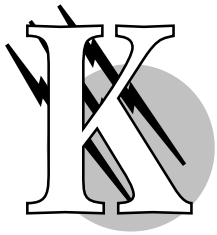


Kalman Filters and Nonlinear Filters

Vytas B. Gylys
Texas Instruments Inc.

I. Introduction
II. Linear Estimation Problem and Kalman Filters
III. Modifications of Kalman Filters and Nonlinear Filtering
IV. Filter Implementation
V. Applications

GLOSSARY

Brownian motion process Process $\{x_t\}, t \geq 0$, is a (scalar) *constant-diffusion Brownian motion process* if (a) $\{x_t\}$ is a process with independent increments and (b) for any nonnegative t_1 and t_2, the increments $(x_{t_1} - x_{t_2})$ are Gaussian random variables with $E[(x_{t_1} - x_{t_2})] = 0$ and $E[(x_{t_1} - x_{t_2})^2] = q|t_1 - t_2|$, where $q > 0$.

Covariance function of a stochastic process If $\{x_t\}$ is a second-order, scalar- or vector-valued process (that is, if $E[x_t x_t^{\mathrm{T}}] < \infty$ for every t in the parameter space T_0), its covariance function is defined as

$$R(t, s) = E[x_t - Ex_t)(x_s - Ex_s)^{\mathrm{T}}]$$

for all t, s in the parameter set T_0. This function is also known as the autocovariance function of the process.

Estimation Process of observing the values taken on by random measurements and using these measurements to compute (with the assistance of an estimator) the values of some unknown parameters expressible as functions of measurements. In this article, the term *estimation* is used synonymously with *filtering*. See also Estimator.

Estimator An estimator $\hat{\theta}$ of some θ from a parameter space Θ is a function of random measurements $M_k \equiv \{m_0, m_1, \ldots, m_k\}$ (and hence itself is a random variable) constructed for inferring the unknown value of θ. The value taken on by the estimator of θ *after* the observation of a particular set of measurements is called an *estimate* of θ. To be consistent with the notational conventions set up for this article, we do not distinguish notationally between an estimator (a random variable) and a resulting estimate (a value taken on by this random variable as a function of actually observed measurements). The term *estimator* in this article also refers to a computational procedure (algorithm) for computing estimates. In general, it is used synonymously with the term *filter*.

Extended Kalman filter (EKF) Nonlinear filter (for a system with nonlinearities in the dynamics equation

or in the measurement equation of its model) that relinearizes at each measurement update time t_k the nonlinear parts of the system model, while using a vector-valued Taylor expansion around some reference trajectory $\{s_{0k}; k = 0, 1, \ldots\}$, and then applies Kalman filter equations. The reference trajectory used is obtained at each t_k from $\hat{s}_{k-1|k-1}$ (for the computation of $\hat{s}_{k|k-1}$ and $P_{k|k-1}$) and then from $\hat{s}_{k|k-1}$ (for the computation of $\hat{s}_{k|k}$ and $P_{k|k}$). Here, $\hat{s}_{k|j}$ and $P_{k|j}$, with $k \geq j$, denote the minimum-mean-square-error estimates of the state and the state error covariance matrix based on measurements $\{m_0, m_1, \ldots, m_j\}$. The discrete-time EKF algorithm is defined in terms of Eqs. (36)–(44) in the text. A slightly different version of the linearized extension of the Kalman filter is obtained by linearizing the model around some known trajectory. This version is often called a linearized Kalman filter (LKF).

Gaussian process A scalar- or vector-valued stochastic process $\{x_t\}$ is said to be Gaussian if for every finite subset $(t_1, t_2, \ldots, t_n)$ of points from the parameter set T_0, random variables $(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$ are jointly Gaussian; that is, jointly they are distributed as a Gaussian random variable. It follows from the properties of Gaussian distribution that all finite-dimensional distributions of a Gaussian process $\{x_t\}$ are completely specified by its means $\bar{x}_t \equiv E[x_t]$ and its covariance functions $R(t, s) = E[(x_t - \bar{x}_t)(x_s - \bar{x}_s)^{\mathrm{T}}]$.

Linear system model Model in which both the stochastic differential (or difference) equations of system dynamics and the measurement equations relating the system state to the noisy measurements of the system are linear equations with respect to the state, measurement, and noise vector. The discrete-time linear system model, considered in this article for the Kalman filter and other related estimation schemes, is defined in the text in terms of Eqs. (2) and (3) and Assumptions 1D and 4D.

Markov process A process $\{x_t\}$ is a kth-order Markov process if for all $t_1 < t_2 < \cdots < t_n$ in the parameter set T_0 and some fixed integer k $(1 \leq k < n)$, the conditional probability $p[x_{t_n} < b_n \mid x_{t_1} = b_1, \ldots, x_{t_{n-1}} = b_{n-1}] = p[x_{t_n} < b_n \mid x_{t_{n-k}} = b_{n-k}, \ldots, x_{t_{n-1}} = b_{n-1}]$ holds for all n-tuples of real numbers $(b_1, b_2, \ldots, b_n)$. The forgetting-of-the-remote-past property of process $\{x_t\}$, defined by the above probability statement, is called the Markov property.

Minimum-mean-square-error (MMSE) estimator An estimator $\hat{\theta}$ of some parameter θ (fixed or random) from a parameter space is said to be a minimum-mean-square-error estimator of θ based (conditioned) on measurements $M_k \equiv \{m_0, m_1, \ldots, m_k\}$, if

$$E\left[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^{\mathrm{T}} \mid M_k\right] \leq E\left[(\hat{\theta}_0 - \theta)(\hat{\theta}_0 - \theta)^{\mathrm{T}} \mid M_k\right],$$

where $\hat{\theta}_0$ is any other estimator of θ. If θ and M_k are two jointly distributed random variables, it can be shown that the MMSE estimator $\hat{\theta}$ must be of the form $E[\theta \mid M_k]$; that is, $\hat{\theta}$ is a conditional mean estimator.

Noise process Zero-mean stochastic process.

Positive definite and positive semidefinite matrices Matrix B is *positive definite* if and only if $\mathrm{x}^{\mathrm{T}}\mathrm{Bx} > 0$ for all nonzero vectors x of compatiable length; B is *positive semidefinite* if and only if $\mathrm{x}^{\mathrm{T}}\mathrm{Bx} \geq 0$ for all nonzero vectors x of compatible size.

Process with independent increments Process $\{x_t\}$ such that, for every ordered set

$$t_0 < t_1 < t_2 < \cdots < t_n$$

of parameter values from the parameter set T_0, the random variables $(x_{t_k} - x_{t_{k-1}})$ are mutually independent for $k = 1, 2, \ldots, n$.

Recursive estimation of system state Estimation scheme that does not require that all past measurements of the system and the past system state data be saved for estimation of the current state. To estimate the current system state. To estimate the current system state, a typical recursive estimator uses only the current measurements and the most recent estimate of the system state. In a more general case, a recursive estimator may use a finite, fixed-size set of most recent measurements and system state estimates. This property greatly facilitates the implementation of a recursive estimator for real-time operation on a digital computer. Nearly all filters considered in this article, including the Kalman filter, are recursive estimators.

Second-order process Scalar- or vector-valued stochastic process $\{x_t\}$ such that $E[x_t x_t^T] < \infty$ for every t in the parameter set T_0.

Square root filtering Collection of related computational algorithms for (measurement) updating, in a numerically stable manner, of estimates P of the state error covariance matrix in a Kalman filter and for computing the filter gain matrix. These algorithms utilize a result from matrix theory, according to which any nonnegative definite symmetric P can be factored as $P = SS^{\mathrm{T}}$, where S is usually a square matrix that is not necessarily nonnegative definite symmetric.

Stationary process A process $\{x_t\}$ is said to be stationary if for any set of values $t_1, t_2, \ldots, t_n$ from the parameter set T_0 and for any t_0 such that $t_j + t_0$ is in T_0 *for all* $j = 1, \ldots, n$, the joint distribution of $(x_{tj}, j = 1, 2, \ldots, n)$ is identical to that of $(x_{t_j + t_0}, j = 1, 2, \ldots, n)$.

System measurement *m* A vector that contains the values of system measurements, which are assumed to be noisy and associated with some instant of time. In this article, the letter m is reserved for the measurement vector. In particular, m_t or $m(t)$ refers to the measurement

vector m at time t; m_k refers to m at a discrete sampling epoch t_k. The measurement vector has two interpretations: Either it represents a scalar- or vector-valued random variable (RV) or it represents actually observed values taken on by this RV. Notationally we do not distinguish between these two meanings of m.

System measurements Sets of noisy measurements of the system. Each measurement set is associated with some instant or interval of time and contains the measured values, corrupted by random noise, of all (or of a selected subset of) observable system outputs. For example, if the sensor tracking an aircraft is a radar, the set of measurements obtained at the sampling epoch t_k may consist of the range and range-rate measurements at t_k.

System model (in recursive filtering) Mathematical specification of those characteristics of a system that pertain to the construction of a recursive filtering scheme for the system. Suppose that at any instant of time the state of a system is specified by a vector-valued quantity, known as the state vector. Then the model of such a system typically consists of (a) vector-valued stochastic differential (or difference) equations that specify the dynamics (evolution) of system state with respect to time, (b) vector-valued measurement equations that relate the system state to the noisy measurements of the system, and (c) probabilistic assumptions that specify the distributions of three stochastic components in the system model (the process noise term in system dynamics equations, the measurement noise term in measurement equations, and the initial distribution of state) and the probabilistic interdependence among these components. Often the process noise and the measurement noise terms additively enter the system dynamics and the measurement equations, respectively.

System state Complete, quantitative, typically time-dependent description of the system at some instant of time. What constitutes a complete description depends on the purpose for which the information about the system state is to be used. For example, if one is interested in predicting into the near-future the position and the velocity of a particle moving in a known gravitational field, it may suffice only to specify for the particle the current values of its position coordinates and its velocity and acceleration vectors relative to some suitable coordinate frame.

(System) state vector s Data array, usually structured as a vector, that contains the quantitative description of the system state at some time point. In this article, the letter s denotes the system state vector. In particular, s_t or $s(t)$ refers to the state vectors at time t; s_k refers to s at a discrete sampling epoch t_k. Note that the state vector has two interpretations: Either it represents a random vector [then it is a vector-valued random variable (RV)] or it represents a vector of specific values taken on by this RV. Notationally we do not distinguish between these two meanings of s.

Unbiased estimator An estimator $\hat{\theta}$ of some θ from a parameter space is said to be unbiased if $E[\hat{\theta}] = \theta$.

White noise process White process $\{x_t\}$ such that $E[x_t] = 0$ for all t.

White process Scalar- or vector-valued, wide-sense stationary stochastic process $\{x_t\}$ with $E[x_t] = \bar{x}$ for all t and the covariance function defined by

$$E\left[(x_t - \bar{x})(x_s - \bar{x})^{\mathrm{T}}\right] = C\delta(t - s),$$

where constant C is a positive scalar or a positive definite matrix and $\delta(\cdot)$ denotes the Dirac delta function defined by the property

$$\delta(t) = 0 \qquad \text{for} \qquad t \neq 0 \quad \text{and}$$

$$\int_{-\infty}^{\infty} \delta(t)\, f(t)\, dt = f(0)$$

for all functions f continuous at 0. Thus, the above definition requires $\{x_t\}$ to be an uncorrelated process with respect to t, while each x_t may be a vector-valued random variable with a nonzero covariance matrix $E[(x_t - \bar{x})(x_t - \bar{x})^{\mathrm{T}}] = C$. This definition of white process is not unique. For example, some authors replace the wide-sense stationarity assumption with the less restrictive "second-order process" property. In such a case, the time-invariant fixed (scalar or matrix) C, appearing above in the definition of covariance structure, is replaced by C_t, that is, with a fixed (i.e., nonrandom) scalar or matrix that may vary with time. In this article, we define (for operational convenience) a discrete-time white process $\{x_t\}$ by requiring only that x_t and x_s be independent random variables for any $t \neq s$.

Wide-sense stationary process Scalar- or vector-valued, second-order stochastic process $\{x_t\}$ such that (a) its mean $\bar{x} \equiv E[x_t]$ is independent of t and (b) its covariance function

$$R(t, s) \equiv E\left[(x_t - \bar{x})(x_s - \bar{x})^{\mathrm{T}}\right]$$

is a function of only the difference $t - s$, that is, $R(t, s) \equiv R_0(t - s)$.

FILTERING of a stochastic dynamic system (i.e., a system the dynamics of which can be adequately described by means of a mathematical model containing stochastic components and possessing certain well-defined properties) is a process of estimating the current state of the system while utilizing the recent estimates as well as

the past and current measurements (observations) of the system. The system model used in the filtering process describes the system dynamics by means of stochastic differential or stochastic difference equations and contains a transformation relating the system state to the measurements. These measurements are assumed to be noisy. The system dynamics part of the model may also be noisy.

A Kalman filter is a recursive, unbiased, minimum-mean-square-error (MMSE) estimator that uses linear models for system dynamics and state-to-measurement transformation. Both the dynamics part of the model and the state-to-measurement transformation incorporate the noise through additive stochastic components (terms). *Recursive* here means that new estimates are produced iteratively throughout the period of observation and that only the current and recent knowledge about the system is used to estimate its current state.

Nonlinear filters considered in this article are estimators of systems that can be adequately described only by means of nonlinear models. Such filters are often constructed by linearizing locally the system model and then applying a Kalman filter to this model.

I. INTRODUCTION

A. Filtering

Originally the term *filtering* was adopted in electrical engineering to refer to that aspect of signal processing which is concerned with the removal from signal of unwanted components such as noise. Until the advent of digital technology, especially of digital computers, filtering was performed exclusively by means of electrical circuits or devices called filters. A filter was tagged as linear or nonlinear depending on the type of circuits or devices (linear or nonlinear) from which it was built.

Presently the term *filtering* also refers to one of the three basic modes of information processing: filtering or estimation, smoothing, and prediction. Consider, for example, the description of the state of a time-varying system expressed in terms of a scalar or a vector quantity s. We call s the system state. Suppose that this system is observed (measured) by means of noisy measurements m, which, like s, may also be scalar- or vector-valued. (The letter t in this article always represents the time viewed as a variable. Specific values of t are indicated by means of a superscript or a subscript appended to t; in particular, t_0 always denotes the initial time.) We assume that s and m are related for all $t \geq t_0$ through the expression

$$m(t) = h[s(t)] + v(t),$$

where h is a function mapping $s(t)$ into $[m(t) - v(t)]$ and v represents the unobservable measurement noise (zero mean random error). For example, consider a particle moving in the xy plane. Assume that the state of the particle at time t is adequately described by means of its position coordinates (x, y) and the corresponding velocity components $(\dot{x}, \dot{y})$. Then the state vector $s \equiv s(t)$ can be written as $s^{\mathrm{T}} = (x, y, \dot{x}, \dot{y})$. Suppose that the system is observed at discrete time epochs $t_0, t_1, t_2, \ldots$ and that at each t_j one measures the polar coordinates (r, θ) of the particle. Thus, the measurement equation can be written

$$r = (x^2 + y^2)^{1/2} + v_1 \qquad \text{and} \qquad \theta = \arctan(y/x) + v_2,$$

where v_1 and v_2 represent the noise associated with the measurements of r and θ, respectively.

We use the terms *filtering* and *estimation* synonymously to describe inference about the value $s(t')$ of s at epoch t', which utilizes all information contained in measurements $m(t)$ accumulated over the period $t_0 \leq t \leq t'$. The term *smoothing* refers to recovery of the value $s(t')$ of s at time $t'(<t_e)$ from measurements $m(t)$ accumulated over the period $t_0 \leq t \leq t_e$. Finally, the term *prediction* refers to inference about the value $s(t')$ of s at time $t'(>t_e)$ made on the basis of measurements $m(t)$ accumulated over the period $t_0 \leq t \leq t_e$. This terminology is not completely universal; some authors use the term *estimation* to refer collectively to all three above-defined activities—filtering, smoothing, and prediction.

In general we write $\hat{s}_{t'|t_e}$ to denote the value of $s(t')$ inferred from the measurements $m(t)$ accumulated over the time period $t_0 \leq t \leq t_e$. Thus, $\hat{s}_{t'|t_e}$ represents an estimate (or a filtered value) of $s(t')$ if $t_e = t'$, a smoothened value of $s(t')$ if $t_e > t'$, and a predicted value of $s(t')$ if $t_e < t'$. In the case of measurements sampled at discrete epochs t_k $(k = 0, 1, 2, \ldots)$, if $t' = t_j$ and $t_e = t_h$, we simplify this notation by writing $\hat{s}_{j|h}$.

B. Historical Connections

The origins of the Kalman filter can be traced back to the ideas on least squares estimation formulated by K. F. Gauss. (Gauss first applied the estimation method around 1795, but his first published account of it appeared only around 1810.) He considered the problem of estimating essentially constant parameters of planetary orbits from noisy observations and derived a recursive estimation procedure for updating past estimates, without having to discard them, on the basis of newly available measurements. However, his recursive least-squares estimation method did not apply to a situation in which the estimated parameters would dynamically evolve over time. The extension of recursive estimation to dynamic system models is relatively recent.

Between the 1930s and the 1960s three major developments made a strong impact on the state of signal processing: (a) the advent of digital technology (digital circuits, devices, and stored-program computers); (b) advances in mathematics and in its application to the theory of linear systems, in particular, to communication and control systems; and (c) the infusion of statistical and probabilistic ideas into the theory of these systems. These advances were further stimulated by the demands of World War II and the geopolitical situation after the war. The Wiener filter is a well-known example from that era.

With the emergence of digital technology after World War II, it became possible for the first time and economically attractive to process complex signals digitally in real time. These advances in hardware had a revolutionary effect on both the theory and the implementation of filters. Software started to replace hardware as the medium of implementation; research shifted toward construction of algorithms for digital computing. The fast Fourier transform and the Kalman filter are examples of algorithms that were specifically designed for or naturally lend themselves to implementation on a digital computer. (To be precise, one should conceptually distinguish between an algorithm, the function it computes or approximates, and the applications for which it is used. Viewed as a mathematical function, the filters considered in this article are stochastic estimators of the system state. However, we often refer to them as algorithms and use expressions such as "the Kalman filter algorithm," especially when the function itself clearly indicates a computational procedure.)

The period between the turn of the century and the beginning of World War II witnessed important advances in the evolution of mathematics, especially in the generalization and abstraction of mathematical concepts, which later had an impact on the development of estimation theory and digital technology. Algebra, topology, functional analysis, and mathematical logic are a few examples of these developments. Two noteworthy developments from that era that were later used in filtering theory were (a) the formulation by A. N. Kolmogorov of the axiomatic theory of probability, based on the new theories of measure and (Lebesque) integration, and (b) the emergence of a generalized theory of linear spaces and operators. Mathematical measure and integration theories unified the conceptual setup of discrete and continuous probability spaces; they also furnished powerful tools, in the form of convergence theorems, for dealing with sequences of random variables. Axiomatic probability theory very soon led to a flurry of results in the theory of stochastic processes and in the application of this theory to practical areas such as time-series analysis.

C. Wiener Filter

The injection of probabilistic and statistical ideas into the filtering and control theories not only changed the view of how a system driven by randomly perturbed signals and observed by means of noisy measurements should be modeled, but also provided the tools for estimating the behavior of such systems. Wiener filtering stands as a major milestone in this process of adoption of the statistical approach to filtering and control. The basic problem considered in the 1940s by N. Wiener was as follows. Suppose that $\{x_t\}$ and $\{y_t\}$, where $-\infty < t < \infty$, represent two zero-mean, jointly wide-sense stationary stochastic processes with *known* joint autocovariances and cross-covariances. More precisely stated,

$$y_t = x_t + v_t, \qquad -\infty < t < t_f,$$

where

$$\begin{aligned} E\left[v_t v_s^{\mathrm{T}}\right] &= I_p \delta(t-s), \\ E\left[x_t v_s^{\mathrm{T}}\right] &= \text{arbitrary}, \qquad t \geq s, \\ &= 0, \qquad\qquad t < s, \end{aligned}$$

and $\delta(\cdot)$ represents the Dirac delta function defined by the property

$$\delta(t) = 0 \qquad \text{for} \qquad t \neq 0 \qquad \text{and}$$

$$\int_{-\infty}^{\infty} \delta(t)\, f(t)\, dt = f(0) \tag{1}$$

for all functions f continuous at 0.

Here y_t, x_t, and v_t represent p-valued vectors (Wiener originally considered only the scalar-valued case with $p = 1$) and I_p is a $p \times p$ identity matrix. Given sample observations $\{y_t(\omega), t_0 < t < t_f\}$ of process $\{y_t\}$, the problem is to find the MMSE estimate $\hat{x}(t_f + \Delta t \mid t_f)$ of x at time $t_f + \Delta t$ for some fixed Δt (≥ 0). It turns out that the task is to find a function $h(t)$ so that, for $t_0 < t_f < \infty$,

$$\hat{x}(t_f + \Delta t) = \int_{t_0}^{t_f} h(t_f - t') y(t')\, dt',$$

and the expectation $E[(\hat{x}(t_f + \Delta t) - x(t_f + \Delta t))^2]$ is minimized.

Wiener explicitly solved the above-stated estimation problem under the assumptions of a scalar observation process ($p = 1$) and a semiinfinite observation time interval ($t_0 = -\infty$). He handled this problem in the frequency domain by showing how to find a closed-form solution to a causal transfer function, say H, of h.

Actually, the optimal filtering theory, now known as the Wiener filter (WF), was independently and, at about the same time, developed by both N. Wiener and A. N. Kolmogorov. Whereas the continuous-time theory is mainly Wiener's work, Kolmogorov is credited with the

development of its discrete-time version. He introduced the notion of innovation sequence (a discrete-time white stochastic process associated with the measurement process, which we elaborate in our discussion of the Kalman filter) as a technique for orthogonalizing the data and then exploited the idea of orthogonal projections to simplify the estimation problem. These ideas were utilized years later to attain another view and deepen the understanding of the Kalman filter.

The publication of Wiener's work after World War II triggered a flurry of research in the 1950s and early 1960s. Attempts were made with some success to weaken (generalize) the underlying mathematical assumptions and to derive filtering schemes that would satisfy the requirements of the practical applications of that time period. Consequently, results were obtained extending the WF to nonstationary processes, to vector-valued processes, and to finite time intervals of observation. Despite these theoretical successes, certain difficulties with the WF remained. More precisely, (a) the results were mathematically complicated and thus were difficult to implement while using the signal-processing hardware of that period; (b) the WF did not lend itself easily to recursive estimation in which the estimates would be updated (improved) with each new batch of data—for example, with each pass of an earth satellite over the view of a tracking station; (c) implementational difficulties with the estimation of vector-valued processes persisted. Furthermore, for some applications, certain the oretical assumptions made in the formulation of the WF (such as the wide-sense stationarity of stochastic processes, the existence of spectral densities, and the knowledge of second-order moments) fail to hold. For example, the stochastic components of dynamic systems typically are nonstationary. Most important, the WF, being a product of the era of analog hardware, did not lend itself easily to the newly emerging mode of processing on a stored-program digital computer with finite memory and limited throughput capacity. The aerospace signal-processing problems in the 1950s started to point to the need for a robust algorithm for recursive estimation of vector-valued processes in dynamic systems, with the observations made and estimates updated at discrete time epochs. Furthermore, such an algorithm was to be implemented for real-time processing on a still primitive digital computer. The Kalman filter (KF) turned out to be an answer to these needs.

D. Transition to State Variable-Based Filtering

Although the use of state variables rather than impulse response is now largely credited to R. Kalman, P. Swerling actually was the first person (in a 1959 journal article) to describe an estimation scheme that essentially was identical to that presented the next year by Kalman. One reason for crediting Kalman with what today is known as the Kalman filter was that subsequently Kalman extended his results considerably beyond those obtained by Swerling.

In addition to practical applications of linear filtering theory developed in the 1960s, foundations for nonlinear estimation were laid during the same period. At the time when Kalman's (and also R. C. Bucy's) work was becoming known in the West, R. L. Stratonovich in the Soviet Union published results on his pioneering work in nonlinear estimation, that is, recursive estimation of the states of a nonlinear system driven by white noise. He also looked into the linearized system model and, while doing this, obtained the KF equations in 1960. Though Stratonovich's work in nonlinear estimation remained unknown for a while in the West, some of his results were independently replicated in early 1960s by H. J. Kushner and W. M. Wonham.

II. LINEAR ESTIMATION PROBLEM AND KALMAN FILTERS

The KF is a recursive, unbiased, MMSE estimator of the state s of a dynamic system observed (i.e., measured) by means of noisy measurements m. In general, both s and m are vector-valued. This system is assumed to be linear in the sense that (a) the dynamics of state s is described in terms of linear differential or difference equations that may be additively perturbed by a random process, called process noise; and (b) noisy measurements m are related to the state s by means of a linear transformation.

A. Discrete-Time Linear System Model

The system model used by Kalman is roughly of the following form. For $t_{k+1} = t_k + \Delta t_k$ $(\Delta t_k > 0$ and $k = 0, 1, 2, \ldots)$,

$$s_k = \Phi_{k-1} s_{k-1} + B_{k-1} u_{k-1} + G_{k-1} w_{k-1}, \qquad k = 1, 2, 3, \ldots, \quad (2)$$

and

$$m_k = H_k s_k + v_k, \qquad k = 0, 1, 2, \ldots. \quad (3)$$

Equation (2) is called the system dynamics equation, and Eq. (3) the measurement equation. Subscript k refers to the kth sampling epoch t_k. Then s_k denotes the value of state at time t_k. $\{w_k, k = 0, 1, 2, \ldots\}$, or simply written $\{w_k\}$, is a random sequence (a discrete-time stochastic process) that represents the process noise, also known as the input, driving, or plant noise of process $\{s_k\}$. The $\{m_k\}$ represents the measurements, also known as the ouput process,

which at time t_k are perturbed by random noise v_k. Matrix (operator) Φ_{k-1}, which propagates the state from time t_{k-1} to t_k, is called the system state transition matrix. The optional term $B_{k-1}u_{k-1}$ represents deterministic control inputs with $\{u_k\}$ being a known sequence. Matrix H_k defines the state-to-measurement transformation at t_k. The subscripts attached to Φ, B, G, and H indicate that these matrices may vary over time. It is assumed that these matrices are a priori known; that is, they do not depend on estimates of s or on measurements m.

As already noted the random processes $\{w_k\}$, $\{v_k\}$, $\{s_k\}$, and $\{m_k\}$ usually are vector-valued. We use n_s to denote the length of vector s_k and n_m to denote the lengths of v_k and m_k. As a special case we may have $n_s = 1$ or $n_m = 1$.

Notationally, we do *not* distinguish between scalar-valued and vector-valued quantities. Uppercase letters, English or Greek, are reserved mainly for matrices (operators). We also notationally do not distinguish between the random variables (RVs), or random functions, and the specific values taken on by these RVs. For example, depending on the context, w_k denotes either the RV representing the process noise at t_k or a specific value taken by this RV.

The system model given by Eqs. (2) and (3) is incomplete unless we specify the underlying probabilistic assumptions, which are as follows.

ASSUMPTION 1D. The process noise $\{w_k\}$ is a zero-mean, Gaussian (normal), white process with the covariance matrix

$$E\left[w_k w_j^{\mathrm{T}}\right] = Q_k\,\delta_{kj}, \tag{4}$$

where Q_k is a *known* $n_s \times n_s$ nonnegative definite matrix and δ_{kj} denotes the Kronecker delta, that is, $\delta_{kj} = 1$ if $k = j$; otherwise, $\delta_{kj} = 0$. (If $\{w_k\}$ is absent from the system model for some or all k, then for each such k, Q_k may be thought of as being a zero matrix.)

ASSUMPTION 2D. The measurement noise $\{v_k\}$ is a zero-mean, Gaussian, white process with the covariance matrix

$$E\left[v_k v_j^{\mathrm{T}}\right] = R_k\,\delta_{kj}, \tag{5}$$

where R_k denotes a *known* $n_m \times n_m$ positive definite matrix, and δ_{kj} is the Kronecker delta.

ASSUMPTION 3D. For all k and j, v_k and w_j are statistically independent RVs.

ASSUMPTION 4D. The value s_0 of system state s at the starting time t_0 is a Gaussian RV whose mean $\bar{s}_0$ and convariance matrix Σ_0, that is,

$$\bar{s}_0 \equiv E[s_0] \qquad \text{and} \qquad \Sigma_0 \equiv E\left[(s_0 - \bar{s}_0)(s_0 - \bar{s}_0)^{\mathrm{T}}\right] \tag{6}$$

are known. Furthermore, the RV s_0 is statistically independent of the the RVs w_k and v_k for all $k \geq 0$.

In Assumptions 1D–4D, the letter "D" indicates discrete time. In this article, a discrete-time white process is defined to be a discrete-time stochastic process, say $\{x_i\}$, for which x_k and x_j are statistically independent RVs whenever $k \neq j$. Collectively, all a priori distributional assumptions concerning the random components (s_0, $\{w_k\}$, and $\{v_k\}$) of the model are referred to as the a priori probability assumptions or, simply, as the a priori probabilities. All above-stated assumptions of Gaussian distribution, taken together, are referred to as the Gaussianity assumption. As a relaxation of the original assumptions in Wiener filtering, Assumptions 1D–4D do not require $\{w_k\}$ or $\{v_k\}$ to be a stationary process in the sense that matrices Q_k and R_k may be time dependent.

B. Properties

The system model defined in terms of Eqs. (2) and (3), restricted by the distributional assumptions (1D–4D), has the following easily verifiable properties.

a. System state s_k (for $k = 1, 2, \ldots$) can be expressed as

$$s_k = \Phi_{k,0}s_0 + \sum_{i=0}^{k-1} \Phi_{k,i+1}(B_i u_i + G_i w_i),$$

where the multistep state transition matrix $\Phi_{k,j}$, defined in terms of the single-step state transition matrix Φ_i as $\Phi_{i+1,i} \equiv \Phi_i$, possesses the following properties:

(i) $\Phi_{k,k} \equiv I_s (= n_s \times n_s$ identity matrix),

(ii) $\Phi_{k,j} = \Phi_{k-1}\Phi_{k-2}\cdots\Phi_j \quad$ for $\quad k > j \geq 0$, and

(iii) $\Phi_{k,i} = \Phi_{k,j}\Phi_{j,i} \quad$ for all $\quad k \geq j \geq i \geq 0$.

b. Each s_k is a multivariate Gaussian RV of length n_s; actually, we can say more: The process $\{s_k\}$ is a Gaussian random process.

c. $\{s_k\}$ is a Markov process, which implies that for any $0 \leq k_1 < k_2 < \cdots < k_n < k$, the conditional probability (density) function of s_k given $s_{k_1}, s_{k_2}, \ldots, s_{k_n}$ has the Markov property

$$p\left(s_k \,\middle|\, s_{k_1}, s_{k_2}, \ldots, s_{k_n}\right) = p\left(s_k \,\middle|\, s_{k_n}\right).$$

d. The measurement process $\{m_k\}$ is a Gaussian process, but it is *not* a Markov process.

e. Processes $\{s_k\}$ and $\{m_k\}$ jointly form a Gaussian process $\{(s_k^{\mathrm{T}}, m_k^{\mathrm{T}})^{\mathrm{T}}\}$.

The preceding properties of the system model are used in the derivation of the KF algorithm and in proving its

optimality properties. Recursiveness of the KF algorithm essentially is a consequence of (Markov) property c.

The first two moments of $\{s_k\}$ and $\{m_k\}$ follow directly from properties a–e. First, consider the process $\{s_k\}$ and, for $k \geq j \geq 0$, define

$$\begin{aligned} \bar{s}_k &\equiv E[s_k], \\ \Sigma_{k,j} &\equiv E\big[(s_k - \bar{s}_k)(s_j - \bar{s}_j)^{\mathrm{T}}\big], \\ \Sigma_k &\equiv \Sigma_{k,k}. \end{aligned}$$

It follows from properties a and b that

$$\begin{aligned} \bar{s}_k &= E[s_k] = \Phi_{k,0}\bar{s}_0 + \sum_{i=1}^{k-1} \Phi_{k,i+1} B_i u_i \quad \text{for} \quad k \geq 0, \\ \bar{s}_{k+1} &= \Phi_k \bar{s}_k + B_k u_k \qquad \text{for} \qquad k \geq 0, \\ \Sigma_k &= \Phi_{k,0}\Sigma_0(\Phi_{k,0})^{\mathrm{T}} + \sum_{i=0}^{k-1} (\Phi_{k,i+1}G_i) Q_i (\Phi_{k,i+1}G_i)^{\mathrm{T}}, \\ \Sigma_{k+1} &= \Phi_k \Sigma_k \Phi_k^{\mathrm{T}} + G_k Q_k G_k^{\mathrm{T}}, \\ \Sigma_{k,i} &= \Phi_{k,i}\Sigma_i \qquad \text{if} \qquad k \geq i. \end{aligned}$$

The first two moments of measurement process $\{m_k\}$ turn out to be

$$\bar{m}_k \equiv E[m_k] = H_k \bar{s}_k$$

and

$$E\big[(m_k - \bar{m}_k)(m_i - \bar{m}_i)^{\mathrm{T}}\big] = H_k \Phi_{k,i} \Sigma_i H_i^{\mathrm{T}} + R_k\, \delta_{kj} \qquad \text{for} \qquad k \geq i \geq 0,$$

where δ_{kj} is the Kronecker delta defined following Eq. (4). The above-listed expressions for the first two moments of s_k and m_k remain valid even if the initial state s_0 or at least one of the noise processes, $\{w_k\}$ or $\{v_k\}$, is non-Gaussian.

C. Discrete-Time Filtering Problem

In this subsection as well as in the next, it is important to view the measurements m_k, the state s_k, and the state estimator as RVs but not as the values taken on by these RVs. Let the RV $M_k \equiv \{m_0, m_1, \ldots, m_k\}$ represent the set of all measurements obtained over the time period $t_0 \leq t \leq t_k$. Use $\hat{s}_{k|j}$ to denote an estimator of s_k that is a function of (is conditioned on) M_j. As noted earlier, we have an estimation problem if $k = j$, a prediction problem if $k > j$ (in such a case, $\hat{s}_{k|j}$ is more appropriately called the predictor of s_k), and a smoothing problem if $k < j$. The KF is concerned only with the first two problems.

The optimality criterion used in the derivation of the KF algorithm is minimization of mean square error. Thus, the KF is an estimator $\hat{s}_{k|j}$ conditioned on M_j, which minimizes

$$E\big[(s_k - \hat{s}_{k|j})^{\mathrm{T}}(s_k - \hat{s}_{k|j}) \,\big|\, M_j\big] \qquad \text{for} \qquad k = 0, 1, 2, \ldots. \tag{7}$$

It can be shown that such $\hat{s}_{k|j}$ is of the form

$$\hat{s}_{k|j} = E[s_k \mid M_j]. \tag{8}$$

In the literature, the MMSE estimator is also alternatively known as the minimum variance (MV) or the least-squares (LS) estimator. The latter name is technically imprecise. Furthermore, these names are somewhat misleading because they fail to indicate that the resulting estimates depend on a priori probability assumptions and that these estimates are conditional in nature. Also note that the MMSE criterion can be applied to a more general system model than one defined by Eqs. (2) and (3) and Assumptions 1D–4D. Even then the minimization of Eq. (7) implies the conditional mean estimator of Eq. (8). However, the resulting estimator may not be mathematically or implementationally tractable.

Given the starting conditions $\bar{s}_0$, Σ_0, and the initial measurements M_0 (=$\{m_0\}$), the KF provides a procedure for computing recursively at each t_k the values of

$$\hat{s}_{k|k-1} = E[s_k \mid M_{k-1}] \quad \text{and} \quad \hat{s}_{k|k} = E[s_k \mid M_K] \tag{9}$$

and the associated error covariance matrices

$$P_{k|k-1} \equiv E\big[(s_k - \hat{s}_{k|k-1})(s_k - \hat{s}_{k|k-1})^{\mathrm{T}} \,\big|\, M_{k-1}\big] \tag{10a}$$

and

$$P_{k|k} \equiv E\big[(s_k - \hat{s}_{k|k})(s_k - \hat{s}_{k|k})^{\mathrm{T}} \,\big|\, M_k\big]. \tag{10b}$$

One noteworthy property of the resulting KF estimator is that it is *unbiased*. This follows from Eq. (8):

$$\begin{aligned} E[s_k - \hat{s}_{k|j}] &= E[s_k] - E[E[s_k \mid M_j]] \\ &= E[s_k] - E[s_k] = 0. \end{aligned}$$

The MMSE and unbiasedness properties of the KF imply that it also is an MV estimator. Another important property of the KF estimator is that it is *linear* in the sense that the estimators $\hat{s}_{k|k-1}$ and $\hat{s}_{k|k}$ are of the form

$$\hat{s}_{k|k-1} = \Phi_{k-1}\hat{s}_{k-1|k-1} + B_{k-1}u_{k-1}$$

and

$$\hat{s}_{k|k} = K_k(m_k - H_k\hat{s}_{k|k-1}) + \hat{s}_{k|k-1},$$

where K_k is the so-called Kalman gain matrix.

What happens if the Gaussianity assumptions for $\{w_k\}$, $\{v_k\}$, or s_0 are dropped? Then the KF algorithm still provides the best and unbiased estimator among all linear

estimators of the state. It is best in the sense that it produces the smallest mean square error among all linear estimators of the state. However, in such a situation there may exist a nonlinear filter which, in the MMSE sense, is better than the KF.

D. Alternative Approaches to Discrete-Time Linear Filtering

The MMSE optimality criterion [Eq. (7)] is not the only one used in recursive estimation. The LS principle of residual minimization, the maximum-likelihood (ML) principle, and various optimality principles used in conjunction with the Bayes method are examples of other possibilities. In the case of a linear estimation problem [i.e., for a system model of the form of Eqs. (2) and (3)], if the initial conditions and the noise processes have Gaussian distributions, as is implied by Assumptions 1D–4D, the methods of LS (with appropriate weights), ML, and Bayes (with a quadratic performance criterion) yield estimators of the same (KF) type.

The weighted LS method, which can be expressed in recursive form for sampled measurements [Eq. (3)], is the simplest of the three above-mentioned methods because it does not require making any explicit assumptions, with one exception, about the distribution of RVs involved in the estimation process. The exception concerns the assumption of a priori knowledge of the process and measurement noise covariance matrices, the inverses of which are used as weights in the LS estimation procedure. The optimality criterion at time t_k is the minimization of the sum of squares of the weighted residuals (i.e., the weighted differences between the measurements and their values predicted by the model) accumulated over the epochs $t_0, t_1, \ldots, t_k$. As already noted, this approach under the Gaussianity assumptions for noise processes leads to an estimator of KF type.

The LS method (recursively or nonrecursively formulated, weighted or unweighted) has been widely used since Gauss to estimate the parameters of static systems. A static system may be thought of as a special case of the system model [Eqs. (2) and (3)] in which Φ_k is an $n_s \times n_s$ identity matrix, and B_k and G_k are zero matrices, that is, the system is time invariant and so Eq. (2) is not needed to describe the system dynamics. The weighted or unweighted LS method, usually expressed in nonrecursive form, constitutes the backbone of regression analysis and currently is one of the principal methods in statistical inference.

The ML method has also been used for years in statistical inference as a point estimation technique. It is based on viewing the probability (density) function $p_s(M_k)$ of measurements $M_k = \{m_0, m_1, \ldots, m_k\}$ as a function $\lambda_s \equiv p_s(M_k)$ of the unknown parameters s (in our case, s represents the unknown system state) on which the form of this function depends. Then, given measurements M_k, the ML estimate of s is that value, say s^*, of s that maximizes λ_s. This principle has been extended to the construction of recursive estimators for linear dynamic systems of the type defined by Eqs. (2) and (3). As already noted, under Assumptions 1D–4D, the resulting recursive estimator is identical to the KF.

In engineering literature, an estimator derived from the Bayes conditional probability inversion formula

$$p(x \mid y) = p(x, y)/p(y) = p(y \mid x)p(x)/\int p(y \mid x)p(x)\,dx \quad (11)$$

by applying some optimality criterion is often, although somewhat imprecisely, called the Bayes estimator of x. More precisely, the resulting estimator is a Bayes estimator only if, in addition, the $p(x)$ appearing on the right-hand side of Eq. (11) is a Bayes prior (also known as Bayes a priori) probability of x. The term *prior* here means that $p(x)$ is postulated prior to observing the value of y. In estimation, y and x would represent the measurements and the unknown parameters (state), respectively. Hence, in the problems considered here, y and x would be replaced by variously subscripted M (or m) and s, respectively. The left-hand side of the Bayes inversion formula would then give the posterior (also known as Bayes a posteriori) probability of unknown parameters (state) s conditioned on measurements M (or m).

Using Eq. (11) and the RV independence conditions implied by Assumptions 1D–4D, and also noting that

$$M_k \equiv \{m_0, m_1, \ldots, m_{k-1}, m_k\} = \{M_{k-1}, m_k\},$$

one can derive the so-called Bayes probability recursion formulas,

$$p(s_k \mid M_k) = p(s_k \mid M_{k-1})p(m_k \mid s_k)/p(m_k \mid M_{k-1}) \quad (12a)$$

and

$$p(s_k \mid M_{k-1}) = \int p(s_k \mid s_{k-1})p(s_{k-1} \mid M_{k-1})\,ds_{k-1}, \quad (12b)$$

with the denominator in Eq. (12a) expressible as

$$p(m_k \mid M_{k-1}) = \int p(m_k \mid s_k)p(s_k \mid M_{k-1})\,ds_k. \quad (12c)$$

Equations (12a) and (12b) recursively define the measurement updating and the time propagation (also called time updating) of probability densities for $k = 0, 1, \ldots$. To get the recursive process started at t_0, one must interpret correctly the meaning of $p(s_0 \mid M_{-1})$. Since M_{-1} constitutes an empty set (i.e., no measurements), $p(s_0 \mid M_{-1})$ represents the Bayes prior $p(s_0)$. As an example, this prior

for a KF is specified in terms of the first two moments, $\bar{s}_0$ and Σ_0, of the Gaussian distribution of state s at the initial time t_0. It is worth noting that, in Eqs. (12a)–(12c), $p(s_k \mid s_{k-1})$ is obtained from the state time-propagation equations (2) with the assistance of the a priori known probability (density) $p(w_{k-1})$ of process noise w_{k-1} at t_{k-1}. Similarly, $p(m_k \mid s_k)$ is determined from the measurement equations (3) and with the assistance of the a priori known probability (density) $p(v_k)$ of measurement noise v_k at t_k. Knowledge of all these probability densities and $p(s_0)$ recursively determines $p(s_k \mid M_k)$ for all $k = 0, 1, \ldots$.

Using the recursive probability equations (12a)–(12c) in conjunction with various optimality criteria produces a variety of recursive estimation schemes. However, due mainly to mathematical and implementational considerations, the MMSE criterion, equivalent to Eq. (7) and yielding an estimator of the type in Eq. (8), is the most widely used criterion at present. As already noted, this criterion is applied to derive the KF equations.

E. Discrete-Time Kalman Filter

Given the discrete-time linear system model (including the distributional Assumptions 1D–4D) and the defining relation of Eq. (8) there are several ways in which the discrete-time KF algorithm can be derived. One technique, which is mathematically straightforward although tedious in detail, is based on the use of recursive probability equations (12a)–(12c) and the repeated application of conditional multivariate Gaussian distribution to express the probability densities $p(s_k \mid M_{k-1})$ and $p(s_k \mid M_k$ for all $k = 0, 1, \ldots$ in terms of m_k, $\hat{s}_{k-1|k-1}$, $P_{k-1|k-1}$, the matrices appearing in the system model, and the covariances of noise processes $\{w_k\}$ and $\{v_k\}$. Due to the regenerative properties of Gaussian distribution, densities $p(s_k \mid M_{k-1})$ and $p(s_k \mid M_k)$ remain Gaussian for all $k \geq 0$. By Eqs. (9), (10a), and (10b), the first two moments of $p(s_k \mid M_{k-1})$ and $p(s_k \mid M_k)$ are $\hat{s}_{k|k-1}$ with $P_{k|k-1}$ and $\hat{s}_{k|k}$ with $P_{k|k}$, respectively. Hence, the mathematical expressions for these quantities, giving the recursive estimation equations (13)–(18), are explicitly obtainable from the mathematical expressions for $p(s_k \mid M_{k-1})$ and $p(s_k \mid M_k)$.

Another technique for deriving the KF algorithm uses the notion of innovations, defined as

$$r_k = m_k - E[m_k \mid m_{k-1}] = m_k - H_k \hat{s}_{k|k-1}.$$

In the innovation process $\{r_k\}$, each r_k constitutes that part of the measurement at t_k that contains new information not provided by $m_0, m_1, \ldots, m_{k-1}$. Innovations $\{r_k\}$ constitute a zero-mean white process. It follows that r_k represents the error in the orthogonal projection of s_k on the subspace generated by $M_k = \{m_0, m_1, \ldots, m_k\}$. This projection, being a linear combination of the vectors in M_k, defines a linear MMSE estimator of s. Using this notion of orthogonality, the whiteness of $\{r_k\}$, and the system model [Eqs. (2) and (3)], especially the Markov property of state propagation, it is then rather simple to derive the KF recursion expressions.

Using the preceding system model and probabilistic assumptions, one can state the KF algorithm for a discrete-time linear system as follows.

Discrete-Time Kalman Filter Algorithm. At each $t_k = t_{k-1} + \Delta t_{k-1} (k = 0, 1, 2, \ldots)$, proceed as follows.

a. If $k = 0$ (i.e., at the initial time t_0), initialize the estimation process by setting

$$\hat{s}_{0|-1} = \hat{s}_0 \qquad \text{and} \qquad P_{0|-1} = \Sigma_0, \tag{13}$$

where $\bar{s}_0$ and Σ_0 are defined by Eq. (6).

b. If $k \geq 1$, propagate the estimates of system state vector s and state error covariance matrix P from t_{k-1} to t_k by computing

$$\hat{s}_{k|k-1} = \Phi_{k-1} \hat{s}_{k-1|k-1} + B_{k-1} u_{k-1} \tag{14}$$

and

$$P_{k|k-1} = \Phi_{k-1} P_{k-1|k-1} \Phi_{k-1}^{\mathrm{T}} + G_{k-1} Q_{k-1} G_{k-1}^{\mathrm{T}}, \tag{15}$$

respectively.

c. Use measurements $m_k = H_k S_k + v_k$ to update the propagated estimate of s by computing the Kalman gain matrix

$$K_k = P_{k|k-1} H_k^{\mathrm{T}} \left[H_k P_{k|k-1} H_k^{\mathrm{T}} + R_k \right]^{-1} \tag{16}$$

and the update of the state estimate

$$\hat{s}_{k|k} = \hat{s}_{k|k-1} + K_k [m_k - H_k \hat{s}_{k|k-1}]. \tag{17}$$

Next, update the estimate P of state error covariance matrix by computing

$$P_{k|k} = [I - K_k H_k] P_{k|k-1}. \tag{18}$$

The preceding algorithm assumes that one knows the values of $\bar{s}_0$ and Σ_0 and also knows how to compute for all k the matrices Φ_k, B_k, R_k, Q_k, and H_k. In the literature, $\hat{s}_{k|k}$ is often called the measurement update of $\hat{s}_{k|k-1}$ and $\hat{s}_{k+1|k}$ the time update of $\hat{s}_{k|k}$.

If the state estimate update equation (17) is rewritten

$$\hat{s}_{k|k} = K_k m_k + [I - K_k H_k] \hat{s}_{k|k-1}, \tag{17$'$}$$

an equivalent expression for $P_{k|k}$ is

$$P_{k|k} = K_k R_k K_k^{\mathrm{T}} + [I - K_k H_k] P_{k|k-1} [I - K_k H_k]^{\mathrm{T}}, \tag{18$'$}$$

which is known as the Joseph form (named after P. D. Joseph) of error covariance update. Since the right-hand side of Eq. (18′) is the sum of two symmetric matrices (the

first a nonnegative definite, the second a positive definite matrix), the state error covariance update in this form numerically is more stable, although computationally more expensive, than the same operation in the short form [Eq. (18)].

EXAMPLE 1. FILTERING OF AN AUTOREGRESSIVE PROCESS. This example illustrates the application of the discrete-time KF to a simple linear system in which both the state s and the measurement m are scalars. The system model is of the form

$$s_k = \phi s_{k-1} + w_k \qquad \text{and} \qquad m_k = h s_k + v_k,$$

where ϕ and h are scalar multipliers, that is, $n_s = n_m = 1$. Both noise processes, $\{w_k\}$ and $\{v_k\}$, and the initial conditions are assumed to satisfy Assumptions 1D–4D. In the present case, the process noise covariance matrices Q and R are 1×1; that is, Q and R represent the time-invariant variances of w_k and v_k, respectively, and are scalar quantities.

The KF equations now reduce to the following procedure.

a. If $k = 0$, initialize the estimation process by specifying s_0 and $P_{0|-1}$.

b. If $k \geq 1$, propagate the estimates by computing

$$\hat{s}_{k|k-1} = \phi \hat{s}_{k-1|k-1},$$
$$P_{k|k-1} = \phi^2 P_{k-1|k-1} + Q.$$

c. For all $k \geq 0$, update the estimates by computing

$$K_k = h P_{k|k-1} \big/ \left[h^2 P_{k|k-1} + R\right],$$
$$\hat{s}_{k|k} = \hat{s}_{k|k-1} + K_k[m_k - h\hat{s}_{k|k-1}],$$
$$P_{k|k} = K_k^2 R + [1 - hK_k]^2 P_{k|k-1}.$$

Figure 1 summarizes the results of a sample simulation of this KF. Figure 1a shows the generated time history of the true system state s. Figure 1b summarizes the time history of estimation error, which, at t_k, is defined as $e_k = \hat{s}_{k|k} - s_k$. The values of parameters of the true system model are

$$\phi = 1.02, \quad h = 100.0, \quad Q = 4.0, \quad \text{and} \quad R = 1.0.$$

The values of the corresponding parameters in the system model assumed by the KF are equal in value to those of the true model. The true initial value of state is $s_0 = 5.0$. The initialization values used by the KF are $\bar{s}_0 = 4.0$ and $P_{0|-1} = 25.0$.

Since, according to the true system model, the measurements are very precise compared with the level of process noise (i.e., the variance of m/h is $R/h^2 = 0.0001 \ll Q$) and the system model assumed in the KF agrees with the true model, the filter practically converges at t_0 after the first measurement update. One can easily see this by observing in Fig. 1b that (for all $k \geq 0$) the points $\pm(P_{k|k})^{1/2}$ lie on two horizontal lines that are symmetrically located with respect to the time axis.

In many respects, the above-described filtering problem for a random walk is trivial. Its only purpose was to illustrate the KF computations for a very simple (scalar) case. The following "reversion" of this problem has practical applications in time-series analysis and is known as a system identification problem. Suppose that the value of scalar multiplier ϕ is continuously undergoing random perturbations and so is not completely known. Assume that the dynamics of ϕ is adequately described by means of an equation of the form

$$\phi_k = \phi_{k-1} + w'_{k-1}$$

and that we are now directly observing the state s via measurement equation of the form

$$s_k = s_{k-1}\phi_k + v'_k,$$

which essentially is the state transition equation of the original problem with a modified noise term. Making the usual probabilistic assumptions for the noise processes $\{w'_k\}$ and $\{v'_k\}$ and the initial distribution of ϕ completes a system model for the KF of the system identification problem (estimation of the unknown and possibly time-varying values of parameters in the system model).

F. Continuous-Time Kalman Filter

A year after the publication of Kalman's seminal work on discrete-time filtering, R. Bucy and R. Kalman laid the foundation for continuous-time KFs. Although presently the KFs, probably without exception, are implemented as computer programs for processing on a digital computer, strong interest in the continuous-time KFs persists. Often the estimation problem for a stochastic dynamic system can be precisely formulated only in terms of stochastic differential equations. Thus, the continuous-time KF is an important modeling tool to a designer.

The system model for a continuous-time KF is summarized next. (As much as possible, the same notation as in the discrete-time model is used here. Also, to simplify the discussion of the continuous-time KF, the term representing the control inputs has been deleted from the system dynamics model; the presence of such a term would affect in an a priori known way only the mean value of the state, not its covariances.) The dynamics of the system state, at

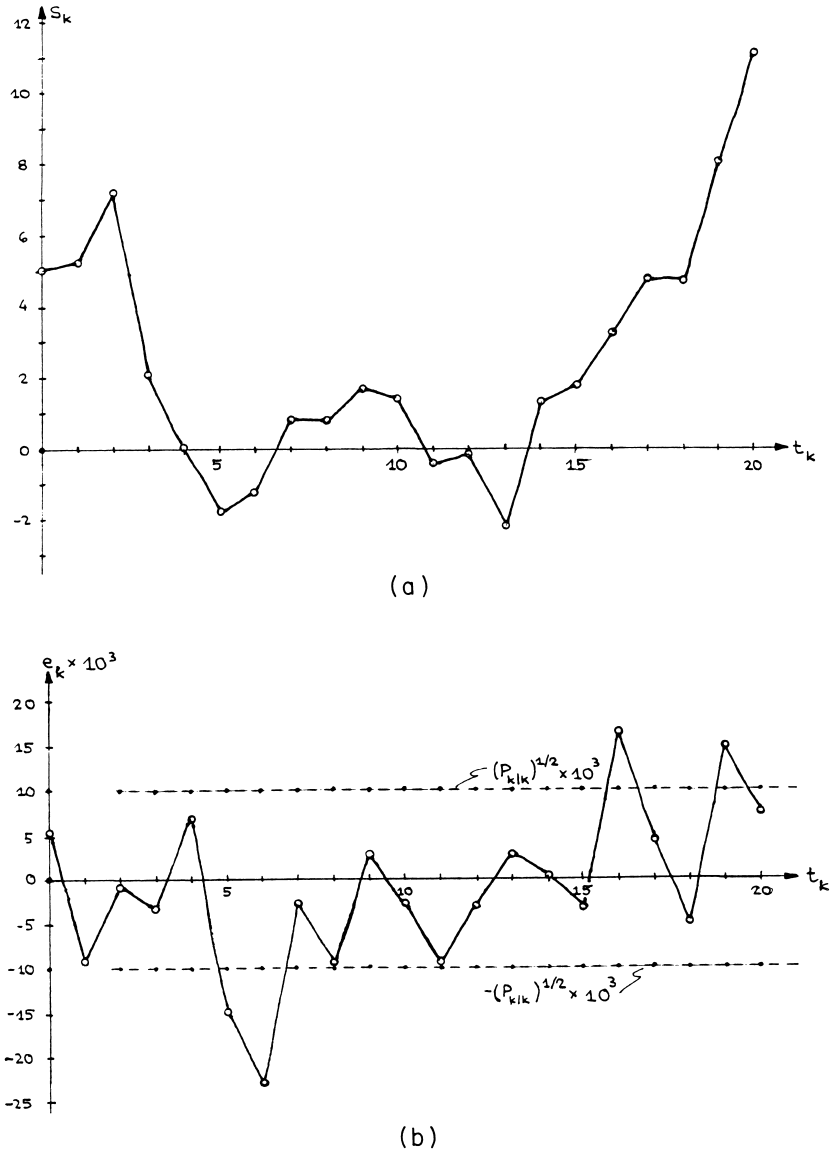


FIGURE 1 Results of a simulation run of the Kalman filter, Example 1. (a) Generated time history of the true system state and (b) time history of estimation error e (multiplied by 10^3).

least *formally*, is described in the continuous-time case by a stochastic differential equation of the form

$$ds_t/dt = F_t s_t + G_{(c)t} w_t, \qquad t \geq t_0. \tag{19}$$

The measurement equation is written

$$m_t = H_t s_t + v_t, \qquad t \geq t_0. \tag{20}$$

The time argument t presently ranges over a continuum (an interval) of real numbers and is, as indicated, finitely bounded from below. Thus, $\{w_t\}$, $\{s_t\}$, $\{v_t\}$, and $\{m_t\}$ are now continuous-time stochastic processes with $t \geq t_0$. With regard to vector and matrix sizes, the assumptions made and the notationused earlier continue to hold. The subscript "c" distinguishes a continuous-time entity from its discrete-time counterpart. The probabilistic assumptions about the model specified by Eqs. (19) and (20) are similar to those stated for the discrete-time version of the KF.

ASSUMPTION 1C. The process noise $\{w_t\}$ is a continuous-time, zero-mean, Gaussian white process with the covariance structure of the form

$$E\left[w_t w_{t'}^{\mathrm{T}}\right] = Q_{(c)t}\, \delta(t - t'), \tag{21}$$

where $Q_{(c)t}$ is a known $n_s \times n_s$ nonnegative definite matrix (which may optionally be a zero matrix) and $\delta(\cdot)$ denotes the Dirac delta function defined by Eq. (1).

ASSUMPTION 2C. The measurement noise $\{v_t\}$ is a continuous-time, zero-mean, Gaussian white process with

$$E\left[v_t v_{t'}^{\mathrm{T}}\right] = R_{(\mathrm{c})t}\,\delta(t-t'), \tag{22}$$

where $R_{(\mathrm{c})t}$ is a known $n_m \times n_m$ positive definite matrix and $\delta(\cdot)$ represents the Dirac delta function defined by Eq. (1).

ASSUMPTION 3C. For all t and $t' \geq t_0$, RVs w_t and $v_{t'}$ are independently distributed RVs.

ASSUMPTION 4C. The value s_0 of s at t_0 is distributed as a Gaussian RV whose mean and covariance matrix, that is,

$$\bar{s}_0 \equiv E[s_0]$$

and

$$\Sigma_0 \equiv E\left[(s_0-\bar{s}_0)(s_0-\bar{s}_0)^{\mathrm{T}}\right], \tag{23}$$

are known. Furthermore, the RV s_0 is statistically independent of the RVs w_t and $v_{t'}$ for all t and $t' \geq t_0$.

Mathematically the stochastic differential equation (19) is not meaningful because of the white noise assumption (Assumption 1C) for the continuous-time process $\{w_t\}$. It is well known, however, that white Gaussian noise is the formal derivative of a Brownian motion process $\{\beta_t\}$, with $t \geq t_0$, which has the property that the increments

$$\left\{\beta_{t_j} - \beta_{t_{j-1}}\right\}, \qquad j \geq 1,$$

are independent Gaussian RVs for any ordered set $t_0 < t_1 < t_2 < \cdots < t_n$ of epochs. Thus, the differential equation (19) can be viewed as being equivalent to

$$ds_t = (F_t s_t)\,dt + (G_{(\mathrm{c})t})\,d\beta_t, \qquad t \geq t_0, \tag{24}$$

and so to

$$s_t - s_{t_0} = \int_{t_0}^{t} F s_\tau\,d\tau + \int_{t_0}^{t} G_{(\mathrm{c})\tau}\,d\beta_\tau. \tag{25}$$

On the right-hand side of Eq. (25), the first integral is ordinary (Riemann or Lebesque), but the second is an Itô stochastic integral. (Integrals of this type were defined in the 1940s by the Japanese mathematician K. Itô.)

Replacement of Eq. (19) by Eq. (24) requires the corresponding modification of Assumptions 1C–4C. In these assumptions, $\{w_t\}$ must be replaced by $\{d\beta_t\}$, where $\{\beta_t, t \geq t_0\}$ is a Brownian motion process and

$$E\left[d\beta_t d\beta_{t'}^{\mathrm{T}}\right] = Q_{(\mathrm{c})}\delta(t-t'). \tag{21$'$}$$

The modified assumption are referred to as Assumptions 1C′–4C′.

Rigorous treatment of continuous-time stochastic processes and, in particular, of stochastic differential equations requires considerably more mathematical sophistication than that of discrete-time stochastic processes (sequences of RVs). Consequently, derivation of a continuous-time KF is technically more difficult than derivation of its discrete-time counterpart. Except for noting that mathematically it is more rigorous to use Eq. (24) in place of Eq. (19), we do not dwell any longer on these theoretical issues and proceed to an intuitive definition of a continuous-time KF.

According to whether the measurements are incorporated continuously over the time or are just sampled at discrete time points, one could distinguish two variations of continuous-time KF: a continuous-dynamics, continuous-measurement KF and a continuous-dynamics, discrete-measurement KF. Since the first is of little interest in practice, we do not consider it in this chapter. Hence, from now on the term *continuous-time KF* implies a continuous-dynamics, discrete-measurement KF.

Continuous-Time (Discrete-Measurement) Kalman Filter Algorithm. For the system model given by Eqs. (24) and (20) and the modified Assumptions 1C′–4C′, the evolution of the conditional mean $\hat{s}_{t|t}$ and the state error covariance matrix $P_{t|t}$, *during* an interval $t_k \geq t > t_{k+1}$ *between* two consecutive measurement sampling epochs t_k and t_{k+1}, satisfies the differential equations

$$d\hat{s}_{t|t}/dt = F_t \hat{s}_{t|t}$$

and

$$dP_{t|t}/dt = F_t P_{t|t} + P_{t|t}F_t^{\mathrm{T}} + G_{(\mathrm{c})t} Q_t G_{(\mathrm{c})t}^{\mathrm{T}}.$$

At a measurement epoch t_k (say $= t$), the estimates of the state vector and state error covariance matrix are

$$\hat{s}_{t|t} = \hat{s}_{t|t-} + K_t[m_t - H_t \hat{s}_{t|t-}] \tag{26}$$

and

$$P_{t|t} = P_{t|t-} - K_t H_t P_{t|t-}, \tag{27}$$

respectively, where

$$K_t = P_{t|t-}H_t^{\mathrm{T}}\left[H_t P_{t|t-} H_t^{\mathrm{T}} + R_{(\mathrm{c})t}\right]^{-1}. \tag{28}$$

Above, $\hat{s}_{t|t-}$ (or $P_{t|t-}$) denotes the value of $\hat{s}$ (or of P) measurement-updated at t_{k-1} and then time-propagated up to $t = t_k$. At the initial time $t = t_0$ the solution process is started while using

$$\bar{s}_0 = E[s_0]$$

and

$$P_{0|0-} = \Sigma_0 = E\left[(s_0-\bar{s}_0)(s_0-\bar{s}_0)^{\mathrm{T}}\right]$$

as the initial conditions, which are assumed to be known.

As in the discrete-time KF, the matrices F_t, $Q_{(\mathrm{c})t}$, and $R_{(\mathrm{c})t}$ are also assumed to be known, but they differ from the corresponding matrices of a discrete-time model even for the same dynamic system.

To handle (on a computer) the integration of differential equations implied by the propagation forward in time of estimates, a common approach is (a) to construct a discrete-time system model corresponding to the continuous-time system model defined by Eqs. (24) and (20) and (b) then to apply the discrete-time KF algorithm to this discrete-time system model. Heuristically this can be done by considering a time interval $t_{k-1} \le t \le t_{k-1} + \Delta t_{k-1} = t_k$ and replacing ds_t, dt, and $d\beta_t$ in Eq. (24) with $s_k - s_{k-1}$, Δt_{k-1}, and $\beta_k - \beta_{k-1}$, respectively, which yields

$$s_k \simeq \Phi_{k-1} s_{k-1} + G_{k-1} w_k$$

if, for all $k = 0, 1, 2, \ldots$, one defines $\Phi_k \equiv \mathrm{I} + \Delta t_k F_k$, $G_k \equiv \Delta t_k G_{(c)k}$, $Q_k \equiv \Delta t_k Q_{(c)k}$, and $w_k \equiv \beta_k - \beta_{k-1}$.

III. MODIFICATIONS OF KALMAN FILTERS AND NONLINEAR FILTERING

A. Filter Divergence

Discrepancies between the actual performance (the actual quality of estimates) of a filter and its apparent performance as indicated by the estimated state error covariance matrices $P_{k|k-1}$ and $P_{k|k}$ are called filter divergence. One variation of this phenomenon, known as apparent divergence, occurs when the actual estimation errors of state, although they remain bounded, become by at least an order of magnitude larger than indicated by P matrices. In such a case, performance of the KF can often be adequately improved by means of small parametric or structural adjustments in the linear system model, by changing the characterization of noise (R and Q matrices) or the initialization values s_0 and P_0 of estimates, or by switching to more accurate numerical procedures. Collectively such improvements are referred to as filter turning. What is important to note is that apparent divergence usually does not require changing from the linear system model or the basic KF algorithm.

A more serious situation, called true divergence, arises when the state error covariance estimates $P_{k|k-1}$ and $P_{k|k}$ remain bounded, while the actual errors in the state estimates start growing without bounds. True divergence may be caused by the numerical instability of filter algorithms or by modeling flaws. For example, certain critical factors or state components have not been included in the model, the system dynamics equation or the measurement-to-state transformation is incorrectly defined, or the noise model is structurally or distributionally incorrect. Numerical instability can usually be removed by resorting to better numerical algorithms. (This problem is briefly discussed in Section IV.) Modeling flaws can be detected by means of simulation in which the filter constructed on the basis of the assumed (say, linear) model is tested against a scenario generated while using a true (say, nonlinear) model. Model inadequacies arising from inherent nonlinearities in the modeled system, if they are not too strong, can often be compensated for by the techniques that do not require complete abandonment of the KF algorithm. One may be able to linearize locally the system model and then use an extended form of the KF algorithm. Otherwise, if they cannot be removed by such techniques, one must resort to nonlinear filtering, which is characterized by the use of nonlinear models for system dynamics and (or) measurements. By comparison with the KF, nonlinear filtering is considerably more difficult to attain both theoretically and practically.

B. Filter Tuning and Modifications

Certain deviations of the assumed linear system model or noise from the actual behavior of the modeled system (from its true model), leading to apparent divergence, can often be compensated for by means of a combination of the following techniques: (a) modifying, typically raising, the assumed levels of process or measurement noise through appropriate changes in noise covariance matrices Q and R; (b) bounding from below the main-diagonal elements of state error covariance matrix P and then appropriately adjusting the off-diagonal elements; (c) introducing the so-called fading (finite) memory filtering or overweighting the most recent measurements; (d) whitening (decorrelating) the process or measurement noise, also removing singularity or near-singularity from matrix R; (e) adding more states; that is, increasing the length of the state vector. The foregoing list of possibly useful measures is by no means complete.

Raising the level of process noise by an appropriate modification of matrix Q often compensates for small nonlinearities and factors, such as states, unaccounted for in the assumed model. Raising or lowering the assumed level of measurement noise by modifying matrix R forces the filter to decrease or increase, respectively, the dependence on measurements.

Bounding from below the estimates of state error variances (the main-diagonal elements of matrix $P_{k|k}$) and then correspondingly adjusting the estimated covariances (the remaining elements of $P_{k|k}$) usually increases the numerical stability of the estimation process and protects the filter from blowing up. For example, the estimation process can easily be destabilized by the cumulative effect of rounding errors that destroys the positive definiteness or symmetry of matrix P. Also, as the estimation process progresses, the elements of matrix P tend to become smaller; consequently, the filter starts to depend too

much on the assumed model and to little on the incoming measurements.

Fading memory techniques and the overweighting of the most recent measurements (discussed in Maybeck's Volume 2) help in situations where, for example, the true system model changes over time, perhaps in a manner dependent on the system state. In such a case we may want to suppress the effects of more remotely past estimates on the current estimates.

Next we briefly examine how to handle several situations that arise when some of the distributional assumptions 1D–4D (or 1C–4C) are violated. To compensate for nonwhiteness (autocorrelation over time) in the measurement noise $\{v_k\}$ or in the process noise $\{w_k\}$ one can resort, at the cost of increased computing load, to one of several known "whitening" techniques. For example, if the measurement noise $\{v_k\}$ is nonwhite so that the covariance matrix $E[v_j v_k^{\mathrm{T}}]$ is not zero for some $j \neq k$, one can conceptually model the process $\{v_k\}$ as a finite-dimensional system driven by white noise. This can be done only by adding new states to the overall system model and, thereby, increasing the processing cost. On the other hand, it saves the optimality properties of the filter.

A different situation arises when both noise processes $\{w_k\}$ and $\{v_k\}$ are white but the measurement noise covariance matrix R_k is nonnegative definite with nullity q (has q zero eigenvalues) for all k. Then (as shown by Anderson and Moore) q functionals of the state are exactly known and the length of the state vector can be shortened by q. If the covariance matrices are nonsingular, one can construct a suboptimal filter of reduced dimension.

Another pathological situation, violating the distributional Assumption 3D or 3C, comes up in feedback control applications when the process and measurement noise processes $\{w_k\}$ and $\{v_k\}$ are correlated or even identical. This problem can be solved by modifying the KF covariance processing equations (15), (16), and (18).

Filter tuning and stabilization techniques mentioned in the preceding paragraphs are critically important in applications work, that is, in designing filters for real-life situations and then making them work properly. Actually, every such development requires at least some tuning.

C. Nonlinear System Model

We now generalize the discrete-time system model defined in terms of Eqs. (2) and (3) to

$$s_k = f_{k-1}(s_{k-1}, u_{k-1}) + G_{k-1}(s_{k-1})w_{k-1} \tag{29}$$

and

$$m_k = h_k(s_k) + v_k. \tag{30}$$

We do it by replacing the linear terms $\Phi_{k-1}s_{k-1}$ and $B_{k-1}u_{k-1}$ of the linear dynamics model by a function $f_k(\cdot, \cdot)$, allowing the process noise-to-state transformation matrix G_{k-1} to depend on system state, and generalizing the state-to-measurement transformation matrix H_k to a function $h_k(\cdot)$. The subscripts attached to f, G, and h emphasize that these functions may be of time-dependent form. Furthermore, f may be nonlinear in s and possibly in u; h may be nonlinear in s. The same distributional assumptions as made for the linear discrete-time model (Assumptions 1D–4D) are used for the nonlinear model defined by Eqs. (29) and (30). Now as before, the noise is included in the model through additive terms; also, functions f and h in general are vector-valued so that their dimensions, respectively, agree with the lengths of vectors s and m. More general nonlinear system models are possible.

One could similarly define a continuous-time nonlinear system model, but we do not do it in this chapter.

D. General Approach to Nonlinear Filtering

Nearly all presently known approaches to nonlinear filtering are based, in one way or another, on the application of Bayes recursive probability equations (12a)–(12c) to a nonlinear system model, say to one of the form defined by Eqs. (29) and (30). The underlying problem is then to estimate from measurements $M_k = \{m_0, m_1, \ldots, m_k\}$ the evolution of the posterior probability densities $p(s_k \mid M_k)$ over t_k for $k = 0, 1, \ldots$. Knowing $p(s_k \mid M_k)$ at each sampling epoch t_k enables one to apply the optimization criterion of a desired type (such as the MMSE, the maximum a posteriori, or one of some other type) to obtain the corresponding estimates of state.

When the low-order moments or some other statistics are sufficient in the sense that they define or adequately approximate $p(s_k \mid M_k)$, optimal or nearly optimal filtering schemes can be derived for certain limited classes of problems. Typically, this can be done when a combination of the following conditions exists: (a) the nonlinearities in Eqs. (29) and (30) locally can be well approximated at each t_k by a linearized model (e.g., by a Taylor series approximation to the original model); (b) an adequate approximation to the state trajectory (known as a nominal state trajectory) is a priori available; (c) the actual noise levels are low and both noise processes are Gaussian. Several versions of the linearized extensions of the KF discussed below, are the most commonly used representatives of such nonlinear estimators. In general, although the manner in which $p(s_k \mid M_k)$ evolves can be described by means of difference (in discrete-time case) or differential (in continuous-time case) equations, these equations usually are too complex to be solved for practical implementation analytically in

closed form, or even numerically. This is especially true for real-time filter implementations. Because of these difficulties, much of the research thrust has been in the direction of developing tractable approximations to $p(s_k \mid M_k)$.

E. Linearized Extensions of Kalman Filters

If the functions f and h are sufficiently smooth and can be locally expanded in Taylor series with respect to the estimates or with respect to an a priori generated nominal time history of system states, then one can apply the KF algorithm to the linearized system model

$$s_k = F_{k-1}s_{k-1} + G_{k-1}w_{k-1} \tag{31}$$

and

$$m_k = H_k s_k + v_k, \tag{32}$$

with the matrices F, G, and H defined in terms of their (i, j)th elements as

$$\begin{aligned} F_k &\equiv F_k[s_{0k}, u_k] \equiv ((F_{i,j)t}))_{t=t_k} \\ &\equiv \left(\left(\delta f_{(i)t}(s, u)\big/\delta s_{(j)t}\right)\right)_{s=s_{0k},u=u_k,t=t_k}, \end{aligned} \tag{33}$$

$$G_k \equiv G_k(s_{0k}) \equiv \left(\left(G_{(i,j)t}(s)\right)\right)_{s=s_{0k},t=t_k}, \tag{34}$$

and

$$H_k \equiv H_k[s_{0k}] \equiv \left(\left(\delta h_{(i)t}(s)\big/\delta s_{(j)}\right)\right)_{s=s_{0k},t=t_k}. \tag{35}$$

Here we introduce the notational convention $X = ((X_{(i,j)t}))$ to denote matrix X at time t in terms of its (i, j)th element $X_{(i,j)t}$. For example, the right-hand side of Eq. (33) indicates that the (i, j)th element of matrix F_k is the partial derivative at $t = t_k$ of the ith component of f_k with respect to the jth component of s; the subscript expression to the right of double parentheses specifies at what point these partial derivatives are to be evaluated. In (33), this point is

$$s = s_{0k}, \qquad u = u_k, \qquad t = t_k,$$

where: $s = s_{0k}$ indicates the value of the nominal state vector s_0 at $t = t_k$; $u = u_k$ is the value of control u at $t = t_k$.

The two most common choices for the nominal state trajectory $\{s_{0k}\}$ are as follows: (a) $s_{ok} \equiv \hat{s}_{k|k}$, or $\hat{s}_{k|k-1}$, which yields the extended Kalman filter (EKF), or (b) $\{s_{0k}\}$ is a priori set to be a nominal state trajectory, which gives the so-called linearized Kalman filter (LKF). As an example, we next state in detail the algorithm for the discrete-time EKF, which is similar in form to the discrete-time KF algorithm defined in terms of Eqs. (13)–(18).

Discrete-Time Extended Kalman Filter Algorithm. At each $t_k = t_{k-1} + \Delta t_{k-1}$ $(k = 0, 1, 2, \ldots)$ proceed as follows.

a. If $k = 0$ (i.e., at the initial time t_0), initialize the estimation process by setting

$$\hat{s}_{0|-1} = \bar{s}_0 \qquad \text{and} \qquad P_{0|-1} = \Sigma_0, \tag{36}$$

where $\bar{s}_0$ and Σ_0 are as defined by Eq. (6).

b. If $k \geq 1$, propagate the estimates of system state vector s and the state error covariance matrix P from t_{k-1} to t_k by computing

$$\hat{s}_{k|k-1} = f_{k-1}(\hat{s}_{k-1|k-1}, u_{k-1}) \tag{37}$$

and

$$P_{k|k-1} = F_{k-1}P_{k-1|k-1}F_{k-1}^{\mathrm{T}} + G_{k-1}Q_{k-1}G_{k-1}^{\mathrm{T}}, \tag{38}$$

respectively.

c. To update the estimates, first compute

$$H_k = H_k[\hat{s}_{k|k-1}] \tag{39}$$

according to Eq. (35) and evaluate the Kalman gain matrix

$$K_k = P_{k|k-1}H_k^{\mathrm{T}}\left[H_k P_{k|k-1}H_k^{\mathrm{T}} + R_k\right]^{-1}. \tag{40}$$

Next use the current measurements m_k to update the estimate of s by means of

$$\hat{s}_{k|k} = \hat{s}_{k|k-1} + K_k[m_k - h_k(\hat{s}_{k|k-1})]. \tag{41}$$

Then update the estimate of state error covariance matrix P by computing

$$P_{k|k} = [I - K_k H_k]P_{k|k-1}. \tag{42}$$

d. Finally, compute the matrix

$$F_k \equiv F_k[\hat{s}_{k|k}, u_k] \tag{43}$$

according to Eq. (33) and evaluate the process noise-to-state transformation matrix

$$G_k \equiv G_k[\hat{s}_{k|k}]. \tag{44}$$

The above-stated discrete-time EKF reduces to the standard discrete-time KF with its all-optimality properties if the *true* system model is linear, that is, if the true model is defined by Eqs. (2) and (3). When the EKF algorithm is applied to a nonlinear system whose true model is of the form given by Eqs. (29) and (30), the resulting estimator is no longer linear or optimal, which makes it more prone to various problems. For example, if the time step Δt used is too large to accommodate nonlinearities in system dynamics or in measurements, the estimates may be biased or even the entire estimation process may be destabilized by filter divergence. Thus, the tuning and the performance validation of an EKF or an LKF typically require considerably more work (simulation and analysis) than performing the same tasks for a KF applied to a truly linear model.

In addition to the LKF and the EKF several other types of linearized extensions of the KF are known. One class

of variations can be derived by including more than the first-order term in the Taylor series expansions of functions $f_k(s_k, u_k)$ or $h_k(s_k)$. For example, the estimation algorithm obtained by including the first- and second-order terms is known as the second-order extended (or linearized) KF. Another variation can be constructed by using several, say I, iterations to perform the measurement update of state with each set of measurements m_k. In such a scheme, step c of the discrete-time EKF algorithm is repeated I times. One enters the ith pass of this step with $\hat{s}_{k|k}(i-1)$, then computes $\hat{s}_{k|k}(i)$ and all other quantities dependent on $\hat{s}_{k|k}(i)$. The first iteration is entered by defining $\hat{s}_{k|k}(0) \equiv \hat{s}_{k|k-1}$. At the end of the Ith iteration, one defines $\hat{s}_{k|k} \equiv \hat{s}_{k|k}(I)$.

Although the measures described in the foregoing two paragraphs often stabilize the EKF and produce satisfactory results, they fail for some nonlinear or non-Gaussian systems. The cause of such failures is usually ascribed to two kinds of approximations used in the construction of EKF: (a) the linearization of system equations relative to some reference values and (b) the assumption of Gaussianity for all a priori distributions in the system model. Approximations of the second kind enforce the retention of a false illusion that the probability densities $p(s_k \mid M_k)$ continue to be Gaussian for all k. For example, if the true conditional distribution of state is multimodal [i.e., densities $p(s_k \mid M_k)$ have multiple peaks], the EKF functions more as a maximum likelihood than as an MMSE estimator and the estimated mean may end up following one (if any) of the peaks.

EXAMPLE 2. ESTIMATING THE MOTION AND THE DRAG CHARACTERISTICS OF A BALLISTIC PROJECTILE ENTERING THE ATMOSPHERE. Consider the problem of estimating the motion and the drag (air resistance) characteristics of a *ballistic projectile* (BP). To simplify the problem without robbing from it its illustrative value, assume that the BP is aimed at the North Pole and is approaching the earth along the polar axis. We denote this axis the Y axis of the one-dimensional coordinate system whose origin is assumed to coincide with the North Pole and whose positive direction points "upward" from the pole.

Notation. In this example, the individual components of a vector are referred to by means of parenthesized subscripts. For example, if z is a column vector of length 2, we write $z = [z_{(1)}, z_{(2)}]^{\mathrm{T}}$. In particular, to refer to the values of these components at time t, we write either $[z_{(1)t}, z_{(2)t}]$ or $[z_{(1)}, z_{(2)}]_t$. If t represents a sampling epoch, say t_k, we write either $[z_{(1)k}, z_{(2)k}]$ or $[z_{(1)}, z_{(2)}]_k$. Another notational convention adopted here is the use of dots above a symbol to denote the derivatives with respect to time of the quantity represented by the symbol. For example, $\dot{y}_k$ and $\ddot{y}_k$, respectively, denote the values of dy/dt and d^2y/dt^2 at time t_k. In addition, we use "hats" over lowercase letters to denote unit vectors. For example, $\hat{v}$ denotes the unit vector in the direction of velocity vector v.

Physical Model. We model the earth as a perfect sphere of radius

$$R_E = 6{,}378{,}135\,(\mathrm{m}).$$

For the present problem, the gravitational field at any point P on the polar (Y) axis, y meters above sea level, is approximated by

$$-g(y)\hat{y} = -(g_0)(C)(1.0 + y/R_E)^{-2}\hat{y}\ (\mathrm{m/sec}^2),$$

where

$$g_0 = 9.7803327\,(\mathrm{m/sec}^2)$$

and

$$C = 1.00530246.$$

The air density at P (y meters above sea level) is approximated by

$$\rho \equiv \rho(y) = \rho_0 \exp(-\gamma_y)\,(\mathrm{kg/m}^3),$$

where

$$\rho_0 = 1.22\,(\mathrm{kg/m}^3),$$
$$\gamma = 1.6404 \times 10^{-4}\ (\mathrm{m}^{-1}).$$

The drag force on the BP, divided by its unknown mass and moving with velocity $\dot{y}$, is approximated by

$$d \equiv d(-\hat{v}) = (\dot{y})^2[\rho(y)](D_F)\ (\mathrm{m/sec}^2),$$

where $\hat{v}$ is the unit vector in the direction of velocity, and D_F is the unknown drag factor (m/kg). In the present case, $\hat{v} = -\hat{y}$. Hence, the drag can be written

$$d(y, \dot{y})\hat{y} = (\dot{y})^2[\rho(y)](D_F)\hat{y}.$$

It follows that the differential equation of motion for the BP is of the form

$$\ddot{y} = -g(y) + d(y, \dot{y}) + \text{process noise}.$$

System Model. Analysis of the physical model suggests the use of a four-dimensional ($n_s = 4$) state vector (SV) defined by letting

$$s_{(1)} \leftrightarrow y, \qquad s_{(2)} \leftrightarrow \dot{y}, \qquad s_{(3)} \leftrightarrow \ddot{y}, \qquad s_{(4)} \leftrightarrow D_F.$$

The state transition equations for this SV can be directly written from the time-propagation equations,

$$\begin{aligned} y_{k+1} &= y_k + h\dot{y}_k + (0.5)h^2\ddot{y}_k + w_{(1)k}, \\ \dot{y}_{k+1} &= \dot{y}_k + h\ddot{y}_k + w_{(2)k}, \\ \ddot{y}_{k+1} &= \ddot{y}_k + h\dddot{y}_k + w_{(3)k}, \\ (D_F)_{k+1} &= (D_F)_k + w_{(4)k}, \end{aligned}$$

where

$$h \equiv \Delta t_k = t_{k+1} - t_k,$$
$$\ddot{y}_k = \{(d/dy)[-g(y)]\dot{y} + (d/dy)[d(y, \dot{y})]\dot{y} + (d/d\dot{y})[d(y, \dot{y})]\ddot{y}\}_k,$$

is a function of y, $\dot{y}$, $\ddot{y}$, and D_F (all evaluated at $t = t_k$); $w_{(j)k}$ are components of process noise vector w_k. Hence writing

$$s_{(j)k+1} = f_{(j)}(s_k) + w_{(j)k} \qquad \text{for} \qquad j = 1, \ldots, n_s,$$

we get

$$f_{(1)}(s_k) = \left(s_{(1)} + hs_{(2)} + (0.5)h^2 s_{(3)}\right)_k,$$
$$f_{(2)}(s_k) = \left(s_{(2)} + hs_{(3)}\right)_k,$$
$$f_{(3)}(s_k) = \left(s_{(3)} + h\dot{s}_{(3)}\right)_k,$$
$$f_{(4)}(s_k) = \left(s_{(4)}\right)_k,$$

where, as it follows from a previous expression for $\ddot{y}_k$, the time derivative $\dot{s}_{(3)k}$ can be expressed in terms of $s_{(1)}$, $s_{(2)}$, $s_{(3)}$, and $s_{(4)}$ (all evaluated at t_k).

We assume that at each observation epoch t_k a ground-based sensor measures both the position (y) and the velocity ($\dot{y}$) of the incoming ballistic missile. Hence the measurement equations are of the form

$$m_{(1)k} = y_k + v_{(1)k},$$
$$m_{(2)k} = \dot{y}_k + v_{(2)k},$$

or (in vector/matrix form)

$$m_k = Hs_k + v_k.$$

Here v_k represents the measurement noise and

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Filter implementation requires a procedure for evaluating the F matrix. As noted in the description of the EKF algorithm, the (i, j)th element of F_k is obtained by differentiating $f_{(i)}$ with respect to $s_{(j)}$ and then evaluating this derivative at t_k. For example, in the present case, we get

$$F_{(1,2)k} = \left(\delta f_{(1)} / \delta s_{(2)}\right)_k = h = \Delta t.$$

In our implementation of the filter for the present problem, we included an algorithm for bounding the state error covariance matrix P. (This algorithm is described in Maybeck's Volume 2.) We found it to be effective in stabilization of the filtering process.

Results of a Simulation Run. Figures 2A–E illustrate the results of a computer simulation run performed for the parameter values and the initial conditions summarized in Fig. 2A. The filter was tuned up to respond rapidly to relatively accurate measurements. Note that the drag factor D_F (expressed as square meters per kilogram) is treated as an unknown constant parameter that is estimated along with the dynamic variables of system state.

F. Other Approaches to Nonlinear Filtering and the Gaussian Sum Approximation Filter

Since it is generally impossible to construct for a nonlinear or non-Gaussian system an exact representation of a posteriori probability density functions (PDFs) $p(s_k \mid M_k)$, various types of finite approximations to $p(s_k \mid M_k)$ have been investigated. As an illustration of this approach we outline next one specific technique, known as the Gaussian sum approximation (GSA).

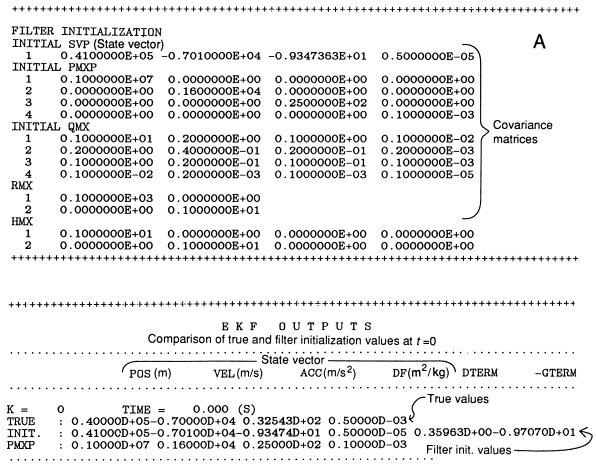
Let R^n denote the n-dimensional Euclidean space. The GSA method is based on a theoretical result, according to which any probability function p of a vector (or scalar-)-valued RV x, say of length n, can be approximated as closely as desired in the $L_1(R^n)$ space (i.e., the integral $\int_{R^n} | p(x) - p_A(x) | dx$ can be made arbitrarily small) by a function of the form

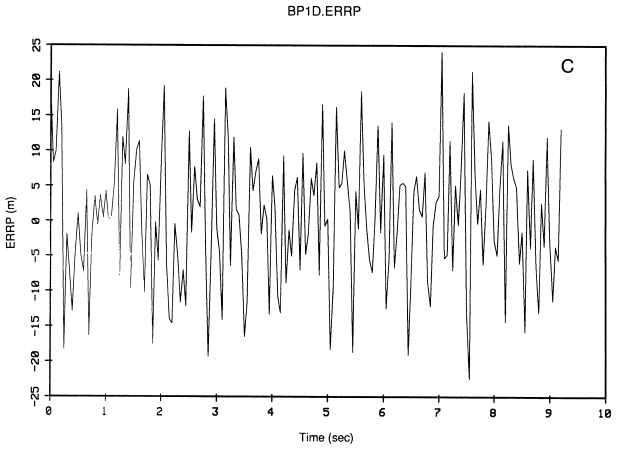
$$p_A(x) = \sum_{i=1}^{L} \alpha_i N[x - a_i, B_i] \tag{45}$$

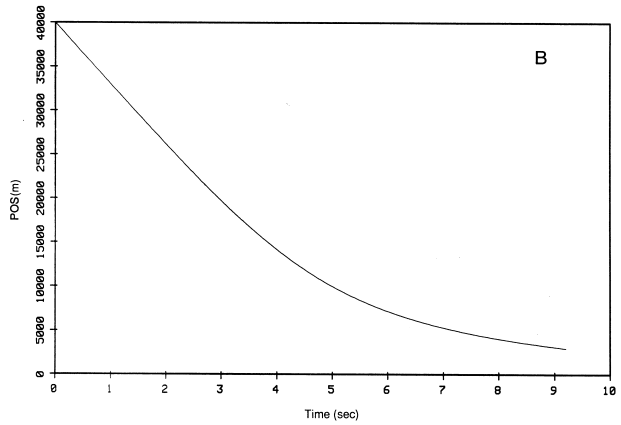
for some positive integer L, positive scalars α_i such that $\alpha_1 + \alpha_2 + \cdots + \alpha_L = 1$, suitably selected n-dimensional vectors a_i, and $n \times n$ positive definite matrices B_i. In Eq. (45), $N[x - a_i, B_i]$ represents the n-variate Gaussian PDF

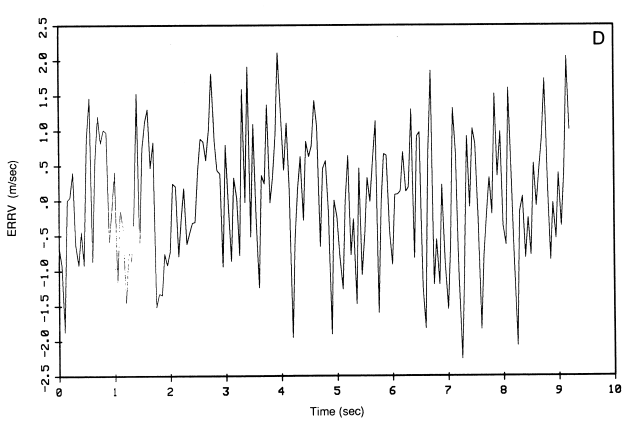
$$(2\pi)^{-n/2} |B_i|^{-1/2} \exp\left\{-\tfrac{1}{2}(x - a_i)^{\mathrm{T}} B_i^{-1} (x - a_i)\right\}, \tag{46}$$

with $|B_i|$ denoting the determinant of B_i. Furthermore, it can be shown that $p_A(x)$ converges uniformly in x to any PDF of practical interest as the covariance matrices B_i tend to a zero matrix and L increases. There are many ways of selecting the parameters a_i, B_i, and α_i. For construction of a filtering scheme, it is convenient (a) to choose the values of a_i so as to build up in the state space a grid covering at least a substantial part of the mass of the PDF to be approximated and (b) to set $B_i = bI$ (with I representing an identity matrix) for some small positive scalar b. Then each term in the Gaussian sum on the right-hand side of Eq. (45) converges to an impulse function centered at a_i as $b(>0)$ converges down to zero. Hence, for a small $b > 0$, each term in the Gaussian sum is practically zero outside a small open neighborhood (hypersphere) in R^n containing a_i. Since $p_A(x) \geq 0$ and $\int p_A(x)\, dx = 1$, p_A is a PDF.









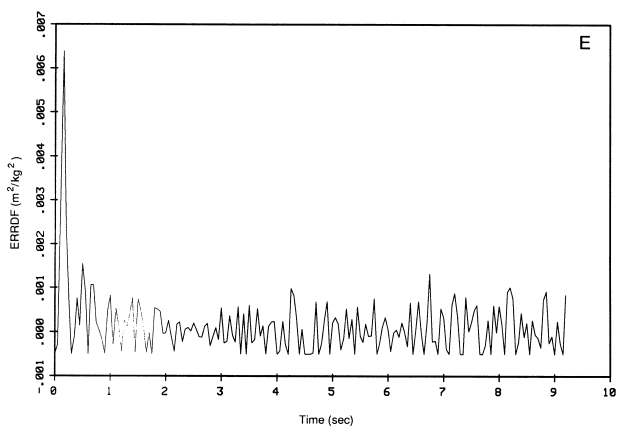


FIGURE 2 (A) The upper part shows the state vector and filter matrices at the initialization time t_0. The lower part compares the true and initialized state vectors at t_0. (B) True position (the Y coordinate) of the ballistic projectile vs time. (C) Error in position estimates vs time. Because of the compression of time scale, the initialization error of 1 km is invisible in the graph. (D) Error in velocity estimates vs time. Because of the compression of time scale, the initialization error of 10 m/sec is not visible on the graph. (E) Error in D_F estimates vs time.

Next suppose that $p(s_k \mid M_k)$ is approximated by

$$p_A(s_k \mid M_k) = \sum_{i=1}^{L} \alpha_{ik} N[s_k - \bar{s}_{ik}, B_{ik}]. \quad (47)$$

Then $p_A(s_k \mid M_k)$ is also a PDF, and while using appropriately selected values of $\bar{s}_{ik}$, one can show that

$$\hat{s}_{k|k} \equiv E[s_k \mid M_k] = \sum_{i=1}^{L} \alpha_{ik} \bar{s}_{ik} \quad (48)$$

and

$$P_{k|k} \equiv E\big[(s_k - \hat{s}_{k|k})(s_k - \hat{s}_{k|k})^{\mathrm{T}} \big| M_k\big] = \sum_{i=1}^{L} \alpha_{ik}\big[B_{ik} + (\hat{s}_{k|k} - \bar{s}_{ik})(\hat{s}_{k|k} - \bar{s}_{ik})^{\mathrm{T}}\big]. \quad (49)$$

This suggests that perhaps the conditional MMSE estimator $\hat{s}_{k|k}$ can be approximated as a weighted sum of conditional mean estimators, each representing an EKF. It can be proved that this is indeed the case. Combining the foregoing ideas with the nonlinear system model given by Eqs. (29) and (30) and Assumptions 1D–4D, the latter now possibly relaxed by dropping some or all Gaussianity assumptions, leads to a filtering scheme, which we outline next.

Discrete-Time Gaussian Sum Approximation (GSA) Filter Algorithm. This algorithm applies to the system model defined by Eqs. (29) and (30) and by the probabilistic assumptions 1D through 4D. Let n_s be the dimension of state vectors.

a. Let $k = 0$ (i.e., initialize the sampling time t_k to t_0) and select appropriate values of the parameters $\bar{s}_{ik}$ and $\alpha_{i,k-1}$ (all $\alpha_{i,k-1} \geq 0$ and $\sum_{i=1}^{L} \alpha i, k-1 = 1$) for $i = 1, \ldots, L$.

b. To complete a GSA,

$$\sum_{i=1}^{L} \alpha_{i,k-1} N[s - \bar{s}_{ik}, \bar{B}_{ik}]$$

to $p(s_k \mid M_{k-1})$, where $M_k = \{m_k, M_{k-1}\}$ and M_{k-1} is an empty measurement set when $k = 0$, initialize (reinitialize) all L covariance matrices $\bar{B}_{ik}$ to bI, with b representing a small positive constant and I the $n_s \times n_s$ identity matrix.

c. Approximate $p(s_k \mid M_k)$ by means of the following measurement-update procedure. For $i = 1, \ldots, L$, compute

$$\begin{aligned} H_{ik} &= ((\delta h_k(s)/\delta s))_{s=\bar{s}_{ik}}, \\ D_{ik} &= \big(H_{ik}\bar{B}_{ik}H_{ik}^{\mathrm{T}} + R_k\big)^{-1}, \\ K_{ik} &= \bar{B}_{ik}H_{ik}^{\mathrm{T}}D_{ik}, \\ B_{ik} &= \bar{B}_{ik} - \bar{B}_{ik}H_{ik}^{\mathrm{T}}D_{ik}H_{ik}\bar{B}_{ik}^{\mathrm{T}}, \\ s_{ik} &= \bar{s}_{ik} + K_{ik}[m_k - h_k(\bar{s}_{ik})], \end{aligned}$$

with the new weights determined by

$$S_k = \sum_{j=1}^{L} \alpha_{j,k-1} N\big[m_k - h_k(\bar{s}_{jk}), D_{ik}^{-1}\big]$$

and

$$\alpha_{ik} = \alpha_{i,k-1} N\big[m_k - h_k(\bar{s}_{ik}), D_{ik}^{-1}\big] \big/ S_k.$$

Finally, compute $\hat{s}_{k|k}$ and $P_{k|k}$ by means of Eqs. (48) and (49), respectively.

d. Approximate $p(s_{k+1} \mid M_k)$ by means of the following time-update procedure:

$$\sum_{i=1}^{L} \alpha_{ik} N[s_{k+1} - \bar{s}_{i,k+1}, \bar{B}_{i,k+1}],$$

where

$$\begin{aligned} F_{ik} &= ((\delta f_k(s)/\delta s))_{s=s_{ik}}, \\ \bar{s}_{i,k+1} &= f_k(s_{ik}), \\ \bar{B}_{i,k+1} &= F_{ik}B_{ik}F_{ik}^{\mathrm{T}} + G_k(s_{ik})Q_kG_k(s_{ik})^{\mathrm{T}}. \end{aligned}$$

e. Let $k \leftarrow k+1$ (i.e., increment the sampling time by setting $t_k \leftarrow t_{k+1}$). If the new $t_k > t_{\max}$ (= estimation end time), then *quit*; otherwise, proceed to step f.

f. Test whether $\bar{B}_{ik} < bI$ (i.e., whether $[bI - \bar{B}_{ik}]$ is a positive definite matrix) for all i. If yes, go to step c; otherwise, go to step b.

In the preceding algorithm, use the same rule for computing matrices H_{ik} and F_{ik} as given for matrices H_k and F_k [see the text following Eqs. (33)–(35)]. As the above algorithm indicates, both the measurement-update and the time-propagation equations for each constituent filter in the Gaussian sum are essentially those of an EKF; besides, when any covariance matrix $\bar{B}_{ik}$ becomes too "big" (in the sense explained in step f), all L covariance matrices $\bar{B}_{ik}$ must be reinitialized. If $L = 1$, the GSA filter reduces to an EKF.

The following two important advantages of the GSA come from the choice of the Gaussian weighting PDFs and the positive weighting coefficients that add to 1: (a) any finite Gaussian sum is a PDF; (b) the presence of Gaussian weighting PDFs allows the utilization of certain Gaussian properties in the computation of approximations to $p(s_k \mid M_k)$, though the resulting approximation function p_A is not Gaussian. However, the GSA approach has one disadvantage: The series obtained from the Gaussian sum expansion cannot be orthogonalized. Consequently, it is difficult to obtain useful error bounds. Thus, as an alternative to the GSA, the use of approximations based on orthogonal series expansions (e.g., the Edgeworth and Gram–Charlier series) has been investigated. Despite the attractiveness of orthogonality, such approaches have one

distinct disadvantage. The resulting truncated sum approximations are not PDFs.

Finally, we should mention the nonlinear filtering methods based on the assumption that $p(s_k|M_k)$ is Gaussian for all k. The EKF was one example of this approach. As already noted, this assumption often leads to discarding a great deal of information present in the actually true PDF, especially in the multimodal case.

IV. FILTER IMPLEMENTATION

Filters of the type considered in this chapter are usually implemented in software form. Besides a filter is typically designed for ultimate implementation as part of a real-time system, which means that it will be subject to timing constraints of that system. Thus, in addition to tasks already discussed (e.g., defining the system model, selecting a filter algorithm, and making a trial filter work), one must (a) select appropriate numerical procedures to be used with the filter algorithm and (b) adapt the program implementing the filter to the real-time processing environment in case the filter is to become part of a real-time system. In the latter case, one must also time and size the algorithms for a particular real-time computer.

A. Numerical Procedures

The accuracy, stability, and computing efficiency of a filter depend critically on the adopted numerical procedures. For example, the unnecessary use of multiple precision arithmetic may be prohibitively expensive in some real-time applications; on the other hand, careless computation of dot products while using operands of insufficient length may quickly lead to an unbounded accumulation of roundoff error and destabilization of the estimation process. As another example, consider a continuous-time nonlinear filter. It often uses numerical integration to propagate the estimates of the state to the next time point, and thus one faces the problem of selecting an appropriate method for propagating numerically the solution of underlying differential equations.

There are several simple rules of thumb with regard to matrix operations in Kalman filtering of linear systems: (a) Dot products should in general be computed using arithmetic of a higher precision than used in the remaining matrix operations; (b) matrices that theoretically are symmetric should be periodically resymmetrized [this simply requires replacing each member of every pair of elements a_{ij} and a_{ji}, where $i \neq j$, of a theoretically symmetric matrix A with $(a_{ij} + a_{ji})/2$]; (c) positive definiteness of theoretically positive definite matrices should be preserved. Furthermore, in many applications, such as in precision navigation, the state must be known to many more significant digits than the gain and the state error covariances. Hence, nearly all gain–covariance processing can be performed at a lower precision than the processing of the state vector, but this approach is safe only if the gain–covariance processing is computationally stable. Square root filtering, discussed next, provides a numerically stable method for preserving the positive definiteness in state error covariance matrices.

B. Square Root Filtering

By far the greatest trouble spot in computer mechanization of the KF is the updating of state error covariance matrix P, that is, the computation of $P_{k|k}$ according to Eq. (18). As the estimation process progresses, the elements of $P_{k|k}$ typically continue to decrease in magnitude and so matrix $P_{k|k}$ keeps approaching the zero matrix, although theoretically it should forever remain positive definite, no matter how small in magnitude its elements become. Hence, unless special measures are taken, accumulation of roundoff error in the repetitive use of Eq. (18) may cause the computed $P_{k|k}$ to lose its positive definiteness. As suggested by the matrix inversion operation appearing on the right-hand side of Eq. (16) for computing the Kalman gain, this situation is aggravated if several components of the measurement vector are very accurate and consequently the positive definite measurement error covariance matrix R is ill conditioned, that is, if R has eigenvalues of both relatively very large and small magnitudes.

Let A be a nonnegative definite symmetric matrix; then there exists a matrix S such that $A = SS^{\mathrm{T}}$. Matrix S is often called the square root of A. The Cholesky decomposition algorithm provides a method of constructing from A the matrix S so that S is lower triangular; that is, all elements of S above the main diagonal are zero. Square root filtering is motivated by the observations that, if the state error covariance matrix $P = SS^{\mathrm{T}}$, then (a) since SS^{T} is always nonnegative definite, matrix P expressed as SS^{T} cannot become negative definite, and (b) matrix S is generally less ill conditioned than matrix P.

Several versions of the square root filtering algorithm are known. The earliest form was developed by J. E. Potter in 1964 for applications in which the process noise is absent (i.e., covariance matrix Q is zero) and the measurements are sequentially processed as scalars. In 1967 J. F. Bellantoni and K. W. Dodge extended Potter's results to vector-valued measurements. A. Andrews in 1968 and then S. F. Schmidt in 1970 published two alternative procedures for handling the process noise. In 1973 N. A. Carlson described a procedure that considerably improved the speed and decreased the memory requirements of square root filtering and in which, as in Potter's algorithm,

vector-valued measurements are processed sequentially as scalars. Finally, the so-called UDU^{T} covariance factorization method is the most recent major milestone in numerical handling of KFs. This method, developed by G. J. Bierman and C. L. Thornton, represents the state error covariances before and after the measurement update step as

$$P_{k|k-1} = (U_{k|k-1})D_{k|k-1}(U_{k|k-1})^{\mathrm{T}}$$

and

$$P_{k|k} = (U_{k|k})D_{k|k}(U_{k|k})^{\mathrm{T}},$$

with D being a diagonal matrix and U an upper triangular matrix with 1's on its main diagonal. In this method, the square root of the covariance matrix, which now would correspond to $UD^{1/2}$, is never computed explicitly, which avoids numerical computation of square roots. Like Carlson's algorithm, the UDU^{T} factorization method maintains the covariance matrix in factored form and so (like Carlson's algorithm) is considerably more efficient in processor time and storage than the original Potter algorithm.

As a quick comparison of computational efficiency, the conventional Kalman method, the less efficient form of Carlson's algorithm, and the UDU^{T} factorization method are roughly equal: The processing of each time step (consisting of one time propagation and one measurement update) requires of the order of $\frac{1}{6}[9n_s^3 + 9n_s^2n_m + 3n_s^2n_w]$ adds and about the same number of multiplies, plus a relatively modest number of divides and square roots (square roots are required only in some, as in Potter's or Carlson's square root algorithms). Here, as before, n_s is the length of the state vector, n_m the length of the measurement vector, and n_w the lenght of the process noise vector w. The faster version of Carlson's algorithm is more efficient and requires only of the order of $\frac{1}{6}[5n_s^3 + 9n_s^2n_m + 3n_s^2n_w]$ adds and $\frac{1}{6}[5n_s^3 + 12n_s^2n_m + 3n_s^2n_w]$ multiplies, plus $2n_sn_m$ divides and n_sn_m square roots, at each time point. The stable (Joseph) form of the KF [as given by Eq. (18′)] fares more poorly: At each time step, it requires of the order of $\frac{1}{6}[18n_s^3 + 15n_s^2n_m + 3n_s^2n_w]$ adds and about the same number of multiplies.

As a summary, (a) a square root filter is a numerically stable form for performing the KF covariance–gain processing defined by Eqs. (15), (16), and (18); (b) the efficiency of its more recent versions roughly compares with that of these three equations; (c) the increased stability allows one to use relatively low-precision arithmetic in the KF gain–covariance processing, with a possible exception of some dot products.

Real-time implementation of a filter involves additional issues that are unimportant in the non-real-time environment. Besides the adequacy of functional performance, the most important of these issues is the requirement to produce timely responses to external stimuli. Thus, resorting to a parallel or concurrent processing may be the only way out. This usually implies the use of special hardware architectures such as parallel, vector pipelined, or systolic processors.

As one example, consider the use of a filter in the tracking of multiple objects in a hard real-time environment characterized by strict deadlines. In such a case one may want to maintain simultaneously many estimation processes, each handling a single object. Parallel processors may seem to be a suitable hardware architecture for this problem, but if separate estimation processes in such an application progress at different rates and at any time some of them require a great amount of special handling, then parallel architecture, such as a single-instruction multiple-data stream computer, may not be the best choice. As another example, consider a KF to be implemented as part of a navigation system on a small airborne computer (uniprocessor). Suppose that the navigation measurements come at a certain fixed rate. If the filtering process cannot keep up with the arrival rate of measurements and so not all of them can be utilized, the estimation performance may deteriorate. In this problem, if there is an upper bound on hardware resources, the only solution may be to decompose the estimation algorithm into concurrently executable processes. For instance, the time-propagation step (which, say, is to be executed at a relatively high rate) may constitute one process and the measurement-update step (which needs to be executed only at some lower rate, say, at the rate of measurement arrivals) may constitute another. Such a decomposition of an estimation algorithm into concurrent procedures often creates a surrogate algorithm that performs more poorly than the original algorithm.

The effects of the finite-length word computing is another issue that must be considered in filter implementation for real-time applications. The computer on which a filter is developed and validated through successive off-line simulations is often more powerful and uses higher-precision arithmetic and number representations than the ultimate real-time processor. Hence, one must in advance determine in advance what effect a shorter word length will have on performance.

V. APPLICATIONS

A good part of present-day technology would be unthinkable without recursive estimators, especially modern computer-based control and communication systems. Kalman and other types of recursive filters have been widely adopted by the defense and aerospace communities for applications such as navigation, aircraft flight control, satellite tracking, and orbit determination.

Early applications of recursive filtering were mainly in the estimation and identification of uncertain dynamic or control systems. Many of these applications, although in general successful, were hastily executed in the sense that insufficient thought was given to various theoretical, numerical, and purely implementational issues. The first two decades were also marred by a flood of papers describing specific applications, published in the technical literature or presented at various engineering conferences. However, during that period many important theoretical issues, especially gaps in theoretical knowledge, were identified and resolved.

In contrast to the quick, widespread adoption of recursive estimators to control and of dynamic systems during the 1960s and 1970s, their penetration of signal processing was considerably less decisive. One reason for that was the inadequacy of computer hardware and software with respect to the stringent timing constraints of on-line signal processing, although there are several application areas of signal processing, such as the analysis of economic time series, in which real-time constraints are virtually nonexistent. Another factor was the technological culture of that period, which was also reinforced by the nature of the problem. In communication problems, one is usually interested in steady-state behavior. Alternative signal processing techniques, notably those based on frequency domain analysis, such as the fast Fourier transform, were developed and then adapted to digital signal processing. However, developments in computer hardware and software technologies (e.g., the introduction of gallium arsenide devices, very large-scale integration, and data-flow architectures such as systolic processors) promise to change this situation.

Several interesting types of functional extensions of filtering are known. We briefly mention only one of them, which is known as adaptive filtering. Adaptive filtering is concerned with the generalization of estimation algorithms, especially of KFs, to state estimation situations in which the system model is not completely known. For example, one may want to estimate the state of a stochastic dynamic system for which some parameters, such as the covariance matrices of noise processes or a few coefficients in the system dynamics equation, are unknown or are changing in an unknown way in time. A filter that estimates the state while concurrently estimating the values of the unknown components of the system model is called an adaptive filter. Example 2 illustrates a relatively simple case of adaptive filtering: estimation of the parameters that define the drag force while estimating the trajectory. In that example, the adaptive filtering problem is solved by introducing an additional component into the state vector.

SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT INSTRUMENTS • ANALOG ELECTRICAL FILTERS • APPROXIMATIONS AND EXPANSIONS • CIRCUIT THEORY • DIGITAL FILTERS • SPATIAL OBJECTIVE ANALYSIS • STOCHASTIC PROCESSES

BIBLIOGRAPHY

Anderson, B. D. O., and Moore, J. B. (1979). "Optimal Filtering," Prentice-Hall, Englewood Cliffs, NJ.

Davis, M. H. A., and Vinter, R. B. (1985). "Stochastic Modelling and Control," Chapman and Hall, New York.

Jazwinski, A. H. (1970). "Stochastic Processes and Filtering Theory," Academic Press, New York.

Kailath, T. (1974). *IEEE Trans. Info. Theory* **IT-20**(2), 146–179.

Kalman, R. E. (1960). *Trans. ASME Ser. D J. Basic Eng*. **82,** 34–35.

Kalman, R. E., and Bucy, R. S. (1961). *Trans. ASME Ser. D J. Basic Eng*. **83,** 95–107.

Maybeck, P. S. (1979, 1982). "Stochastic Models, Estimation, and Control," Vol. 1, Vols. 2 and 3, Academic Press, New York.

Schmidt, G. T. (1976). "Linear and Nonlinear Filtering Techniques," Control, and Dynamic Systems, Vol. 12, pp. 63–98, Academic Press, New York.

Thornton, C. L., and Bierman, G. J. (1980). "UDU^T Covariance Factorization for Kalman Filtering," Control and Dynamic Systems, Vol. 16, pp. 177–248, Academic Press, New York.