1

INTRODUCTION

Once you get the physics right, the rest is mathematics.

—Rudolf E. Kalman
Kailath Lecture, Stanford University, May 11, 2009

1.1 CHAPTER FOCUS

This chapter presents a preview of where we are heading, some history of how others got there before us, an overview showing how all the material fits together, and a common notation and nomenclature to make it more apparent.

1.2 ON KALMAN FILTERING

1.2.1 First of All: What Is a Kalman Filter?

Theoretically, it has been called the *linear least mean squares estimator* (LLSME) because it minimizes the mean-squared estimation error for a linear stochastic system using noisy linear sensors. It has also been called the *linear quadratic estimator* (LQE) because it minimizes a quadratic function of estimation error for a linear dynamic system with white measurement and disturbance noise. Even today, more than half a century after its discovery, it remains a unique accomplishment in

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

the history of estimation theory. It is the only practical finite-dimensional solution to the real-time optimal estimation problem for stochastic systems, and it makes very few assumptions about the underlying probability distributions except that they have finite means and second central moments (covariances). Its mathematical model has been found to represent a phenomenal range of important applications involving noisy measurements for estimating the current conditions of dynamic systems with less-than-predictable disturbances. Although many approximation methods have been developed to extend its application to less-than-linear problems, and despite decades of dedicated research directed at generalizing it for nonlinear applications, no comparable general solution[1] for nonlinear problems has been found.

Practically, the Kalman filter is one of the great discoveries of *mathematical engineering*, which uses mathematical modeling to solve engineering problems—in the much same way that mathematical physics is used to solve physics problems, or computational mathematics is used for solving efficiency and accuracy problems in computer implementations.

Its early users would come to consider the Kalman filter to be the greatest discovery in practical estimation theory in the twentieth century, and its reputation has continued to grow over time. As an indication of its ubiquity, a *Google*®; web search for "Kalman filter" or "Kalman filtering" produces more than a million hits. One reason for this is that the Kalman filter has enabled human kind 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 monitoring and 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 the mathematical framework for inferring the unmeasured variables 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 and securities. It has become a universal tool for integrating different sensor and/or data collection systems into an overall optimal solution.

As an added bonus, the Kalman filter model can be used as a tool for assessing the relative accuracy of alternative sensor system designs for likely scenarios of dynamic system trajectories. Without this capability, development of many complex sensor systems (including Global Navigation Satellite Systems) may not have been possible.

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

1. *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. Mathematical tools make mental work more efficient, just as mechanical tools make physical work more efficient. As with any tool, it is important to

[1]However, a somewhat limited finite-dimensional nonlinear solution has been found [1].

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.

2. *It is a computer program*. It has been called "ideally suited to digital computer implementation" [2], 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.
3. *It is a consistent statistical characterization of an estimation problem*. It is much more than an *estimator*, because it propagates the current *state of knowledge* of the dynamic system, including the mean-squared uncertainties arising from random dynamic perturbations and sensor noise. These properties are extremely useful for statistical analysis and the predictive design of sensor systems.

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 effectively, then please read on!

1.2.2 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 and was used to denote the material that was used as a filter for liquids.) Originally, a filter solved the problem of separating unwanted components of liquid–solid mixtures. In the era of crystal radios and vacuum tubes, the term was applied to analog circuits that "filter" electronic signals. These 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.

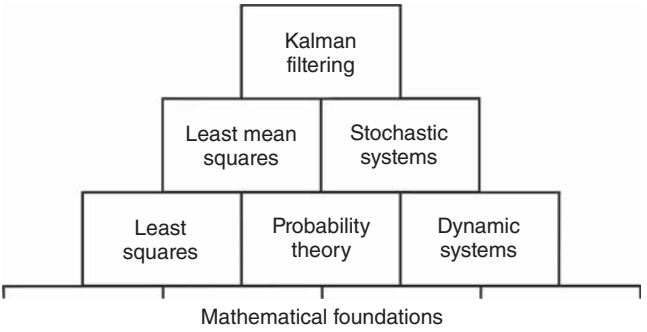

Figure 1.1 Foundational concepts in Kalman filtering.

1.2.3 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[2] (matrix theory, in Appendix B on the Wiley web site).

1.2.4 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. *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 dwarf planet 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 expressing 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.

[2]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.

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 GNSS[a] receiver

[a]*Abbreviation*: GNSS, Global Navigation Satellite System.

2. *Performance 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 performance criteria. These criteria are typically related to estimation accuracy and system cost.

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

1. the types of sensors to be used,
2. the locations and orientations of the various sensor types with respect to the system to be estimated,
3. the allowable noise characteristics of the sensors,
4. the prefiltering methods for smoothing sensor noise,
5. the data sampling rates for the various sensor types, and
6. 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.

1.3 ON OPTIMAL ESTIMATION METHODS

The Kalman filter is the result of an evolutionary process of ideas from many creative thinkers over many centuries. We present here some of the seminal ideas in this process, the discoverers of which are listed in historical perspective in Figure 1.2. This list is by no means exhaustive. There are far too many people involved to show them all, but the figure should give some idea of the time periods involved. The figure covers only half a millennium, and the study and development of mathematical concepts goes back beyond history. Readers interested in more detailed histories of optimal estimation are referred to the survey articles by Kailath [8, 30], Lainiotis [3], Mendel and Gieseking [4], and Sorenson [55, 56] and the personal accounts of Battin [5] and Schmidt [6]. More recent contributions from the last five discoverers on this list are discussed in Chapters 7 and 8.

1.3.1 Beginnings of Optimal 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 (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 an interesting event in the history of astronomy. The following

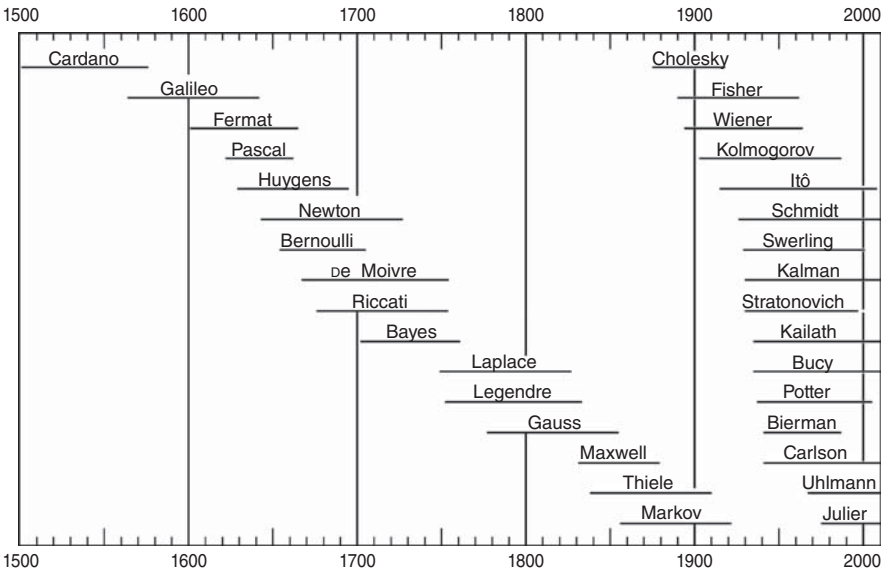

Figure 1.2 Lifelines of some important contributors to estimation technology.

account was put together from several sources, including the account by Baker and Makemson [7].

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, it included an error by the printer. While searching for the "missing" star, Piazzi discovered, instead, something that moved. It was the "dwarf planet" *Ceres*—the largest body in the asteroid belt 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 before it moved too close to the sun and disappeared.

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

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

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 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 sound philosophical justification for there being only seven. The new celestial object became a subject of conversation in intellectual circles nearly everywhere. Fortunately, the problem caught the attention of a 24-year-old mathematician at Góttingen 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" (later reclassified as an asteroid), 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 had faded into obscurity as an object of intellectual interest until the 2007 launch of scientific probe Dawn for a 2015 rendezvous with Ceres. 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.

1.3.2 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

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

or

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

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:

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

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

[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 [8]. 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–Kolmogorov filter and the Kalman filter.

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 are 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

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

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.

1.3.2.1 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

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

is *nonsingular* (i.e., invertible). The matrix product $\mathcal{G} = H^{\mathrm{T}}H$ 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.

Example 1.2 (The Gramians of Guier and Weiffenbach) Development of satellite navigation started just after the world's first artificial satellite, Sputnik I, was launched from the Soviet Union on Friday, October 4, 1957. On the following Monday, William Guier (1926–2011) and George Weiffenbach (1921–2003), two scientists at the Applied Physics Laboratory (APL) of Johns Hopkins University, started recording and analyzing the 20 MHz carrier signals from Sputnik I. These signals exhibited noticeable patterns of Doppler shift as the satellite passed from horizon to horizon. Weiffenbach was able to use a spectrum analyzer to track the Doppler frequency shift as the satellite passed from horizon to horizon, generally within a period of several minutes. Curious to understand how the satellite orbit influenced the observed patterns of Doppler shift, Guier and Weiffenbach calculated partial derivatives of Doppler shift with respect to orbital parameters.

For any parameter p_k of the satellite orbit, Guier and Weiffenbach could obtain numerical partial derivatives of the measurable Doppler frequency shift $f_{\mathrm{Dop}}(t)$ at the known receiver location to that parameter be generating an orbit with perturbed value $p_k + \delta_{p,k}$ and calculating the resulting perturbations $\delta_{f,k}(t_i)$ in the Doppler shifts at the receiver at sample times t_i during a satellite pass by, as

$$\frac{\partial f_{\mathrm{Dop}}(t_i)}{\partial p_k} \approx \frac{\delta_{f,k}(t_i)}{\delta_{p,k}}.$$

Small variations $\Delta_{p,k}$ in the orbit parameters should then be approximately related to observable deviations $\Delta_{Dop}(t_i)$ of Doppler shift during one satellite pass by the

[4]Named after the Danish mathematician Jørgen 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 probability 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.

linear system of equations

$$\begin{bmatrix} \Delta_{\text{Dop}}\left(t_1\right) \\ \Delta_{\text{Dop}}(t_2) \\ \Delta_{\text{Dop}}(t_3) \\ \vdots \\ \Delta_{\text{Dop}}(t_N) \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial f_{\text{Dop}}(t_1)}{\partial p_1} & \frac{\partial f_{\text{Dop}}(t_1)}{\partial p_2} & \frac{\partial f_{\text{Dop}}(t_1)}{\partial p_3} & \cdots & \frac{\partial f_{\text{Dop}}(t_1)}{\partial p_n} \\ \frac{\partial f_{\text{Dop}}(t_2)}{\partial p_2} & \frac{\partial f_{\text{Dop}}(t_2)}{\partial p_2} & \frac{\partial f_{\text{Dop}}(t_2)}{\partial p_3} & \cdots & \frac{\partial f_{\text{Dop}}(t_2)}{\partial p_n} \\ \frac{\partial f_{\text{Dop}}(t_3)}{\partial p_1} & \frac{\partial f_{\text{Dop}}(t_3)}{\partial p_2} & \frac{\partial f_{\text{Dop}}(t_3)}{\partial p_3} & \cdots & \frac{\partial f_{\text{Dop}}(t_3)}{\partial p_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_{\text{Dop}}(t_N)}{\partial p_1} & \frac{\partial f_{\text{Dop}}(t_N)}{\partial p_2} & \frac{\partial f_{\text{Dop}}(t_N)}{\partial p_3} & \cdots & \frac{\partial f_{\text{Dop}}(t_N)}{\partial p_n} \end{bmatrix}}_{H} \begin{bmatrix} \Delta_{p,1} \\ \Delta_{p,2} \\ \Delta_{p,3} \\ \vdots \\ \Delta_{p,n} \end{bmatrix},$$

where the matrix H is $N \times n$, N is the number of Doppler frequency shifts observed during one pass, and n is the number of satellite orbit parameters.

The problem of estimating the n unknown orbital parameters $p_1, p_2, p_3, \ldots, p_n$, given the N observations $\Delta_{\text{Dop}}(t_i)$ is similar to the problem faced by Gauss in 1801 in solving for the "Keplerian" orbital parameters of Ceres, in that the available observations span only a small fraction of a complete orbit. Guier and Weiffenbach then did what Gauss had done: they tried Gauss's method of least squares to see whether a suitable estimate could be obtained. That would depend on whether the associated $n \times n$ Gramian matrix

$$G = H^T H$$

is invertible. Gauss had taken several months to obtain his solution, but Guier and Weiffenbach had something Gauss did not have: the use of a Univac 1103A computer.[5]

At first, following the approach of Gauss, the partial derivatives were with respect to the six Keplerian parameters of the Sputnik I orbit. However, effects of gravitational anomalies on satellite orbits were found to be more significant than anticipated, and partial derivatives with respect to the dominant gravitational anomalies were then added to the linearized model. Additional partial derivatives with respect to ionospheric propagation effects and satellite transmitter frequency were also added. In all cases, it could be shown that the associated Gramian matrices G are nonsingular; indicating that the satellite orbit is determinable from the Doppler shift pattern from a single satellite pass by of a receiver with known location. As an added bonus, the solution also provided estimates of anomalies in the gravitational field at satellite altitudes.

In March of 1958, these results were reviewed by Frank McClure (1916–1973), director of APL's Research Center. McClure asked whether this relationship could be inverted to determine the horizontal receiver location, given the Doppler shift history and the satellite ephemeris (orbit description), and Guier and Weiffenbach were able

[5] A vacuum-tube computer with roughly the same capabilities as the IBM 704. It could have one to four banks of 18 kB magnetic core random-access memories in addition to magnetic drum memory, a 36-bit data word, and multiply times in the order of a few tenths of a millisecond.

to show that the 2×2 Gramian matrix for this problem is also nonsingular. That is, given the Doppler frequency shift pattern from one pass by a satellite with known orbit, one can obtain a least-squares solution for the longitude and latitude of the receiver antenna.

This discovery would result in the development of the world's first satellite navigation system: the US Navy's Transit Navigation System. In December of 1956, the US Navy had committed to develop a new class of nuclear powered ballistic missile submarines to be launched in the 1960s, but these submarines would need an accurate position fix before launching their missiles. The Transit Navigation System would fulfill that need. It became operational in the 1960s and remained in operation until it was eclipsed by Global Positioning System (GPS) in the 1990s.

1.3.2.2 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.

1.3.2.3 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.

Example 1.3 (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

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

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

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

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

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

is nonsingular.

1.3.2.4 Gramian Matrices 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 ($\mathcal{G}$), 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.

1.3.3 Mathematical Modeling of Uncertainty

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[6] (1501–1576) performed an accurate analysis of probabilities for games involving dice. He assumed that successive tosses of the dice were statistically independent events. Like the pioneering Indian mathematician Brahmagupta (589–668), Cardano 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 (1622–1662), 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.

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. Andrei Andreyevich Markov (1856–1922) would develop much of the theory of what is today called a *Markov process* (in continuous time) or *Markov chain* (in discrete time), a random process with the property that the evolution over time of its probability distribution can be treated as an initial-value problem. That is, the instantaneous variation with time of the probability distribution of possible states of the process is determined by the current distribution, which includes the effects of all past history of the process.

An important figure in probability theory and the theory of random processes in the twentieth century was the Russian academician Andrei Nikolayevich Kolmogorov (1903–1987). Starting around 1925, working with Aleksandr Yakovlevich Khinchin and others, he reestablished the foundations of probability theory on *measure theory,* which had originated as the basis for integration theory and became the accepted mathematical basis of probability and random processes. Along with Norbert Wiener, 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 theory was the first formal theory of optimal estimation for systems involving random processes.

[6]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*, *Cardan shaft*, or *universal joint*.

1.3.4 The Wiener–Kolmogorov 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 3 years at the Tufts University. He then entered graduate school at the 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 6 more years. Then, in 1919, he obtained a teaching appointment at the Massachusetts Institute of Technology (MIT). He remained on the faculty at the 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–Kolmogorov 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.[7]

1.3.4.1 Wiener–Kolmogorov 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 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.

In his derivation of an optimal estimator, Wiener would use probability measures on function spaces to represent uncertain dynamics. He 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 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.

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

1.3.5 The 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 the Youngstown College there for 3 years before entering the MIT.

Kalman received his bachelor's and master's degrees in electrical engineering at the 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 [9]. 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. Du Pont Company, Kalman obtained an appointment as lecturer and graduate student at the Columbia University. At that time, Columbia was well known for the work in control theory by John R. Ragazzini, Lotfi A. Zadeh,[8] 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 6 years after that at the research center of the Glenn L. Martin Company in Baltimore, the Research Institute for Advanced Studies (RIAS).

To head its mathematics division, RIAS had lured mathematician Solomon Lefschetz (1884–1972) from Princeton. Lefschetz had been a classmate with rocket pioneer Robert H. Goddard (1882–1945) at the Clark University, and thesis advisor to Richard E. Bellman (1920–1984) at Princeton. Lefschetz hired Kalman on the recommendation of Robert W. Bass, who had been a postdoc under Lefschetz at Princeton before coming to the RIAS in 1956. Kalman recommended Richard S. Bucy, who would join him at the RIAS.

1.3.5.1 Discovery In 1958, the Air Force Office of Scientific Research (AFOSR) was funding Kalman and Bucy to do advanced research in estimation and control at the RIAS.

In late November of 1958, not long after coming to the 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–Kolmogorov filtering*

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

[9] Although frequency-domain methods were then the preferred approach to the filtering problem, the use of time-domain state-space models for time-varying systems had already been introduced (e.g., by Laning and Battin. [10] in 1956).

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. The rest is history.

The Kalman filter is the culmination of a progression of models and associated optimal estimation methods for dynamic processes.

1. Wiener–Kolmogorov models use the PSD in the frequency domain to characterize the dynamic and statistical properties of a dynamic process. Optimal Wiener–Kolmogorov estimators are derivable from the PSD, which can be estimated from measured system outputs. This assumes the dynamic process model is time invariant.
2. Control theorists use linear differential equations as dynamic system models. This led to the development of mixed models, in which the dynamic system functions as a "shaping filter" excited by white noise. Coefficients of the linear differential equations determine the shape of the output PSD, and the shape of the PSD defines the Wiener–Kolmogorov estimator. This approach allows the dynamic system model to be time varying. These linear differential equations can be modeled as a system of first-order differential equations in what has come to be called *state space*.

The next step in this progression would be to develop the equivalent estimation methods right from a time-varying state-space model—and that is what Kalman did.

According to Robert W. Bass (1930–2013) [11], who was at the RIAS in that period, it was Richard S. Bucy who recognized that—if one assumes a finite-dimensional state-space model—the Wiener–Hopf equation used in deriving the Wiener–Kolmogorov filter is equivalent to a nonlinear matrix-valued differential equation. Bucy also recognized that the nonlinear differential equation in question was of the same type as one studied by Jacopo Francesco Riccati (1676–1754) more than two centuries earlier, now called *the Riccati equation.* 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.

With the additional assumption of finite dimensionality, Kalman was able to derive the Wiener–Kolmogorov 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.

Earlier results The Danish astronomer Thorvald Nicolai Thiele (1838–1910) had derived what is essentially the Kalman filter for scalar processes, and some of the seminal ideas in the Kalman filter had been published by Peter Swerling (1929–2001) in 1959 [12] and Ruslan Leont'evich Stratonovich (1930–1997) in 1960 [35].

1.3.5.2 Introduction of the Kalman Filter Kalman's ideas were met with some skepticism among his peers, and he chose a mechanical 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."[10] His second paper, on the continuous-time case and coauthored with Bucy, 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 hundreds of doctoral theses in electrical engineering over the next decade or so.

1.3.5.3 Early Applications: The Influence of Stanley F. Schmidt 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 [13]. Schmidt had known Kalman from meetings at technical conferences and had invited him to Ames to further explain his approach. Schmidt had 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* (*EKF*),which has been used ever since for many 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, then shortened to Draper Laboratory). Battin was already using state-space methods for the design and implementation of astronautical guidance systems, and he made the Kalman filter as part of the Apollo onboard guidance, 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.

1.3.5.4 Other Accomplishments of Kalman 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.

[10]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 the UCLA and sponsored jointly by the UCLA and the National Aeronautics and Space Administration (NASA) Dryden Laboratory.

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.

For his many contributions to mathematical engineering, Kalman was awarded the IEEE Medal of Honor in 1974, the IEEE Centennial Medal in 1984, the Steele Prize of the American Mathematical Society in 1987, and the Bellman Prize of the American Automatic Control Council in 1997.

In 1985, the first year the Inamori Foundation awarded its Kyoto Prizes, Kalman was awarded the Kyoto Prize in Advanced Technology. 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.

Kalman is a member of the US National Academy of Sciences, the US National Academy of Engineering, the American Academy of Arts and Sciences, and a foreign member of the French, Hungarian, and Russian Academies of Sciences.

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 [14].

On February 19, 2008, the US National Academy of Engineering awarded Kalman the Draper Prize, the Nation's most prestigious award in engineering, at an evening ceremony in Washington, DC.

In a ceremony at the White House on October 7, 2009, Kalman was awarded the National Medal of Science by US President Barak Obama.

1.3.5.5 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.

1.3.5.6 Relative Advantages of Kalman and Wiener–Kolmogorov Filtering

1. The Wiener–Kolmogorov 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 when 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–Kolmogorov 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–Kolmogorov 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.

1.3.6 Implementation Methods

1.3.6.1 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.

1.3.6.2 Early ad hoc Fixes It was discovered early on[11] 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 [15]. 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. This approach was too easily abused for covering up modeling errors, however.

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

A symmetrized form of the discrete-time Riccati equation was developed by Joseph [16] 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 [17].

1.3.6.3 James E. Potter (1937–2005) and Square-Root Filtering The first big breakthrough for improving the numerical stability of Kalman filtering occurred at the Instrumentation Laboratory at MIT, the prime contractor for guidance and control of the Apollo moon project. The Kalman filter for Apollo navigation could be implemented in 36-bit floating-pointing arithmetic on an IBM 7000-series mainframe computer, but it would eventually have to run on a flight computer using 15-bit fixed-point arithmetic. The main problem was implementing the Riccati equation solution. James Potter was then an MIT graduate student working part-time at the laboratory. He took the problem home with him on a Friday and came back on the following Monday with the solution.

Potter introduced the idea of factoring the covariance matrix as

$$P = GG^{\mathrm{T}} \tag{1.14}$$

and expressing the observational update equations in terms of G, rather than P. The result was better numerical stability of the filter implementation. An even more efficient implementation—in terms of *triangular* factors—was published by Bennet in 1967 [18], and the solution was generalized to vector-valued measurements by Andrews in 1968 [19].

Cholesky Factors André-Louis Cholesky[12] (1875–1918) derived an algorithm for solving least-squares problems that included factoring a symmetric positive-definite matrix P as the symmetric product of a triangular matrix C with positive diagonal elements and its transpose:

$$P = CC^{\mathrm{T}}, \tag{1.15}$$

called the *Cholesky decomposition* of P. The triangular factor C is called a Cholesky factor of P.

Generalized Cholesky Factors By convention, only triangular matrices with positive diagonal elements are considered to be Cholesky factors. Otherwise, the solution of Equation 1.14 is not unique. If C is the Cholesky factor of P and M is any orthogonal matrix (so that $MM^{\mathrm{T}} = I$), the matrix

$$G = CM \tag{1.16}$$

[12]Because Cholesky was French, his last name should perhaps be pronounced something like "show-less-KEY," with the accent on the last syllable. Cholesky was a French artillery officer killed in action in World War I, and his algorithm was published posthumously by fellow officer Commandant Benoit [20]. Cholesky may not have been the first to derive the factoring algorithm, but his name was soon attached to it as a matter of respect.

also satisfies the equation

$$GG^{\mathrm{T}} = (CM)(CM)^{\mathrm{T}} \tag{1.17}$$

$$= CMM^{\mathrm{T}}C^{\mathrm{T}} \tag{1.18}$$

$$= CIC^{\mathrm{T}} \tag{1.19}$$

$$= CC^{\mathrm{T}} \tag{1.20}$$

$$= P. \tag{1.21}$$

But, because G is not necessarily triangular with positive diagonal elements, we will call any solution G of $GG^{\mathrm{T}} = P$ a *generalized Cholesky factor* of P.

Matrix Square Roots A square root S of a matrix P satisfies the equation $P = SS$ (i.e., without the transpose on the second factor).

Square-Root Filtering Potter's derivation used a special type of symmetric matrix called an *elementary matrix*, a concept introduced by Householder [21]. Potter factored an elementary matrix as the square of another elementary matrix. In this case, the factors were truly square roots of the factored matrix.

The application on which Potter was working on was for dynamics in space, where there is no appreciable dynamic disturbance noise. In that case, the propagation over a discrete time interval of the covariance matrix P of navigation could be implemented as the double matrix product $\Phi P \Phi^{\mathrm{T}}$, where Φ is a known *state transition matrix* for trajectories in space. Potter could then propagate his generalized Cholesky factor G of P forward in time with a single matrix multiply as ΦG. In doing so, G would no longer remain either a square root or a Cholesky factor of P (unless it remained symmetric). However, this "square-root" appellation has stuck with extensions of Potter's approach, even though the factors involved are generalized Cholesky factors, not matrix square roots.

1.3.6.4 Improved 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 [22] on temporal update methods for Cholesky factors. Extensions of square-root covariance and information filters were introduced in Kaminski's 1971 thesis [23] at Stanford University. The first of the triangular factoring algorithms for the observational update was due to Agee and Turner [24], 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 [25], followed by the "square-root-free" algorithm of Bierman in 1974 [26] and the associated temporal update method introduced by Thornton [27]. The computational complexity of the square-root filter for time-invariant systems was greatly simplified by Morf and Kailath [28] soon after that. Specialized

parallel processing architectures for fast solution of the square-root filter equations were developed by Jover and Kailath [29] and others over the next decade, and much simpler derivations of these and earlier square-root implementations were discovered by Kailath [30].

1.3.6.5 Matrix Decomposition, Factorization, and Triangularization These terms are bandied about in square-root filtering, often interchangeably. There are some distinctions, however.

Matrix Decomposition The term *decomposition* is perhaps the most broad. It generally refers to decomposing a matrix into a representation composed of different parts with some useful properties. For example, the "singular value decomposition" (SVD) of a symmetric positive-definite $n \times n$ matrix P yields the product decomposition of P as $P = EDE^{\mathrm{T}}$, where the column vectors of the orthogonal matrix E are the eigenvectors of P and the diagonal matrix D has the corresponding eigenvalues on its diagonal, leading to the alternative representation of $P = EDE^{\mathrm{T}}$ as the "eigenvalue-eigenvector decomposition" of P:

$$P = \sum_{i=1}^{n} \lambda_i e_i e_i^{\mathrm{T}},$$

where the λ_i are the (positive) eigenvalues of P and the e_i are the associated eigenvectors.[13] The SVD, like many other factorization methods used in "square-root" filtering, is also used for solving least-squares problems [31]. The so-called "QR decomposition" of a matrix is another used for solving least-squares problems. It factors a matrix as the product of an orthogonal matrix (Q) and a 'triangular"[14] matrix R (i.e., with zeros either above or below the main diagonal). (However, this notation does conflict with standard notation for Kalman filtering.) The Cholesky decomposition also produces triangular factors, but the term *decomposition* by itself does not imply matrix factoring. For any square matrix S, for example, the symmetric–antisymmetric decomposition

$$S = \underbrace{\frac{1}{2}(S + S^{\mathrm{T}})}_{\text{sym.}} + \underbrace{\frac{1}{2}(S - S^{\mathrm{T}})}_{\text{anti.}}$$

decomposes S as the sum of its symmetric and antisymmetric parts.

Matrix Factorization *Factorization* is a term used by Gerald Bierman (1941–1987) for methods to factor a matrix into a product of matrices with more useful properties for Kalman filtering implementation [32]. For example, the so-called "UD

[13] This eigenvalue-eigenvector decomposition is a property of all "normal" matrices, defined as the square matrices S such that $SS^{\mathrm{T}} = S^{\mathrm{T}}S$.

[14] See Chapter 7 and Appendix B (on the Wiley web site) for further discussions of triangular forms.

decomposition" used by Bierman factors a symmetric positive-definite matrix P as

$$P = UDU^{\mathrm{T}},$$

where D is diagonal with positive diagonal entries and U is the "unit triangular matrix" (i.e., triangular with ones along its main diagonal). "Factorization" generally refers to the algorithmic methods used for obtaining the result, often (but not always) done "in-place" (i.e., in memory, overwriting the input matrix with the factor(s)). For example, UD factorization can overwrite the diagonal of the input matrix with D and off-diagonal terms with those of U (because the diagonal of U is known to contain only ones).

Matrix Triangularization The term *triangularization* refers to factorization in which the resulting factor is triangular. It is used for "QR decompositions" performed in-place, destroying the original matrix and replacing it with its triangular factor (R). The orthogonal transformation Q is not saved, but the operations used to render the effect of Q tend to be well conditioned numerically. The sequence of operations performed in-place is called *triangularization* of the original matrix. Triangularization methods derived by Givens [33], Householder [21], and Gentleman [33] are used to make Kalman filtering implementations more robust against roundoff errors.

The more useful factorization and triangularization methods for Kalman filtering are described in Chapter 7.

1.3.6.6 Generalizations Linear estimation theory has been extended to nonquadratic error criteria, as well. Optimization with respect to the "sup norm" or H_∞ norm minimizes the maximum error, which is advantageous for applications in which the associated risk is decidedly nonquadratic. The first major application of Kalman filtering (for Apollo navigation to the moon and back) had very hard constraints on atmospheric entry on the return to Earth. Large excursions of the entry angle could result in spacecraft burn-up (too steep) or skip-out (too shallow). An H_∞ estimator might have been more appropriate under those circumstances, but it had not been developed yet.

For a more expansive view of linear estimation methods, see Kailath et al. [34] and the references therein.

1.3.7 Nonlinear Approximations

It is human nature to use successful approaches to problem solving within a limited context on problems outside that context. The Kalman filter is no exception to this rule. Those experienced with Kalman filtering often find themselves morphing problems to resemble the Kalman filtering model.

This is especially so with nonlinear problems, for which there is no practical and mathematically correct approach comparable to the Kalman filter. Although it was

originally derived for linear problems, the Kalman filter is habitually applied to nonlinear problems by using various approximation methods. This approach has worked remarkably well for a number of nonlinear problems, but there will always be limits to how far it can be pushed.

We mention here some approaches that have been used to extend the applicability of Kalman filtering methodologies to nonlinearly problems. The more successful of these are described in greater detail in Chapter 8.

1.3.7.1 Extended Kalman Filtering (EKF) for Quasilinear Problems EKF was used in the very first application of Kalman filtering: the space navigation problem for the Apollo missions to the moon and back. The approach has been successfully applied to many nonlinear problems ever since.

Success depends on the problem being *quasilinear* and sufficiently dominated by linearity within the expected range of variation that unmodeled errors due to linear approximation are insignificant compared to the modeled errors due to dynamic uncertainty and sensor noise. Methods for verifying whether a problem is sufficiently quasilinear are presented in Chapter 8.

In EKF, linear approximation is used only for solving the Riccati equation, a partial result of which is the Kalman gain. The full nonlinear model is used in propagation of the estimate and in computing predicted sensor outputs.

The approach uses partial derivatives as linear approximations of nonlinear relations. Schmidt [13] introduced the idea of evaluating these partial derivatives at the *estimated* value of the state variables. This and other methods for approximate linear solutions to nonlinear problems are discussed in Chapter 8.

1.3.7.2 Higher Order Approximations Approaches using higher order expansions of the filter equations (i.e., beyond the linear terms) have been derived by Stratonovich [35], Kushner [36], Bucy [37], Bass et al. [38], and others for quadratic nonlinearities, and by Wiberg and Campbell [39] for terms through third order. However, none of these has proven to be very practical.

1.3.7.3 Sampling-Based Methods for Nonlinear Estimation The Kalman filtering methodology has been further extended to problems for which EKF exhibits unacceptable errors. The general approach to approximating nonlinear propagation of the Riccati equation solution is by using representative samples of state variables—as opposed to linearized propagation of the mean (i.e., the estimated state) and covariance matrix of the distribution.

In the 1940s, mathematician Stanislaw Ulam conceived the idea of using pseudorandom sampling to characterize the evolution of neutron distributions in thermonuclear devices. Colleague Nicholas Metropolis coined the term *Monte Carlo*[15] for such methods. Much of the initial development of Monte Carlo techniques occurred at the Los Alamos Laboratory, where adequate computer resources were then becoming

[15]The name refers to the Monaco Monte Carlo gambling casino, which uses pseudorandom methods to transform the distribution of wealth among its players.

available. Others involved in this development at Los Alamos included Enrico Fermi and John von Neumann.

The computational burden of sample-based analysis can be reduced significantly by using more judicious sampling rules, in place of random sampling:

1. In *sequential Monte Carlo* methods, the samples are selected in the order of their relative importance for representing the significant features of the distribution.
2. In *sigma point*, the samples can be based on the eigenvectors and eigenvalues (usually represented by the symbol σ^2) of the covariance matrix.
3. *Unscented transform methods* select samples using the Cholesky decomposition of the covariance matrix. The resulting filter implementation is called *unscented Kalman filtering,* a terminology introduced by Jeffrey Uhlmann. This approach also includes a nonlinear approximation for the cross-covariance of the predicted state vector and the predicted measurement and weighting parameters that can be adjusted for "tuning" the filter to the particular nonlinearities of the application. Unscented transformations for n-dimensional distributions may use $n + 1$ or $2n + 1$ samples, which is about minimal for sample-based methods.

The term *particle filter* is also used to denote extensions of the Kalman filter that use sample-based methods, because the sampled values can be viewed as "particles" carried along by the nonlinear system dynamics.

In all cases, the samples of state vector values are chosen to represent the mean and covariance structure of the ensemble a posteriori distribution (i.e., after the measurement information has been used for refining the estimate). These sample points are then propagated forward in time by simulating the known nonlinear system dynamics, and the resulting a priori covariance at the next measurement opportunity is inferred from the resulting distribution after the nonlinear transformations of individual samples. The resulting covariance structure is then used in computing the Kalman gains to use the measured sensor outputs.

The more successful of these methods are described in Chapter 8. The unscented Kalman filter, in particular, has been shown to be efficient and effective for some of the more nonlinear applications—including system identification (i.e., estimation of dynamic model parameters), a notoriously nonlinear and difficult problem.

1.3.8 Truly Nonlinear Estimation

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 [40], Fokker [41], Planck [42], Kolmogorov [43], Stratonovich [35], Baras and Mirelli [44], and others. Stratonovich modeled the effect on the probability distribution of information obtained through noisy measurements of the dynamic

system, an effect he called *conditioning*. The partial differential equation that includes these effects is called the *conditioned Fokker–Planck equation*. It has also been studied by Kushner [36], Bucy [37], and others using the *stochastic calculus* of Stratonovich or Itô. The theoretical basis for stochastic differential equations was long been hampered by the fact that white noise is not a Riemann-integrable function, but the non-Riemannian *stochastic integrals* of Stratonovich or Itô fixed that.

The general approach results in a stochastic partial differential equation describing the evolution over time of the probability distribution over a "state space" of the dynamic system under study. The resulting models do *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.

For a concise but readable treatment of the stochastic calculus for Kalman filtering, see Jazwinski [45].

1.3.9 The Detection Problem for Surveillance

Surveillance problems include the detection, identification, and tracking of objects within a certain region of space. The Kalman filter helps in solving the tracking problem and may be of some utility (as a nonlinear filter) in solving the identification problem. However, the *detection problem* must usually be solved before identification and tracking can begin. The Kalman filter requires an initial state estimate for each object, and that initial 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 *point process* is a type of random process for modeling events or objects that are distributed over time and/or space, such as the arrivals[16] 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 in time and space of aircraft or spacecraft under surveillance by a radar installation, or the locations of submarines under sonar surveillance in the ocean.

A unified approach combining detection and tracking into one optimal estimation method was derived by John M. Richardson (1918–1996) and specialized to several applications [46]. 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 Gaussian-like closure assumption about the relationships between moments. 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.

[16]In these applications, a point process is also called *arrival process*.

1.4 COMMON 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 basic phonetic elements of standard spoken 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.

1.4.1 "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.

1.4.2 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 [50]. 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.

1.4.2.1 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

TABLE 1.2 Common Symbols Used in Kalman Filtering

Sources*			
(a)	(b)	(c)	Symbol Definition
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 the 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 and References 2, 47, 48 and 49.
(b) Reference 50.
(c) References 51–53, and 54.

TABLE 1.3 Special State-Space Notation

This Book	Other Sources	Definition of Notational Usage
x	$\underline{x}$, $\vec{x}$, $\mathbf{x}$	State 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
$\hat{x}$	$\mathrm{E}\langle x\rangle$, $\overline{x}$	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)

lists the notation used in this book (first 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.

1.4.3 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. [2] 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.

TABLE 1.4 Common Notation for Array Dimensions

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

[17] This use of 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.

1.5 SUMMARY

The *Kalman filter* is an estimator used to estimate the *state* of a *linear dynamic system* perturbed by *white noise* using measurements that are *linear* functions of the system state but corrupted by *additive* 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, discovered by Gauss (and others) around the end of the eighteenth century. 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.

Sequential Monte Carlo methods and particle filtering can be used to extend Kalman filtering beyond the *quasilinear* estimation problems that are solvable by *extended Kalman filtering.*

The *Unscented Kalman Filter* has about the same computational complexity as the extended Kalman filter and essentially the same numerical stability as square-root filtering but with potentially greater robustness against nonlinear effects.

PROBLEMS

1.1 Derive the least-squares equations for finding a and b that provide the best fit to the three equations

$$2 = a + b$$
$$4 = 3a + b$$
$$1 = 2a + b.$$

(a) Express the system of equations in matrix form as

$$z = A \begin{bmatrix} a \\ b \end{bmatrix},$$

where z is a column vector with three rows and the matrix A is 3×2.

(b) Take the matrix product $A^{\mathrm{T}}A$ for A derived in (a).

(c) Take the 2×2 matrix inverse

$$[A^{\mathrm{T}}A]^{-1}$$

for the A derived in (a). [Hint: Use the general formula

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{bmatrix}^{-1} = \frac{1}{m_{11}m_{22} - m_{12}^2} \begin{bmatrix} m_{22} & -m_{12} \\ -m_{12} & m_{11} \end{bmatrix}$$

for inverting symmetric 2×2 matrices.]

(d) Take the matrix product

$$A^{\mathrm{T}} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

for the A derived in (a).

(e) Calculate the least-squares solution

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = [A^{\mathrm{T}}A]^{-1}A^{\mathrm{T}} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

for the $[A^{\mathrm{T}}A]^{-1}$ derived in (c) and

$$A^{\mathrm{T}} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

derived in (d).

1.2 Find the least-squares solution for a and b in the four equations

$$\begin{aligned} 2 &= a + b \\ 4 &= 3a + b \\ 1 &= 2a + b \\ 4 &= 4a + b. \end{aligned}$$

1.3 The "straight-line fit" problem with uniform sampling is to find a bias b and ramp coefficient a to fit a set of N measured values $z_1, z_2, z_3, \ldots, z_N$ sampled at uniform time intervals Δt. The problem can modeled by a system of N linear equations

$$\begin{aligned} z_1 &= a \times 1\Delta t + b \\ z_2 &= a \times 2\Delta t + b \\ z_3 &= a \times 3\Delta t + b \\ &\vdots \\ z_N &= a \times N\Delta t + b, \end{aligned}$$

where the "unknowns" are a and b.

(a) Express the system of equations in matrix form, using dots to represent A in the form

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ \vdots & \vdots \\ a_{N1} & a_{N2} \end{bmatrix},$$

but with formulas for the matrix elements a_{ij}.

(b) Derive a symbolic formula for the 2×2 matrix $A^{\mathrm{T}}A$ for A as defined in (a).

(c) Derive a general formula for $[A^{\mathrm{T}}A]^{-1}$, for the $A^{\mathrm{T}}A$ defined in (b).

(d) Use the above results to derive formulas for the least-squares estimates $\hat{a}$, $\hat{b}$ for the general system of N linear equations.

1.4 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:

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

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 $1 \le j \le n$, given the least integrated squared approximation error

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

You may assume the equalities

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

as given.

REFERENCES

[1] V. Beneš and R. J. Elliott, "Finite-dimensional solutions of a modified Zakai equation," *Mathematics of Control, Signals, and Systems*, Vol. 9, pp. 341–351, 1996.

[2] A. Gelb, J. F. Kasper Jr., R. A. Nash Jr., C. F. Price, and A. A. Sutherland Jr., *Applied Optimal Estimation*, MIT Press, Cambridge, MA, 1974.

[3] D. G. Lainiotis, "Estimation: a brief survey," *Information Sciences*, Vol. 7, pp. 191–202, 1974.

[4] J. M. Mendel and D. L. Geiseking, "Bibliography on the linear-quadratic-Gaussian problem," *IEEE Transactions on Automatic Control*, Vol. AC-16, pp. 847–869, 1971.

[5] R. H. Battin, "Space guidance evolution—a personal narrative," *AIAA Journal of Guidance and Control*, Vol. 5, pp. 97–110, 1982.

[6] S. F. Schmidt, "The Kalman Filter: its recognition and development for aerospace applications," *AIAA Journal of Guidance and Control*, Vol. 4, No. 1, pp. 4–7, 1981.

[7] R. M. L. Baker and M. W. Makemson, *An Introduction to Astrodynamics*, Academic Press, New York, 1960.

[8] T. Kailath, "A view of three decades of linear filtering theory," *IEEE Transactions on Information Theory*, Vol. IT-20, No. 2, pp. 146–181, 1974.

[9] R. E. Kalman, *Phase-Plane Analysis of Nonlinear Sampled-Data Servomechanisms*, M.S. thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 1954.

[10] J. H. Laning Jr. and R. H. Battin, *Random Processes in Automatic Control*, McGraw-Hill, New York, 1956.

[11] R. W. Bass, "Some reminiscences of control and system theory in the period 1955–1960," talk presented at *The Southeastern Symposium on System Theory*, University of Alabama at Huntsville, 18 March 2002.

[12] P. Swerling, "First order error propagation in a stagewise differential smoothing procedure for satellite observations," *Journal of Astronautical Sciences*, Vol. 6, pp. 46–52, 1959.

[13] L. A. McGee and S. F. Schmidt, *Discovery of the Kalman Filter as a Practical Tool for Aerospace and Industry*, Technical Memorandum 86847, National Aeronautics and Space Administration, Ames Research Center, Moffett Field, California, 1985.

[14] A. C. Antoulas, Ed., *Mathematical System Theory, The Influence of R. E. Kalman*, Springer-Verlag, Berlin, 1991.

[15] M. Verhaegen and P. Van Dooren, "Numerical aspects of different Kalman filter implementations," *IEEE Transactions on Automatic Control*, Vol. AC-31, pp. 907–917, 1986.

[16] R. S. Bucy and P. D. Joseph, *Filtering for Stochastic Processes, with Applications to Guidance*, John Wiley & Sons, Inc., New York, 1968.

[17] C. L. Thornton and G. J. Bierman, *A Numerical Comparison of Discrete Kalman Filtering Algorithms: An Orbit Determination Case Study*, JPL Technical Memorandum, Pasadena, CA, pp. 33–771, 1976.

[18] J. M. Bennet, "Triangular factors of modified matrices," *Numerische Mathematik*, Vol. 7, pp. 217–221, 1963.

[19] A. Andrews, "A square root formulation of the Kalman covariance equations," *AIAA Journal*, Vol. 6, pp. 1165–1166, 1968.

[20] Commandant Benoit, "Note sur une méthode de résolution des équations normales provenant de l'application de la méthode des moindres carrés à un systéme d'équations linéaires en nombre infèrieur à celui des inconnues," *(Procédé du Commandant Cholesky), Bulletin Géodésique*, Vol. 2, pp. 67–77, 1924.

[21] A. S. Householder, "Unitary triangularization of a nonsymmetric matrix," *Journal of the Association for Computing Machinery*, Vol. 5, pp. 339–342, 1958.

[22] P. Dyer and S. McReynolds, "Extension of square-root filtering to include process noise," *Journal of Optimization Theory and Applications*, Vol. 3, pp. 444–458, 1969.

[23] P. G. Kaminski, *Square Root Filtering and Smoothing for Discrete Processes*, PhD thesis, Stanford University, 1971.

[24] W. S. Agee and R. H. Turner, *Triangular Decomposition of a Positive Definite Matrix Plus a Symmetric Dyad, with Applications to Kalman Filtering*, White Sands Missile Range Tech. Rep. No. 38, Oct. 1972.

[25] N. A. Carlson, "Fast triangular formulation of the square root filter," *AIAA Journal*, Vol. 11, No. 9, pp. 1259–1265, 1973.

[26] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, Academic Press, New York, 1977.

[27] C. L. Thornton, *Triangular Covariance Factorizations for Kalman Filtering*, PhD thesis, University of California at Los Angeles, School of Engineering, 1976.

[28] M. Morf and T. Kailath, "Square root algorithms for least squares estimation," *IEEE Transactions on Automatic Control*, Vol. AC-20, pp. 487–497, 1975.

[29] J. M. Jover and T. Kailath, "A parallel architecture for Kalman filter measurement update and parameter update," *Automatica*, Vol. 22, pp. 783–786, 1986.

[30] T. Kailath, "State-space modelling: square root algorithms," in *Systems and Control Encyclopedia* (M. G. Singh, Ed.), Pergamon, Elmsford, NY, 1984.

[31] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische Mathematik*, Vol. 14, No. 5, pp. 403–420, 1970.

[32] W. Givens, "Computation of plane unitary rotations transforming a general matrix to triangular form," *Journal of the Society for Industrial and Applied Mathematics*, Vol. 6, pp. 26–50, 1958.

[33] W. M. Gentleman, "Least squares computations by Givens transformations without square roots," *Journal of the Institute for Mathematical Applications*, Vol. 12, pp. 329–336, 1973.

[34] T. Kailath, A. H. Sayed and B. Hassibi, *Linear Estimation,* Prentice-Hall, Upper Saddle River, NJ, 2000.

[35] R. L. Stratonovich, *Topics in the Theory of Random Noise*, (R. A. Silverman, Ed.), Gordon & Breach, New York, 1963.

[36] H. J. Kushner, "On the differential equations satisfied by conditional probability densities of Markov processes," *SIAM Journal on Control, Series A*, Vol. 2, pp. 106–119, 1964.

[37] R. S. Bucy, "Nonlinear filtering theory," *IEEE Transactions on Automatic Control*, Vol. AC-10, pp. 198–206, 1965.

[38] R. W. Bass, V. D. Norum, and L. Schwartz, "Optimal multichannel nonlinear filtering," *Journal of Mathematical Analysis and Applications*, Vol. 16, pp. 152–164, 1966.

[39] D. M. Wiberg and L. A. Campbell, "A discrete-time convergent approximation of the optimal recursive parameter estimator," *Proceedings of the IFAC Identification and System Parameter Identification Symposium*, Vol. 1, pp. 140–144, 1991.

[40] A. Einstein, "Über die von molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen," *Annelen der Physik*, Vol. 17, pp. 549–560, 1905.

[41] A. D. Fokker, "Die mittlerer Energie rotierender elektrischer Dipole im Strahlungsfeld," *Annelen der Physik*, Vol. 43, pp. 810–820, 1914.

[42] M. Planck, "Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie," *Sitzungsberichte d. König. Preussischen Akademie der Wissenschaft*, Vol. 24, pp. 324–341, 1917.

[43] A. A. Kolmogorov, "Über die analytichen Methoden in der Wahrscheinlichkeitsrechnung," *Mathematische Annelen*, Vol. 104, pp. 415–458, 1931.

[44] J. Baras and V. Mirelli, *Recent Advances in Stochastic Calculus*, Springer-Verlag, New York, 1990.

[45] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[46] J. M. Richardson and K. A. Marsh, "Point process theory and the surveillance of many objects," in *Proceedings of the 1991 Symposium on Maximum Entropy and Bayesian Methods*, Seattle University, 1991.

[47] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

[48] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering: With MATLAB Exercises and Solutions*, 4th ed., John Wiley & Sons, Inc., New York, 2007.

[49] D. E. Catlin, *Estimation, Control, and the Discrete Kalman Filter*, Springer-Verlag, New York, 1989.

[50] R. E. Kalman, "A new approach to linear filtering and prediction problems," *ASME Journal of Basic Engineering*, Vol. 82, pp. 34–45, 1960.

[51] A. V. Balakrishnan, *Kalman Filtering Theory*, Optimization Software, New York, 1987.

[52] K. Brammer and G. Siffling, *Kalman-Bucy Filters*, Artech House, Norwood, MA, 1989.

[53] C. K. Chui and G. Chen, *Kalman Filtering with Real-Time Applications*, Springer-Verlag, New York, 1987.

[54] T. Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[55] H. W. Sorenson, "Least-squares estimation: From Gauss to Kalman," *IEEE Spectrum*, Vol. 7, pp. 63–68, 1970.

[56] H. W. Sorenson, "On the development of practical nonlinear filters," *Information Sciences*, Vol. 7, pp. 253–270, 1974.