonlinear partial differential equation called the *Fokker–Planck equation.* It has been studied by Einstein [157], Fokker [160], Planck [207], Kolmogorov [187], Stratonovich [78], Baras and Mirelli [52], and others. Stratonovich modeled the effect on the probability distribution of information obtained through noisy measurements of the dynamic system, an effect called *conditioning.* The partial differential equation that includes these effects is called the *conditioned Fokker–Planck equation.* It has also been studied by Kushner [191], Bucy [147], and others using the *stochastic calculus* of Kiyosi Itô—also called the "Itô calculus." It is a non-Riemannian calculus developed specifically for stochastic differential systems with noise of infinite bandwidth. This general approach results in a stochastic partial differential equation describing

[16]See Chapter 6 and Appendix B for discussions of triangular forms.

the evolution over time of the probability distribution over a "state space" of the dynamic system under study. The resulting model does *not* enjoy the finite representational characteristics of the Kalman filter, however. The computational complexity of obtaining a solution far exceeds the already considerable burden of the conventional Kalman filter. These methods are of significant interest and utility but are beyond the scope of this book.

Point Processes and the Detection Problem. A *point process* is a type of random process for modeling events or objects that are distributed over time or space, such as the arrivals of messages at a communications switching center or the locations of stars in the sky. It is also a model for the initial states of systems in many estimation problems, such as the locations of aircraft or spacecraft under surveillance by a radar installation or the locations of submarines in the ocean. The *detection problem* for these surveillance applications must usually be solved before the *estimation problem* (i.e., tracking of the objects with a Kalman filter) can begin. The Kalman filter requires an initial state for each object, and that initial state estimate must be obtained by detecting it. Those initial states are distributed according to some point process, but there are no technically mature methods (comparable to the Kalman filter) for estimating the state of a point process. A unified approach combining detection and tracking into one optimal estimation method was derived by Richardson [214] and specialized to several applications. The detection and tracking problem for a *single object* is represented by the conditioned Fokker–Planck equation. Richardson derived from this one-object model an infinite hierarchy of partial differential equations representing *object densities* and truncated this hierarchy with a simple closure assumption about the relationships between orders of densities. The result is a single partial differential equation approximating the evolution of the density of objects. It can be solved numerically. It provides a solution to the difficult problem of detecting dynamic objects whose initial states are represented by a point process.

1.3 ON THE NOTATION USED IN THIS BOOK

1.3.1 Symbolic Notation

The fundamental problem of symbolic notation, in almost any context, is that there are never enough symbols to go around. There are not enough letters in the Roman alphabet to represent the sounds of standard English, let alone all the variables in Kalman filtering and its applications. As a result, some symbols must play multiple roles. In such cases, their roles will be defined as they are introduced. It is sometimes confusing, but unavoidable.

"Dot" Notation for Derivatives. Newton's notation using $\dot{f}(t)$, $\ddot{f}(t)$ for the first two derivatives of f with respect to t is used where convenient to save ink.

TABLE 1.2 Standard Symbols of Kalman Filtering

Symbols			Symbol Definition
I[a]	II[b]	III[c]	
F	F	A	Dynamic coefficient matrix of continuous linear differential equation defining dynamic system
G	I	B	Coupling matrix between random process noise and state of linear dynamic system
H	M	C	Measurement sensitivity matrix, defining linear relationship between state of the dynamic system and measurements that can be made
$\overline{K}$	Δ	K	Kalman gain matrix
P	P		Covariance matrix of state estimation uncertainty
Q	Q		Covariance matrix of process noise in the system state dynamics
R	0		Covariance matrix of observational (measurement) uncertainty
x	x		State vector of a linear dynamic system
z	y		Vector (or scalar) of measured values
Φ	Φ		State transition matrix of a discrete linear dynamic system

[a] This book [1, 13, 16, 21]. [b] Kalman [23, 179]. [c] Other sources [4, 10, 18, 65].

Standard Symbols for Kalman Filter Variables. There appear to be two "standard" conventions in technical publications for the symbols used in Kalman filtering. The one used in this book is similar to the original notation of Kalman [179]. The other standard notation is sometimes associated with applications of Kalman filtering in control theory. It uses the first few letters of the alphabet in place of the Kalman notation. Both sets of symbol usages are presented in Table 1.2, along with the original (Kalman) notation.

State Vector Notation for Kalman Filtering. The state vector x has been adorned with all sorts of other appendages in the usage of Kalman filtering. Table 1.3 lists the notation used in this book (left column) along with notations found in some other sources (second column). The state vector wears a "hat" as the estimated value, $\hat{x}$, and subscripting to denote the sequence of values that the estimate assumes over time. The problem is that it has two values at the same time: the *a priori*[17] value (before the measurement at the current time has been used in refining the estimate) and the *a posteriori* value (after the current measurement has been used in refining the estimate). These distinctions are indicated by the signum. The negative sign $(-)$ indicates the a priori value, and the positive sign $(+)$ indicates the *a posteriori* value.

[17]This use of the full Latin phrases as adjectives for the prior and posterior statistics is an unfortunate choice of standard notation, because there is no easy way to shorten it. (Even their initial abbreviations are the same.) If those who initiated this notation had known how commonplace it would become, they might have named them otherwise.

TABLE 1.3 Special State-Space Notation

This book	Other sources	Definition of Notational Usage
x	$\underline{x}$ $\vec{x}$ $\mathbf{x}$	Vector
x_k		The kth component of the vector x
x_k	$x[k]$	The kth element of the sequence $\ldots, x_{k-1}, x_k, x_{k+1}, \ldots$ of vectors
$\hat{x}$	$E\langle x \rangle$ $\bar{x}$	An estimate of the value of x
$\hat{x}_k(-)$	$\hat{x}_{k\|k-1}$ $\hat{x}_{k-}$	A priori estimate of x_k, conditioned on all prior measurements except the one at time t_k
$\hat{x}_k(+)$	$\hat{x}_{k\|k}$ $\hat{x}_{k+}$	A posteriori estimate of x, conditioned on all available measurements at time t_k
$\dot{x}$	x_t dx/dt	Derivative of x with respect to t (time)

TABLE 1.4 Common Notation for Array Dimensions

					Dimensions	
Symbol	Vector Name	Dimensions	Symbol	Matrix Name	Row	Column
x	System state	n	Φ	State transition	n	n
w	Process noise	r	G	Process noise coupling	n	r
u	Control input	s	Q	Process noise covariance	r	r
z	Measurement	ℓ	H	Measurement sensitivity	ℓ	n
v	Measurement noise	ℓ	R	Measurement noise covariance	ℓ	ℓ

Common Notation for Array Dimensions. Symbols used for the *dimensions* of the "standard" arrays in Kalman filtering will also be standardized, using the notation of Gelb et al. [21] shown in Table 1.4. These symbols are not used exclusively for these purposes. (Otherwise, one would soon run out of alphabet.) However, whenever one of these arrays is used in the discussion, these symbols will be used for their dimensions.

1.4 SUMMARY

The *Kalman filter* is an estimator used to estimate the *state* of a *linear dynamic system* perturbed by *Gaussian white noise* using measurements that are *linear* functions of the system state but corrupted by *additive* Gaussian white noise. The mathematical model used in the derivation of the Kalman filter is a reasonable representation for many problems of practical interest, including *control problems* as

well as *estimation problems.* The Kalman filter model is also used for the *analysis* of *measurement and estimation problems.*

The *method of least squares* was the first "optimal" estimation method. It was discovered by Gauss (and others) around the end of the eighteenth century, and it is still much in use today. If the associated *Gramian matrix* is *nonsingular*, the method of least squares determines the *unique* values of a set of unknown variables such that the *squared deviation* from a set of constraining equations is minimized.

Observability of a set of unknown variables is the issue of whether or not they are *uniquely determinable* from a given set of *constraining equations.* If the constraints are *linear* functions of the unknown variables, then those variables are *observable if and only if* the associated Gramian matrix is nonsingular. If the Gramian matrix is *singular,* then the unknown variables are *unobservable*.

The *Wiener–Kolmogorov filter* was derived in the 1940s by Norbert Wiener (using a model in continuous time) and Andrei Kolmogorov (using a model in discrete time) working independently. It is a *statistical estimation method.* It estimates the state of a dynamic process so as to minimize the *mean-squared estimation error.* It can take advantage of *statistical* knowledge about random processes in terms of their power spectral densities in the *frequency domain*.

The "*state-space" model* of a dynamic process uses differential equations (or difference equations) to represent both deterministic and random phenomena. The *state variables* of this model are the variables of interest and their derivatives of interest. Random processes are characterized in terms of their statistical properties in the *time domain,* rather than the frequency domain. The Kalman filter was derived as the solution to the Wiener filtering problem using the state-space model for dynamic and random processes. The result is easier to derive (and to use) than the Wiener–Kolmogorov filter.

Square-root filtering is a reformulation of the Kalman filter for better numerical stability in finite-precision arithmetic. It is based on the same mathematical model, but it uses an equivalent statistical parameter that is less sensitive to roundoff errors in the computation of optimal filter gains. It incorporates many of the more numerically stable computation methods that were originally derived for solving the least-squares problem.

PROBLEMS

1.1 Jean Baptiste Fourier (1768–1830) was studying the problem of approximating a function $f(\theta)$ on the circle $0 \leq \theta < 2\pi$ by a linear combination of trigonometric functions:

$$f(\theta) \approx a_0 + \sum_{j=1}^{n} [a_j \cos(j\theta) + b_j \sin(j\theta)]. \tag{1.15}$$

See if you can help him on this problem. Use the method of least squares to demonstrate that the values

$$\hat{a}_0 = \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\, d\theta,$$
$$\hat{a}_j = \frac{1}{\pi}\int_0^{2\pi} f(\theta)\cos(j\theta)\, d\theta,$$
$$\hat{b}_j = \frac{1}{\pi}\int_0^{2\pi} f(\theta)\sin(j\theta)\, d\theta$$

of the coefficients a_j and b_j for $1 \le j \le n$ give the least integrated squared approximation error

$$\begin{aligned}
\varepsilon^2(a,b) &= \|f - \hat{f}(a,b)\|_{\mathscr{L}_2}^2 \\
&= \int_0^{2\pi}\left[\hat{f}(\theta) - f(\theta)\right]^2 d\theta \\
&= \int_0^{2\pi}\left\{a_0 + \sum_{j=1}^{n}[a_j\cos(j\theta) + b_j\sin(j\theta)]\right\}^2 d\theta \\
&\quad - 2\int_0^{2\pi}\left\{a_0 + \sum_{j=1}^{n} t[a_j\cos(j\theta) + b_j\sin(j\theta)]\right\} f(\theta)\, d\theta \\
&\quad + \int_0^{2\pi} f^2(\theta)\, d\theta.
\end{aligned}$$

You may assume the equalities

$$\int_0^{2\pi} d\theta = 2\pi$$
$$\int_0^{2\pi}\cos(j\theta)\cos(k\theta)\, d\theta = \begin{cases} 0, & j \ne k \\ \pi, & j = k, \end{cases}$$
$$\int_0^{2\pi}\sin(j\theta)\sin(k\theta)\, d\theta = \begin{cases} 0, & j \ne k \\ \pi, & j = k \end{cases}$$
$$\int_0^{2\pi}\cos(j\theta)\sin(k\theta)\, d\theta = 0, \qquad 0 \le j \le n, \qquad 1 \le k \le n$$

as given.