SIAM Review
Vol. 54, No. 4, pp. 801–823

A Fresh Look at the Kalman Filter*

Jeffrey Humpherys†
Preston Redd†
Jeremy West‡

Abstract. In this paper, we discuss the Kalman filter for state estimation in noisy linear discrete-time dynamical systems. We give an overview of its history, its mathematical and statistical formulations, and its use in applications. We describe a novel derivation of the Kalman filter using Newton's method for root finding. This approach is quite general as it can also be used to derive a number of variations of the Kalman filter, including recursive estimators for both prediction and smoothing, estimators with fading memory, and the extended Kalman filter for nonlinear systems.

Key words. Kalman filter, state estimation, control theory, systems theory, Newton's method

AMS subject classifications. 93E10, 60G35, 49M15

DOI. 10.1137/100799666

1. Introduction and History.

The Kalman filter is among the most notable innovations of the 20th century. This algorithm recursively estimates the state variables—for example, the position and velocity of a projectile—in a noisy linear dynamical system by minimizing the mean-squared estimation error of the current state as noisy measurements are received and as the system evolves in time. Each update provides the latest unbiased estimate of the system variables together with a measure on the uncertainty of those estimates in the form of a covariance matrix. Since the updating process is fairly general and relatively easy to compute, the Kalman filter can often be implemented in real time.

The Kalman filter is used widely in virtually every technical or quantitative field. In engineering, for example, the Kalman filter is pervasive in the areas of navigation and global positioning [39, 25], tracking [29], guidance [49], robotics [35], radar [33], fault detection [19], and computer vision [34]. It is also utilized in applications involving signal processing [36], voice recognition [11], video stabilization [10], and automotive control systems [27]. In purely quantitative fields, the Kalman filter also plays an important role in time-series analysis [15], econometrics [5], mathematical finance [31], system identification [30], and neural networks [16].

*Received by the editors on June 22, 2010; accepted for publication (in revised form) November 16, 2011; published electronically on November 8, 2012. This work was supported in part by National Science Foundation grant DMS-CSUMS-0639328.
http://www.siam.org/journals/sirev/54-4/79966.html
†Department of Mathematics, Brigham Young University, Provo, UT 84602 (jeffh@math.byu.edu, ptredd@byu.net). The first author was supported in part by National Science Foundation grant DMS-CAREER-0847074.
‡Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 (jeremymw@umich.edu).

It has been just over 50 years since Rudolf Kalman's first seminal paper on state estimation [22], which launched a major shift toward state-space dynamic modeling. This tour de force in mathematical systems theory, together with his two other groundbreaking papers [24, 23], helped secure him a number of major awards, including the IEEE Medal of Honor in 1974, the Kyoto Prize in 1985, the Steele Prize in 1986, the Charles Stark Draper Prize in 2008, and the U.S. National Medal of Science in 2009.

As is sometimes the case with revolutionary advances, some were initially slow to accept Kalman's work. According to Grewal and Andrews [14, Chapter 1], Kalman's second paper [24] was actually rejected by an electrical engineering journal, since, as one of the referees put it, "it cannot possibly be true." However, with the help of Stanley F. Schmidt at the NASA Ames Research Center, the Kalman filter ultimately gained acceptance as it was used successfully in the navigation systems for the Apollo missions, as well as several subsequent NASA projects and a number of military defense systems; see [32, 14] for details.

Today the *Kalman family* of state estimation methods, which includes the Kalman filter and its many variations, are the de facto standard for state estimation. At the time of this writing, there have been over 6000 patents awarded in the U.S. on applications or processes involving the Kalman filter. In academia, its influence is no less noteworthy. According to Google Scholar, the phrase "Kalman filter" is found in over 100,000 academic papers. In addition, Kalman's original paper [22] is reported to have over 7500 academic citations. Indeed, the last 50 years have seen phenomenal growth in the variety of applications of the Kalman filter.

In this paper, we show that each step of the Kalman filter can be derived from a single iteration of Newton's method on a certain quadratic form with a judiciously chosen initial guess. This approach is different from those found in the standard texts [14, 38, 21, 7, 12], and it provides a more general framework for recursive state estimation. It has been known for some time that the extended Kalman filter (EKF) for nonlinear systems can be viewed as a Gauss–Newton method, which is a cousin of Newton's method [3, 1, 4]. This paper shows that the original Kalman filter and many of its variants also fit into this framework. Using Newton's method, we derive recursive estimators for prediction and smoothing, estimators with fading memory, and the EKF.

The plan of this paper is as follows: In section 2, we review the statistical and mathematical background needed to set up our derivation of the Kalman filter. Specifically, we recall the relevant features of linear estimation and Newton's method. In section 3, we describe the state estimation problem and formulate it as a quadratic optimization problem. We then show that this optimization problem can be solved with a single step of Newton's method with any arbitrary initial guess. Next, we demonstrate that, with a specific initial guess, the Newton update reduces to the standard textbook form of the Kalman filter. In section 4, we consider variations on our derivation, specifically predictive and smoothed state estimates, and estimators with fading memory. In section 5, we expand our approach to certain nonlinear systems by deriving the EKF. Finally, in section 7, we suggest some classroom activities.

To close the introduction, we note that while this paper celebrates Kalman's original groundbreaking work [22], other people have also made major contributions in this area. In particular, we mention the work of Thiele [44, 45, 28], Woodbury [47, 48], Swerling [42, 43], Stratonovich [40, 41], Bucy [24], Schmidt [37, 32], and, more recently, Julier and Uhlmann [20] and Wan and van der Merwe [46] for their development of the unscented Kalman filter. We also acknowledge that there are important numerical considerations when implementing the Kalman filter which are not addressed in this paper; see [21] for details.

2. Background.

In this section, we recall some facts about linear least squares estimation and Newton's method; for more information, see [13, 26]. Suppose that data are generated by the linear model

$$b = Ax + \varepsilon, \tag{2.1}$$

where A is a known $m \times n$ matrix of rank n, ε is an m-dimensional random variable with zero mean and known positive-definite covariance $Q = \mathbb{E}[\varepsilon\varepsilon^T] > 0$, and $b \in \mathbb{R}^m$ represents known, but inexact, measurements with errors given by ε. The vector $x \in \mathbb{R}^n$ contains the parameters to be estimated.

Recall that a *linear estimator* $\hat{x} = Kb$ is said to be *unbiased* if, for all x, we have

$$x = \mathbb{E}[\hat{x}] = \mathbb{E}[Kb] = \mathbb{E}[K(Ax + \varepsilon)] = KAx,$$

where K is some $n \times m$ matrix. The following theorem states that, among all linear unbiased estimators, there is a unique choice that minimizes the mean-squared error and that this estimator also minimizes the covariance.

THEOREM 2.1 (Gauss–Markov [13]). *The linear unbiased estimator for* (2.1) *that minimizes the mean-squared error is given by*

$$\hat{x} = (A^T Q^{-1} A)^{-1} A^T Q^{-1} b \tag{2.2}$$

and has variance

$$\mathbb{E}[(\hat{x} - x)(\hat{x} - x)^T] = (A^T Q^{-1} A)^{-1}. \tag{2.3}$$

Moreover, any other unbiased linear estimator $\hat{x}_L$ of x has larger variance than (2.3)*. Thus we call* (2.2) *the* best linear unbiased estimator*. It is also called the* minimum variance linear unbiased estimator *and the* Gauss–Markov estimator*.*

For a proof of this theorem, see Appendix A. We remark that (2.2) is also the solution of the weighted least squares problem

$$\hat{x} = \underset{x}{\operatorname{argmin}} \frac{1}{2}\|Ax - b\|^2_{Q^{-1}}, \tag{2.4}$$

where the objective function

$$J(x) = \frac{1}{2}\|Ax - b\|^2_{Q^{-1}} = \frac{1}{2}(Ax - b)^T Q^{-1}(Ax - b)$$

is a positive-definite quadratic form. The minimizer is found by root finding on the gradient of the objective function,

$$\nabla J(x) = A^T Q^{-1}(Ax - b),$$

which reduces to solving the normal equations

$$A^T Q^{-1} A\hat{x} = A^T Q^{-1} b. \tag{2.5}$$

Note that the Hessian of $J(x)$ is given by

$$D^2 J(x) = A^T Q^{-1} A$$

and is nonsingular by hypothesis; in fact, the inverse Hessian equals the covariance, and is given by (2.3).

Moreover, since $J(x)$ is a positive-definite quadratic form, a single iteration of Newton's method yields the minimizing solution, irrespective of the initial guess, that is,

$$\hat{x} = x - D^2 J(x)^{-1} \nabla J(x) = x - (A^T Q^{-1} A)^{-1} A^T Q^{-1}(Ax - b) \tag{2.6}$$

for all x. This observation is a key insight that we use repeatedly throughout the remainder of this paper. Indeed, the derivation of the Kalman filter presented herein follows by using Newton's method on a certain positive-definite quadratic form, with a judiciously chosen initial guess that greatly simplifies the form of the solution.

3. The Kalman Filter.

We consider the discrete-time linear system

$$x_k = F_k x_{k-1} + G_k u_k + w_k, \tag{3.1a}$$

$$y_k = H_k x_k + v_k, \tag{3.1b}$$

where $x_k \in \mathbb{R}^n$ denotes the state, $y_k \in \mathbb{R}^q$ are the measurements (outputs), $u_k \in \mathbb{R}^p$ is a known sequence of inputs, and w_k and v_k are uncorrelated zero-mean random noise processes with positive-definite covariances $Q_k > 0$ and $R_k > 0$, respectively. We assume that the initial state of the system is $x_0 = \mu_0 + w_0$ for some known $\mu_0 \in \mathbb{R}^n$.

3.1. State Estimation.

We formulate the state estimation problem: Given m known observations $y_1, \ldots, y_m$ and k known inputs $u_1, \ldots, u_k$, where $k \geq m$, we find the best linear unbiased estimate of the states $x_1, \ldots, x_k$ by writing (3.1) as a large linear system

$$\begin{aligned}
\mu_0 &= x_0 && - w_0, \\
G_1 u_1 &= x_1 - F_1 x_0 && - w_1, \\
y_1 &= H_1 x_1 && + v_1, \\
&\ \vdots && \ \vdots \\
G_m u_m &= x_m - F_m x_{m-1} && - w_m, \\
y_m &= H_m x_m && + v_m, \\
G_{m+1} u_{m+1} &= x_{m+1} - F_{m+1} x_m && - w_{m+1}, \\
&\ \vdots && \ \vdots \\
G_k u_k &= x_k - F_k x_{k-1} && - w_k.
\end{aligned} \tag{3.2}$$

In abbreviated form, (3.2) is written as the linear model

$$b_{k|m} = A_{k|m} z_k + \varepsilon_{k|m},$$

where

$$z_k = \begin{bmatrix} x_0 \\ \vdots \\ x_k \end{bmatrix} \in \mathbb{R}^{(k+1)n}$$

and $\varepsilon_{k|m}$ is a zero-mean random variable whose inverse covariance is the positive-definite block-diagonal matrix

$$W_{k|m} = \operatorname{diag}(Q_0^{-1}, Q_1^{-1}, R_1^{-1}, \ldots, Q_m^{-1}, R_m^{-1}, Q_{m+1}^{-1} \ldots, Q_k^{-1}).$$

Thus, the best linear unbiased estimate $\hat{z}_{k|m}$ of z_k is found by solving the normal equations

$$A_{k|m}^T W_{k|m} A_{k|m} \hat{z}_{k|m} = A_{k|m}^T W_{k|m} b_{k|m}.$$

Since each column of $A_{k|m}$ is a pivot column, it follows that $A_{k|m}$ is of full column rank, and thus $A_{k|m}^T W_{k|m} A_{k|m}$ is nonsingular—indeed, it is positive definite. Hence, the best linear unbiased estimator $\hat{z}_{k|m}$ of z_k is given by the weighted least squares solution

$$\hat{z}_{k|m} = (A_{k|m}^T W_{k|m} A_{k|m})^{-1} A_{k|m}^T W_{k|m} b_{k|m}. \tag{3.3}$$

As with (2.1), the weighted least squares solution is the minimizer of the positive-definite objective function

$$J_{k|m}(z_k) = \frac{1}{2}\|A_{k|m} z_k - b_{k|m}\|^2_{W_{k|m}}, \tag{3.4}$$

which can also be written as the sum

$$\begin{aligned} J_{k|m}(z_k) = \frac{1}{2}\|x_0 - \mu_0\|^2_{Q_0^{-1}} + \frac{1}{2}\sum_{i=1}^{m}\|y_i - H_i x_i\|^2_{R_i^{-1}} \\ + \frac{1}{2}\sum_{i=1}^{k}\|x_i - F_i x_{i-1} - G_i u_i\|^2_{Q_i^{-1}}, \end{aligned} \tag{3.5}$$

since $W_{k|m}$ is block diagonal. In the following, we derive the Kalman filter and some of its variants using Newton's method. Since the state estimation problem is reduced to minimizing a positive-definite quadratic form (3.5), one iteration of Newton's method determines our state estimate $\hat{z}_{k|m}$. Moreover, we show that with a judiciously chosen initial guess, we can find a simplified recursive expression for (3.3).

3.2. The Two-Step Kalman Filter. We now consider the state estimation problem in the case that $m = k$, that is, when the number of observations equals the number of inputs. It is of particular interest in applications to estimate the current state x_k given the observations $y_1, \ldots, y_k$ and inputs $u_1, \ldots, u_k$. The Kalman filter gives a recursive algorithm, which is the best linear unbiased estimate $\hat{x}_{k|k}$ of x_k in terms of the previous state estimate $\hat{x}_{k-1|k-1}$ and the latest data u_k and y_k up to that point in time.

The first step of the Kalman filter, called the predictive step, is to determine $\hat{x}_{k|k-1}$ from $\hat{x}_{k-1|k-1}$. For notational convenience throughout this paper, we denote $\hat{z}_{k|k}$ as $\hat{z}_k$, $\hat{x}_{k|k}$ as $\hat{x}_k$, and $J_{k|k}$ as J_k. In addition, we define $\mathcal{F}_k = \begin{bmatrix} 0 & \cdots & 0 & F_k \end{bmatrix} \in \mathbb{R}^{n \times kn}$, with the entry F_k lying in the block corresponding to x_{k-1}, so that $\mathcal{F}_k z_{k-1} = F_k x_{k-1}$. Using this notation, we may write $J_{k|k-1}$ recursively as

$$J_{k|k-1}(z_k) = J_{k-1}(z_{k-1}) + \frac{1}{2}\|x_k - \mathcal{F}_k z_{k-1} - G_k u_k\|^2_{Q_k^{-1}}. \tag{3.6}$$

The gradient and Hessian are given by

$$\nabla J_{k|k-1}(z_k) = \begin{bmatrix} \nabla J_{k-1}(z_{k-1}) + \mathcal{F}_k^T Q_k^{-1}(\mathcal{F}_k z_{k-1} - x_k + G_k u_k) \\ -Q_k^{-1}(\mathcal{F}_k z_{k-1} - x_k + G_k u_k) \end{bmatrix}$$

and

$$D^2 J_{k|k-1}(z_k) = \begin{bmatrix} D^2 J_{k-1}(z_{k-1}) + \mathcal{F}_k Q_k^{-1} \mathcal{F}_k & -\mathcal{F}_k^T Q_k^{-1} \\ -Q_k^{-1} \mathcal{F}_k & Q_k^{-1} \end{bmatrix}, \tag{3.7}$$

respectively. Since $D^2 J_{k|k-1}$ is positive definite, a single iteration of Newton's method yields the minimizer

$$\hat{z}_{k|k-1} = z_k - D^2 J_{k|k-1}(z_k)^{-1} \nabla J_k(z_k) \tag{3.8}$$

for any $z_k \in \mathbb{R}^{(k+1)n}$. Now, $\nabla J_{k-1}(\hat{z}_{k-1}) = 0$ and $\mathcal{F}_k \hat{z}_{k-1} = F_k \hat{x}_{k-1}$, so a judicious initial guess is

$$z_k = \begin{bmatrix} \hat{z}_{k-1} \\ F_k \hat{x}_{k-1} + G_k u_k \end{bmatrix}.$$

With this starting point, the gradient reduces to $\nabla J_{k|k-1}(z_k) = 0$, that is, the optimal estimate of x_k given the measurements $y_1, \ldots, y_{k-1}$ is the bottom row of $\hat{z}_{k|k-1}$ and the covariance, $P_{k|k-1}$, is the bottom right block of the inverse Hessian $D^2 J_{k|k-1}^{-1}$, which by Lemma B.2 in Appendix B is

$$P_{k|k-1} = F_k P_{k-1} F_k^T + Q_k, \tag{3.9a}$$

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1} + G_k u_k. \tag{3.9b}$$

After measuring the output y_k, we perform the second step, which *corrects* the prior estimate-covariance pair $(\hat{x}_{k|k-1}, P_{k|k-1})$ giving a posterior estimate-covariance pair $(\hat{x}_k, P_k)$; this is called the *update* step. In analogy to $\mathcal{F}_k$, we introduce the notation $\mathcal{H}_k = \begin{bmatrix} 0 & \cdots & H_k \end{bmatrix} \in \mathbb{R}^{q \times n(k+1)}$, with the entry H_k lying in the block corresponding to x_k, so that $\mathcal{H}_k z_k = H_k x_k$. The objective function now becomes

$$J_k(z_k) = J_{k|k-1}(z_k) + \frac{1}{2} \|y_k - \mathcal{H}_k z_k\|^2_{R_k^{-1}} \tag{3.10}$$

and the gradient and Hessian are, respectively,

$$\nabla J_k(z_k) = \nabla J_{k|k-1}(z_k) + \mathcal{H}_k^T R_k^{-1}(\mathcal{H}_k z_k - y_k) = \nabla J_{k|k-1}(z_k) + \mathcal{H}_k^T R_k^{-1}(H_k x_k - y_k)$$

and

$$D^2 J_k(z_k) = D^2 J_{k|k-1} + \mathcal{H}_k^T R_k^{-1} \mathcal{H}_k.$$

The Hessian is clearly positive definite. Again, applying a single iteration of Newton's method yields the minimizer. If we choose the initial guess $z_k = \hat{z}_{k|k-1}$, then, since $\nabla J_{k|k-1}(\hat{z}_{k|k-1}) = 0$, the gradient becomes

$$\nabla J_k(\hat{z}_{k|k-1}) = \mathcal{H}_K^T R_k^{-1}(H_k \hat{x}_{k|k-1} - y_k).$$

The estimate $\hat{x}_k$ of x_k, together with its covariance P_k, is then obtained from the bottom row of the Newton update and the bottom right block of the covariance matrix $D^2 J_k(z_k)^{-1}$. Again, using Lemma B.2, this is

$$P_k = (P_{k|k-1}^{-1} + H_k^T R_k^{-1} H_k)^{-1}, \tag{3.11a}$$

$$\hat{x}_k = \hat{x}_{k|k-1} - P_k H_k^T R_k^{-1}(H_k \hat{x}_{k|k-1} - y_k). \tag{3.11b}$$

In some presentations of the Kalman filter, the update step is given as

$$K_k = P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1}, \tag{3.12a}$$

$$\hat{x}_k = \hat{x}_{k|k-1} + K_k(y_k - H_k\hat{x}_{k|k-1}), \tag{3.12b}$$

$$P_k = (I - K_kH_k)P_{k|k-1}, \tag{3.12c}$$

where K_k is called the *Kalman gain*. We remark that (3.11) and (3.12) are equivalent by Lemma B.4.

We refer to (3.9) and (3.11) as the *two-step Kalman filter*. The Kalman filter is often derived, studied, and implemented as a two-step process. However, it is possible to combine the predictive and update steps into a single step by inserting (3.9) into (3.11). This yields the *one-step Kalman filter*

$$P_k = [(Q_{k-1} + F_{k-1}P_{k-1}F_{k-1}^T)^{-1} + H_k^TR_k^{-1}H_k]^{-1}, \tag{3.13a}$$

$$\hat{x}_k = F_{k-1}\hat{x}_{k-1} + G_{k-1}u_{k-1} - P_kH_k^TR_k^{-1}[H_k(F_{k-1}\hat{x}_{k-1} + G_{k-1}u_{k-1}) - y_k]. \tag{3.13b}$$

We can also derive (3.13) directly by considering the objective function

$$J_k(z_k) = J_{k-1}(z_{k-1}) + \frac{1}{2}\|x_k - \mathcal{F}_kz_{k-1} - G_ku_k\|^2_{Q_k^{-1}} + \frac{1}{2}\|y_k - H_kx_k\|^2_{R_k^{-1}}$$

and then applying Newton's method with the same initial guess used in the predictive step; see [18] for details.

An alternate derivation, which closely resembles the derivation presented originally by Kalman in 1960 [22], is given in Appendix C. This approach, which assumes that the noise is normally distributed, computes the optimal estimates by tracking the evolution of the mean and covariance of the distribution of x_k under iteration of the dynamical system.

4. Variations of the Kalman Filter. The Kalman filter provides information about the most recent state given the most recent observation. Specifically, it provides the estimates $\hat{x}_{k-1}$, $\hat{x}_k$, $\hat{x}_{k+1}$, etc. In this section we derive some of the standard variations of the Kalman filter, which primarily deal with the more general cases $\hat{x}_{k|m}$; see [21, 38] for additional variations.

4.1. Predictive Estimates. We return to the system (3.2) and now consider the case that the outputs are known up to time m and the inputs are known up to time $k > m$. We show how to recursively represent the estimate $\hat{x}_{k|m}$ of the state x_k in terms of the previous state estimate $\hat{x}_{k-1|m}$ and the input vector u_k. This allows us to *predict* the states $x_{m+1}, \ldots, x_k$ based on only the estimate $\hat{x}_m$ and the inputs $u_{m+1}, \ldots, u_k$.

We begin by writing the positive-definite quadratic objective (3.5) in recursive form,

$$J_{k|m}(z_k) = J_{k-1|m}(z_{k-1}) + \frac{1}{2}\|x_k - \mathcal{F}_kz_{k-1} - G_ku_k\|^2_{Q_k^{-1}}.$$

Assuming that $\hat{z}_{k-1|m}$ minimizes $J_{k-1|m}(z_{k-1})$ and following the approach in section 3.2, we find that the minimizer of $J_{k|m}(z_k)$ is given by

$$\hat{z}_{k|m} = \begin{bmatrix} \hat{z}_{k-1|m} \\ F_k\hat{x}_{k-1|m} + G_ku_k \end{bmatrix}. \tag{4.1}$$

By taking the bottom row of (4.1), we have the recursive equation

$$\hat{x}_{k|m} = F_k \hat{x}_{k-1|m} + G_k u_k.$$

From this, we see that the best linear unbiased estimate $\hat{x}_{k|m}$ of the state x_k, when only the input u_k and the previous state's estimate $\hat{x}_{k-1|m}$ are known, is found by evolving $\hat{x}_{k-1|m}$ forward by the state equation (3.1a), in expectation, that is, with the noise term w_k set to zero. Note also that the first k vectors in (4.1) do not change. In other words, as time marches on, the estimates of past states $x_0, \ldots, x_{k-1}$ remain unchanged. Hence, we conclude that future inputs affect only the estimates on future states.

To find the covariance on the estimate $\hat{x}_{k|m}$, we compute the Hessian of $J_{k|m}(z_k)$, which is given by

$$D^2 J_{k|m} = \begin{bmatrix} D^2 J_{k-1|m} + \mathcal{F}_k^T Q_k^{-1} \mathcal{F}_k & -\mathcal{F}_k^T Q_k^{-1} \\ -Q_k^{-1} \mathcal{F}_k & Q_k^{-1} \end{bmatrix}. \tag{4.2}$$

Using Lemma B.2 in Appendix B, we invert (4.2) and obtain a closed-form expression for the covariance:

$$D^2 J_{k|m}^{-1} = \begin{bmatrix} D^2 J_{k-1|m}^{-1} & D^2 J_{k-1|m}^{-1} \mathcal{F}_k^T \\ \mathcal{F}_k D^2 J_{k-1|m}^{-1} & Q_k + \mathcal{F}_k D^2 J_{k-1|m}^{-1} \mathcal{F}_k^T \end{bmatrix}. \tag{4.3}$$

To compute the covariance at time k, we look at the bottom right block of (4.3). Since $\mathcal{F}_k D^2 J_{k-1|m}^{-1} \mathcal{F}_k^T = F_k P_{k-1|m} F_k^T$, where $P_{k-1|m}$ is the bottom right block of $D^2 J_{k-1|m}^{-1}$ and represents the covariance of $\hat{x}_{k-1|m}$, we have that

$$P_{k|m} = F_k P_{k-1|m} F_k^T + Q_k.$$

Note also that the upper-left block of (4.3) is given by $D^2 J_{k-1|m}^{-1}$, or rather the covariance of $\hat{z}_{k-1|m}$. Hence, the covariances of the estimates $\hat{x}_{1|m}, \ldots, \hat{x}_{k-1|m}$ also remain unchanged as more inputs are added.

To summarize, we have the following recursive algorithm for predicting the future states and covariances of a system:

$$P_{k|m} = F_k P_{k-1|m} F_k^T + Q_k, \tag{4.4a}$$
$$\hat{x}_{k|m} = F_k \hat{x}_{k-1|m} + G_k u_k. \tag{4.4b}$$

4.2. Smoothed Estimates. The standard Kalman filter provides estimates of only the current state x_k at each time k. In this section we show how to update the best linear unbiased estimates on all of the previous states as new outputs are observed. Specifically, we estimate the entire vector z_k of states by recursively expressing $\hat{z}_{k|m}$ in terms of $\hat{z}_{k|m-1}$ and y_m. This process is called *smoothing*.

Consider the positive-definite quadratic objective (3.5) in recursive form,

$$J_{k|m}(z_k) = J_{k|m-1}(z_k) + \frac{1}{2}\|y_m - \mathcal{H}_m z_k\|^2_{R_m^{-1}},$$

where $\mathcal{H}_m$ is the $q \times (k+1)n$ matrix

$$\mathcal{H}_m = \begin{bmatrix} 0 & \cdots & H_m & \cdots & 0 \end{bmatrix}.$$

Note that H_m occurs on the $m+1$ block of $\mathcal{H}_m$, since the indexing of z_k begins at $k=0$; therefore, $\mathcal{H}_m z_k = H_m x_m$. To find the best linear unbiased estimate $\hat{z}_{k|m}$ of z_k, we compute the gradient and Hessian of $J_{k|m}(z_k)$ as

$$\nabla J_{k|m}(z_k) = \nabla J_{k|m-1}(z_k) + \mathcal{H}_m^T R_m^{-1}(\mathcal{H}_m z_k - y_m)$$

and

$$D^2 J_{k|m} = D^2 J_{k|m-1} + \mathcal{H}_m^T R_m^{-1} \mathcal{H}_m,$$

respectively. As with the Kalman filter case (3.10) above, we see immediately that $D^2 J_{k|m} > 0$. Thus, the estimator is found by taking a single Newton step

$$\hat{z}_{k|m} = z_k - D^2 J_{k|m}^{-1} \nabla J_{k|m}(z_k),$$

for an arbitrary choice $z_k \in \mathbb{R}^{(k+1)n}$. Setting $z_k = \hat{z}_{k|m-1}$ gives $\nabla J_{k|m-1}(z_k) = 0$. Hence Newton's method implies that

$$\hat{z}_{k|m} = \hat{z}_{k|m-1} - D^2 J_{k|m}^{-1} \mathcal{H}_m^T R_m^{-1}(\mathcal{H}_m \hat{z}_{k|m-1} - y_m).$$

Noting that $\mathcal{H}_m \hat{z}_{k|m-1} = H_m \hat{x}_{m|m-1}$, it follows that

$$\hat{z}_{k|m} = \hat{z}_{k|m-1} - L_{k|m}^{(m)} H_m^T R_m^{-1}(H_m \hat{x}_{m|m-1} - y_m),$$

where $L_{k|m}^{(\ell)}$ denotes the $(\ell+1)$th column of blocks of $D^2 J_{k|m}^{-1}$ (recall that $D^2 J_{k|m}^{-1}$ is a block $(k+1) \times (k+1)$ matrix since the first column corresponds to $\ell = 0$). Noting the identity $D^2 J_{k|m}^{-1} \mathcal{H}_\ell^T = L_{k|m}^{(\ell)} H_\ell^T$ and Lemma B.3, we have

$$\begin{aligned}
D^2 J_{k|m}^{-1} &= (D^2 J_{k|m-1} + \mathcal{H}_m^T R_m^{-1} \mathcal{H}_m)^{-1} \\
&= D^2 J_{k|m-1}^{-1} - D^2 J_{k|m-1}^{-1} \mathcal{H}_m^T (R_m + \mathcal{H}_m D^2 J_{k|m-1}^{-1} \mathcal{H}_m^T)^{-1} \mathcal{H}_m D^2 J_{k|m-1}^{-1} \\
&= D^2 J_{k|m-1}^{-1} - L_{k|m-1}^{(m)} H_m^T (R_m + H_m P_{k|m-1}^{(m,m)} H_m^T)^{-1} H_m L_{k|m-1}^{(m)}{}^T,
\end{aligned}$$

where $P_{k|m}^{(i,j)}$ is the $(i+1, j+1)$-block of $D^2 J_{k|m}^{-1}$ and hence also the $(i+1)$-block of $L_{k|m}^{(j)}$. Therefore the term $L_{k|m}^{(\ell)}$ is recursively defined as

$$L_{k|m}^{(\ell)} = L_{k|m-1}^{(\ell)} - L_{k|m-1}^{(m)} H_m^T (R_m + H_m P_{k|m-1}^{(m,m)} H_m^T)^{-1} H_m P_{k|m-1}^{(\ell,m)}{}^T.$$

Of particular value is the case when $\ell = m$. Then we have

$$L_{k|m}^{(m)} = L_{k|m-1}^{(m)} - L_{k|m-1}^{(m)} H_m^T (R_m + H_m P_{k|m-1}^{(m,m)} H_m^T)^{-1} H_m P_{m|m-1}^{(m,m)}, \tag{4.5a}$$

$$\hat{z}_{k|m} = \hat{z}_{k|m-1} - L_{k|m}^{(m)} H_m^T R_m^{-1} (H_m \hat{x}_{k|m-1} - y_m). \tag{4.5b}$$

This allows us to update the entire vector z_k of state estimates.

4.3. Fixed-Lag Smoothing. In the previous subsection, we showed how to recursively estimate the entire vector z_k as new outputs are observed. Suppose instead that we only want to estimate some of the last few states. Fixed-lag smoothing is the process of updating the previous ℓ states, that is, $x_{k-\ell}, x_{k-\ell+1}, \ldots, x_{k-2}, x_{k-1}$, as the latest output y_k is measured.

When $m = k$, (4.5b) becomes

$$\hat{z}_{k|k} = \begin{bmatrix} \hat{z}_{k-1|k} \\ \hat{x}_k \end{bmatrix} = \begin{bmatrix} \hat{z}_{k-1} \\ \hat{x}_{k|k-1} \end{bmatrix} - \begin{bmatrix} L_k \\ P_k \end{bmatrix} H_k^T R_k^{-1}(H_k \hat{x}_{k|k-1} - y_k),$$

where $\hat{x}_{k|k-1} = F_k \hat{x}_{k-1} + G_k u_k$, P_k is the covariance of the estimator $\hat{x}_k$, which is found in the bottom right corner of the inverse Hessian $D^2 J_k^{-1}$ of the objective function J_k, and L_k is the upper-right corner of $D^2 J_k^{-1}$. Using (3.7) and Lemma B.2, we see that

$$L_k = (D^2 J_{k-1} + \mathcal{F}_k^T Q_k^{-1} \mathcal{F}_k)^{-1} \mathcal{F}_k^T Q_k^{-1} P_k = \begin{bmatrix} \bar{L} \\ \bar{P} \end{bmatrix} F_k^T Q_k^{-1} P_k,$$

where $\bar{L}$ and $\bar{P}$ are, respectively, the upper-right and lower-right blocks of

$$\begin{aligned} &(D^2 J_{k-1} + \mathcal{F}_k^T Q_k^{-1} \mathcal{F}_k)^{-1} \\ &\quad = \begin{bmatrix} D^2 J_{k-2} + \mathcal{F}_{k-1}^T Q_{k-1}^{-1} \mathcal{F}_{k-1} & -\mathcal{F}_{k-1}^T Q_{k-1}^{-1} \\ -Q_{k-1}^{-1} \mathcal{F}_{k-1} & Q_{k-1}^{-1} + H_{k-1}^T R_{k-1}^{-1} H_{k-1} + F_k^T Q_k^{-1} F_k \end{bmatrix}^{-1}. \end{aligned}$$

Thus, by Lemmas B.2 and B.3 we have that

$$\begin{aligned} \bar{P} &= (Q_{k-1}^{-1} + H_{k-1}^T R_{k-1}^{-1} H_{k-1} + F_k^T Q_k^{-1} F_k \\ &\qquad - Q_{k-1}^{-1} \mathcal{F}_{k-1} (D^2 J_{k-2} + \mathcal{F}_{k-1}^T Q_{k-1}^{-1} \mathcal{F}_{k-1})^{-1} \mathcal{F}_{k-1}^T Q_{k-1}^{-1})^{-1} \\ &= ((Q_{k-1}^{-1} + F_{k-1}^T P_{k-1}^{-1} F_{k-1}^T)^{-1} + H_{k-1}^T R_{k-1}^{-1} H_{k-1} + F_k^T Q_k^{-1} F_k)^{-1} \\ &= (P_{k-1}^{-1} + F_k^T Q_k^{-1} F_k)^{-1}, \end{aligned}$$

and from Lemma B.2 it follows that

$$\bar{L} = (D^2 J_{k-2} + \mathcal{F}_{k-1}^T Q_{k-1}^{-1} \mathcal{F}_{k-1})^{-1} \mathcal{F}_{k-1}^T Q_{k-1}^{-1} \bar{P}.$$

Notice that the forms of $\bar{P}$ and $\bar{L}$ are the same as those of P_k and L_k. Thus we can find $\hat{x}_{k-i|k}$ recursively, for $i = 1, \ldots, \ell$, by the system

$$\Omega_{k-i} = (P_{k-i}^{-1} + F_{k-i+1}^T Q_{k-i+1}^{-1} F_{k-i+1})^{-1} F_{k-i+1}^T Q_{k-i+1}^{-1} \Omega_{k-i+1}, \tag{4.6a}$$

$$\hat{x}_{k-i|k} = \hat{x}_{k-i|k-1} - \Omega_{k-i} H_k^T R_k^{-1}(H_k \hat{x}_{k|k-1} - y_k), \tag{4.6b}$$

where $\Omega_k = P_k$, $\hat{x}_0 = \mu_0$, and $P_0 = Q_0$.

4.4. Fading Memory. Although the Kalman filter is a recursive algorithm that updates the current state variable as the latest inputs and measurements become known, the estimation is based on the least squares solution of all the previous states where all measurements are weighted according to their covariance. We now consider the case that the estimator discounts the error in older measurements leading to a greater emphasis on recent observations. This is particularly useful in situations where there is some modeling error in the system.

We minimize the positive-definite objective

$$\begin{aligned} J_k(z_k) = {} & \frac{\lambda^k}{2} \|x_0 - \mu_0\|_{Q_0^{-1}}^2 + \frac{1}{2} \sum_{i=1}^{k} \lambda^{k-i} \|y_i - H_i x_i\|_{R_i^{-1}}^2 \\ & + \frac{1}{2} \sum_{i=1}^{k} \lambda^{k-i} \|x_i - F_i x_{i-1} - G_i u_i\|_{Q_i^{-1}}^2, \end{aligned} \tag{4.7}$$

where $0 < \lambda \leq 1$ is called the *forgetting factor*. We say that the state estimator has perfect memory when $\lambda = 1$, reducing to (3.5); it becomes increasingly forgetful as λ decreases.

Note that (4.7) can be written recursively as

$$J_k(z_k) = \lambda J_{k-1}(z_{k-1}) + \frac{1}{2}\|y_k - H_k x_k\|^2_{R_k^{-1}} + \frac{1}{2}\|x_k - \mathcal{F}_k z_{k-1} - G_k u_k\|^2_{Q_k^{-1}}.$$

The gradient and Hessian are computed to be

$$\nabla J_k = \begin{bmatrix} \lambda \nabla J_{k-1}(z_{k-1}) + \mathcal{F}_k^T Q_k^{-1}(\mathcal{F}_k z_{k-1} - x_k + G_k u_k) \\ H_k^T R_k^{-1}(H_k x_k - y_k) - Q_k^{-1}(\mathcal{F}_k z_{k-1} - x_k + G_k u_k) \end{bmatrix}$$

and

$$D^2 J_k = \begin{bmatrix} \lambda D^2 J_{k-1}(z_{k-1}) + \mathcal{F}_k^T Q_k^{-1}\mathcal{F}_k & -\mathcal{F}_k^T Q_k^{-1} \\ -Q_k^{-1}\mathcal{F}_k & Q_k^{-1} + H_k^T R_k^{-1} H_k \end{bmatrix},$$

respectively. As with (3.10), one can show inductively that $D^2 J_k > 0$. Thus we can likewise minimize (4.7) using (3.8). Since $\nabla J_{k-1}(\hat{z}_{k-1}) = 0$ and $\mathcal{F}_k \hat{z}_{k-1} = F_k \hat{x}_{k-1}$, we make use of the initializing choice

$$z_k = \begin{bmatrix} \hat{z}_{k-1} \\ F_k \hat{x}_{k-1} + G_k u_k \end{bmatrix},$$

thus resulting in

$$\nabla J_k(z_k) = \begin{bmatrix} 0 \\ H_k^T R_k^{-1}[H_k(F_k \hat{x}_{k-1} + G_k u_k) - y_k] \end{bmatrix}.$$

Hence the bottom row of (3.8) simplifies to

$$\hat{x}_k = F_k \hat{x}_{k-1} + G_k u_k - P_k H_k^T R_k^{-1}[H_k(F_k \hat{x}_{k-1} + G_k u_k) - y_k].$$

This update is exactly the same as that of the one-step Kalman filter and depends on λ only in the update for P_k, which is obtained by taking the bottom right block of the inverse Hessian $D^2 J_k^{-1}$ and using Lemma B.2 as appropriate. We find that the covariance is given by

$$P_k = ((Q_k + \lambda F_k P_{k-1} F_k^T)^{-1} + H_k^T R_k^{-1} H_k)^{-1}.$$

Thus, to summarize, we have the recursive estimate for $\hat{x}_k$ given by

$$\begin{aligned} P_k &= ((Q_k + \lambda F_k P_{k-1} F_k^T)^{-1} + H_k^T R_k^{-1} H_k)^{-1}, \\ \hat{x}_k &= F_k \hat{x}_{k-1} + G_k u_k - P_k H_k^T R_k^{-1}[H_k(F_k \hat{x}_{k-1} + G_k u_k) - y_k], \end{aligned}$$

where $\hat{x}_0 = \mu_0$ and $P_0 = Q_0$.

5. The Extended Kalman Filter. We derive the extended Kalman filter (EKF), which is a recursive state estimation algorithm for noisy nonlinear systems of the form

$$x_k = f_k(x_{k-1}, u_k) + w_k, \tag{5.1a}$$

$$y_k = h_k(x_k) + v_k, \tag{5.1b}$$

where, as in (3.1), $x_k \in \mathbb{R}^n$ denotes the state, $y_k \in \mathbb{R}^q$ are the outputs, and $u_k \in \mathbb{R}^p$ the inputs. The noise processes w_k and v_k are uncorrelated and zero mean with positive-definite covariances $Q_k > 0$ and $R_k > 0$, respectively.

In this section, we show that the EKF can be derived by using Newton's method, as in previous sections. However, we first present the classical derivation of the EKF, which follows by approximating (5.1) with a linear system and then using the Kalman filter on that linear system. This technique goes back to Schmidt [37], who is often credited as being the first to implement the Kalman filter.

In contrast to the linear Kalman filter, the EKF is neither the unbiased minimum mean-squared error estimator nor the minimum variance unbiased estimator of the state. In fact, the EFK is generally biased. However, the EKF is the best linear unbiased estimator of the linearized dynamical system, which can often be a good approximation of the nonlinear system (5.1). As a result, how well the local linear dynamics match the nonlinear dynamics determines in large part how well the EKF will perform. For decades, the EKF has been the de facto standard for nonlinear state estimation; however, in recent years other contenders have emerged, such as the unscented Kalman filter (UKF) [20, 46] and particle filters [9]. Nonetheless, the EKF is still widely used in applications. Indeed, even though the EKF is the right answer to the wrong problem, some problems are less wrong than others. George Box was often quoted as saying "all models are wrong, but some are useful" [6]. Along these lines, one might conclude that the EKF is wrong, but sometimes it is useful!

5.1. Classical EKF Derivation. To linearize (5.1), we take the first-order expansions of f at $\hat{x}_{k-1}$ and h at $\hat{x}_{k|k-1}$,

$$x_k = f_k(\hat{x}_{k-1}, u_k) + Df_k(\hat{x}_{k-1}, u_k)(x_k - \hat{x}_{k-1}) + w_k, \tag{5.2a}$$
$$y_k = h_k(\hat{x}_{k|k-1}) + Dh_k(\hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1}) + v_k, \tag{5.2b}$$

which can be equivalently written as the linear system

$$x_k = F_k x_{k-1} + \tilde{u}_k + w_k, \tag{5.3a}$$
$$y_k = H_k x_k + z_k + v_k, \tag{5.3b}$$

where

$$\begin{aligned} F_k &= Df_k(\hat{x}_{k-1}, u_k), \\ H_k &= Dh_k(\hat{x}_{k|k-1}), \end{aligned} \tag{5.4}$$

and

$$\begin{aligned} \tilde{u}_k &= f_k(\hat{x}_{k-1}, u_k) - F_k \hat{x}_{k-1}, \\ z_k &= h_k(\hat{x}_{k|k-1}) - H_k \hat{x}_{k|k-1}. \end{aligned} \tag{5.5}$$

We treat $\tilde{u}_k$ and $y_k - z_k$ as the input and output of the linear approximate system (5.3), respectively. For simplicity, we use the two-step Kalman filter on (5.3), which is given by the prediction and innovation equations (3.9) and (3.11), respectively, or rather

$$P_{k|k-1} = F_k P_{k-1} F_k^T + Q_k, \tag{5.6a}$$
$$\hat{x}_{k|k-1} = f_k(\hat{x}_{k-1}, u_k), \tag{5.6b}$$
$$P_k = (P_{k|k-1}^{-1} + H_k^T R_k^{-1} H_k)^{-1}, \tag{5.6c}$$
$$\hat{x}_k = \hat{x}_{k|k-1} - P_k H_k^T R_k^{-1}(h_k(\hat{x}_{k|k-1}) - y_k). \tag{5.6d}$$

This is the EKF. Notice the use of (5.5) in writing the right-hand sides of (5.6b) and (5.6d).

As with the two-step Kalman filter, we can use Lemma B.4 to write the innovation equations (5.6c)–(5.6d) in the alternative form

$$K_k = P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1}, \tag{5.7a}$$

$$\hat{x}_k = \hat{x}_{k|k-1} + K_k(h_k(\hat{x}_{k|k-1}) - y_k), \tag{5.7b}$$

$$P_k = (I - K_kH_k)P_{k|k-1}. \tag{5.7c}$$

5.2. The Newton-EKF Derivation. We now derive the EKF via Newton's method. By generalizing the objective function in the linear case (3.5) to

$$J_{k|m}(z_k) = \frac{1}{2}\sum_{i=1}^{m}\|y_i - h_i(x_i)\|^2_{R_i^{-1}} + \frac{1}{2}\sum_{i=0}^{k}\|x_i - f_i(x_{i-1}, u_i)\|^2_{Q_i^{-1}}, \tag{5.8}$$

where $f_0 \equiv \mu_0$, we can determine the estimate $\hat{z}_k$ of z_k by minimizing the objective. Assume $m = k$; then we obtain the EKF, as in (5.6), by minimizing (5.8) by a single Newton step with a carefully chosen initial guess. As in the two-step Kalman filter, we write (5.8) as the two recursive equations

$$J_{k|k-1}(z_k) = J_{k-1}(z_{k-1}) + \frac{1}{2}\|x_k - f_k(x_{k-1}, u_k)\|^2_{Q_k^{-1}} \tag{5.9}$$

and

$$J_k(z_k) = J_{k|k-1}(z_k) + \frac{1}{2}\|y_k - h_k(x_k)\|^2_{R_k^{-1}}, \tag{5.10}$$

representing the prediction and update steps, respectively.

We begin with the prediction step, writing (5.9) as

$$J_{k|k-1}(z_k) = J_{k-1}(z_{k-1}) + \frac{1}{2}\|x_k - \mathcal{F}_k(z_{k-1})\|^2_{Q_k^{-1}},$$

where $\mathcal{F}_k(z_{k-1}) = f_k(x_{k-1}, u_k)$; for notational convenience we suppress writing the inputs u_k. The gradient takes the form

$$\nabla J_{k|k-1}(z_k) = \begin{bmatrix} \nabla J_{k-1}(z_{k-1}) - D\mathcal{F}_k(z_{k-1})^T Q_k^{-1}(x_k - \mathcal{F}_k(z_{k-1})) \\ Q_k^{-1}(x_k - \mathcal{F}_k(z_{k-1})) \end{bmatrix}.$$

We define the estimate $\hat{z}_{k|k-1}$ of z_k to be the minimizer of $J_{k|k-1}(z_k)$. This holds when the gradient is zero, which occurs when

$$\hat{z}_{k|k-1} = \begin{bmatrix} \hat{z}_{k-1} \\ \mathcal{F}_k(\hat{z}_{k-1}) \end{bmatrix} = \begin{bmatrix} \hat{z}_{k-1} \\ f_k(\hat{x}_{k-1}, u_k) \end{bmatrix}.$$

Thus, the bottom row yields

$$\hat{x}_{k|k-1} = f_k(\hat{x}_{k-1}, u_k),$$

which is (5.6b). To compute the covariance of the estimate $\hat{x}_{k|k-1}$, we consider the Hessian of $J_{k|k-1}(z_k)$, which is given by

$$D^2 J_{k|k-1}(z_k) = \begin{bmatrix} D^2 J_{k-1}(z_{k-1}) + \Theta(z_k) & D\mathcal{F}_k(z_{k-1})^T Q_k^{-1} \\ Q_k^{-1} D\mathcal{F}_k(z_{k-1}) & Q_k^{-1} \end{bmatrix},$$

where

$$\Theta(z_k) = D^2\mathcal{F}_k(z_{k-1})Q_k^{-1}(x_k - \mathcal{F}_k(z_{k-1})) + D\mathcal{F}_k(z_{k-1})^T Q_k^{-1} D\mathcal{F}_k(z_{k-1}).$$

Note that when $z_k = \hat{z}_{k|k-1}$, it follows that

$$\Theta(\hat{z}_{k|k-1}) = D\mathcal{F}_k(\hat{z}_{k-1})^T Q_k^{-1} D\mathcal{F}_k(\hat{z}_{k-1}).$$

Thus, by taking the lower-right block of the inverse of $D^2 J_{k|k-1}(\hat{z}_k)$, we have

$$\begin{aligned} P_{k|k-1} &= Q_k^{-1} + D\mathcal{F}_k(\hat{z}_{k-1}) D^2 J_{k-1}(\hat{z}_{k-1})^{-1} D\mathcal{F}_k(\hat{z}_{k-1})^T \\ &= Q_k^{-1} + DF_k(\hat{z}_{k-1}) P_{k-1} DF_k(\hat{z}_{k-1})^T \\ &= Q_{k-1}^{-1} + F_k P_{k-1} F_k^T, \end{aligned}$$

which is (5.6a).

Now we find the innovation equations (5.6c)–(5.6d) by rewriting (5.9) as

$$J_k(z_k) = J_{k|k-1}(z_k) + \|y_k - \mathcal{H}_k(z_k)\|^2_{R_k^{-1}},$$

where $\mathcal{H}_k(z_k) = h_k(x_k)$. The gradient and Hessian are written as

$$\nabla J_k(z_k) = \nabla J_{k|k-1}(z_k) - D\mathcal{H}_k(z_k)^T R_k^{-1}(y_k - \mathcal{H}_k(z_k))$$

and

$$D^2 J_k(z_k) = D^2 J_{k|k-1}(z_k) - D^2\mathcal{H}_k(z_k) R_k^{-1}(y_k - \mathcal{H}_k(z_k)) + D\mathcal{H}_k(z_k)^T R_k^{-1} D\mathcal{H}_k(z_k),$$

respectively. Setting $z_k = \hat{z}_{k|k-1}$, we have

$$\nabla J_k(z_k) = \begin{bmatrix} 0 \\ -H_k^T R_k^{-1}(y_k - h_k(\hat{x}_{k|k-1})) \end{bmatrix}$$

and

$$D^2 J_k(z_k) = D^2 J_{k|k-1}(z_k) + D\mathcal{H}_k(z_k)^T R_k^{-1} D\mathcal{H}_k(z_k).$$

Thus, from Newton's method we have

$$\hat{z}_k = \hat{z}_{k|k-1} - D^2 J_k(\hat{z}_{k|k-1})^{-1} \nabla J_k(\hat{z}_{k|k-1}),$$

where as before the bottom right block of $D^2 J_k(\hat{z}_{k|k-1})^{-1}$ is given by

$$\begin{aligned} P_k &= P_{k|k-1} + P_{k|k-1} H_k^T (R_k - H_k P_{k|k-1} H_k^T)^{-1} H_k P_{k|k-1} \\ &= (P_{k|k-1}^{-1} + H_k^T R_k^{-1} H_k)^{-1}, \end{aligned}$$

which is (5.6c). Taking the bottom row of the Newton equation gives

$$\hat{x}_k = \hat{x}_{k|k-1} - P_k H_k^T R_k^{-1}(h_k(\hat{x}_{k|k-1}) - y_k),$$

which is (5.6d).

6. Conclusions. In this paper, we have shown how the Kalman filter and some of its variants can be derived using Newton's method. One advantage of this approach is that it requires little more than multivariable calculus and linear algebra, and it should be more accessible to students. We would be interested to know if other variants of the Kalman filter can also be derived in this way, for example, [2, 8].

7. Activity for the Classroom. In this section, we consider the problem of estimating the position and velocity of a projectile, say, an artillery round, given a few noisy measurements of its position. We show through a series of exercises that the Kalman filter can average out the noise in the system and provide a relatively smooth profile of the projectile as it passes by a radar sensor. We then show that one can effectively predict the point of impact as well as the point of origin, so that troops on the ground can both duck for cover and return fire before the projectile lands. Although we computed the figures below in MATLAB, one could easily reproduce this work in another computing environment.

7.1. Problem Formulation. Assume that the state of the projectile is given by the vector $x = \begin{pmatrix} s_x & s_y & v_x & v_y \end{pmatrix}^T$, where s_x and s_y are the horizontal and vertical components of position, respectively, with corresponding velocity components v_x and v_y. We suppose that state evolves according to the discrete-time dynamical system

$$x_{k+1} = Fx_k + u + w_k,$$

where

$$F = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1-b & 0 \\ 0 & 0 & 0 & 1-b \end{pmatrix} \quad \text{and} \quad u = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -g\Delta t \end{pmatrix}.$$

In this model, Δt is the interval of time between measurements, $0 \le b \ll 1$ is the drag coefficient, g is the gravitational constant, and the noise process w_k has zero mean with covariance $Q_k > 0$. Since the radar device is only able to measure the position of the projectile, we write the observation equation as

$$y_k = Hx_k + v_k, \quad \text{where} \quad H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

and the measurement noise v_k has zero mean with covariance $R_k > 0$.

7.2. Exercises. Throughout this classroom activity, we assume that $b = 10^{-4}$, $g = 9.8$, and $\Delta t = 10^{-1}$. For simplicity, we also assume $Q = 0.1 \cdot I_4$ and $R = 500 \cdot I_2$, though in practice these matrices would probably not be diagonal.

Exercise 1. Using an initial state $x_0 = \begin{pmatrix} 0 & 0 & 300 & 600 \end{pmatrix}^T$, evolve the system forward 1200 steps. Hint: The noise term w_k can be generated by using the transpose of the Cholesky factorization of Q; specifically, set `w = chol(Q)'*randn(4,1)`.

Exercise 2. Now assume that your radar system can only detect the projectile between the 400th and 600th time steps. Using the measurement equation, produce a plot of the projectile path and the noisy measurements. The noise term v_k can be generated similarly to the exercise above.

Exercise 3. Initialize the Kalman filter at $k = 400$. Use the position coordinates y_{400} to initialize the position and take the average velocity over 10 or so measurements to provide a rough velocity estimate. Use a large initial estimator covariance such as $P_{400} = 10^6 \cdot Q$. Then, using the Kalman filter, compute the next 200 state estimates using (3.9) and (3.12). Plot the estimated position over the graph of the previous exercise. Your image should be similar to Figures 7.1(a) and 7.1(b), the latter being a zoomed version of the former. In Figure 7.1(c), we see the errors generated by the measurement error as well as the estimation error. Note that the estimation error is much smaller than the measurement error.

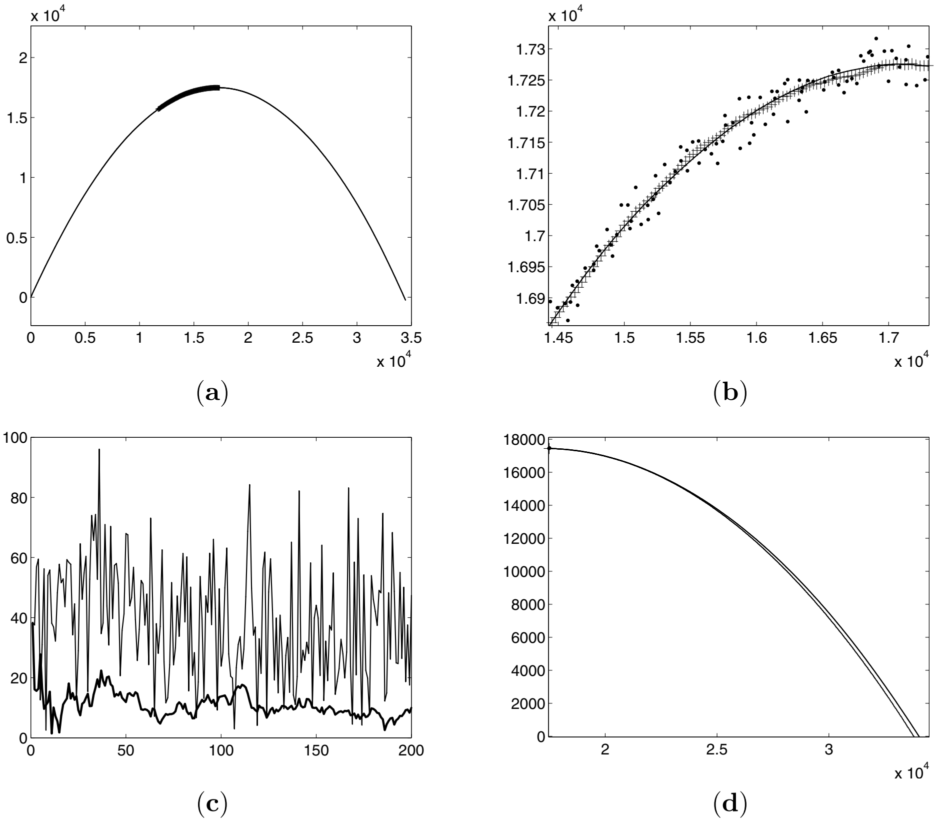

Fig. 7.1 *A plot of* (a) *the trajectory, the measurement points, and the estimates given by the Kalman filter. The dark region on the upper-left part of the trajectory corresponds to the part of the profile where measurement and state estimations are made. In* (b) *a zoom of the previous image is given. The measurements are peppered dots with* (+) *denoting the estimated points from the Kalman filter and the dotted line being the true profile. Note that the estimated and true positions are almost indistinguishable. In* (c) *a plot of errors produced by measurements and the Kalman filter is given. Notice that the estimation error is much smaller than the measurement error. Finally, in* (d) *a plot of the predicted trajectory is given. These lines are close, with the estimation error at the point of impact being roughly half a percent.*

Exercise 4. Using the last state estimate from the previous exercise $\hat{x}_{600}$, use the predictive estimation method described in (4.4) to trace out a projectile until the y-component of position crosses zero, that is, $s_y \approx 0$. Plot the path of the estimate against the true (noiseless) path and see how close the estimated point of impact is when compared with the true point of impact. Your graph should look like Figure 7.1(d).

Exercise 5 (bonus problem). Estimate the point of origin of the projectile by reversing the system and considering the problem

$$x_k = F^{-1}x_{k+1} - F^{-1}u - F^{-1}w.$$

This allows one to iterate backward in time.

Appendix A. Proof of the Gauss–Markov Theorem. In this section, we present a proof of Theorem 2.1, which states that among all linear unbiased estimators $\hat{x} =$

Kx, the choice of K that minimizes the mean-squared error $\mathbb{E}[\|\hat{x} - x\|^2]$ and the covariance $\mathbb{E}[(\hat{x} - x)(\hat{x} - x)^T]$ is

$$K = (A^T Q^{-1} A)^{-1} A^T Q^{-1}.$$

Recall that, since $\hat{x}$ is unbiased, $KA = I$. Therefore,

$$\|\hat{x} - x\|^2 = \|Kb - x\|^2 = \|K(Ax + \varepsilon) - x\|^2 = \|K\varepsilon\|^2 = \varepsilon^T K^T K \varepsilon.$$

Since $\varepsilon^T K^T K \varepsilon$ is a scalar and $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, we have

$$\mathbb{E}[\|\hat{x} - x\|^2] = \mathbb{E}[\mathrm{tr}(\varepsilon^T K^T K \varepsilon)] = \mathbb{E}[\mathrm{tr}(K\varepsilon\varepsilon^T K^T)] = \mathrm{tr}(K\mathbb{E}[\varepsilon\varepsilon^T]K^T) = \mathrm{tr}(KQK^T).$$

Thus the linear unbiased estimator $\hat{x} = Kb$ with minimum mean-squared error satisfies the optimization problem

$$\begin{array}{ll} \underset{K \in \mathbb{R}^{n \times m}}{\text{minimize}} & \mathrm{tr}(KQK^T) \\ \text{subject to} & KA = I. \end{array}$$

This is a convex optimization problem, that is, the objective function is convex and the feasible set is convex. Hence, it suffices to find the unique critical point for the Lagrangian

$$\mathcal{L}(K, \lambda) = \mathrm{tr}(KQK^T) - \mathrm{vec}(\lambda)^T \mathrm{vec}(KA - I),$$

where λ is an $n \times n$ matrix. Simplifying gives

$$\mathcal{L}(K, \lambda) = \mathrm{tr}(KQK^T - \lambda^T(KA - I)).$$

Thus the minimum occurs when

$$0 = \nabla_K \mathcal{L}(K, \lambda) = \frac{\partial}{\partial K} \mathrm{tr}(KQK^T - \lambda^T(KA - I)) = KQ^T + KQ - \lambda A^T.$$

Since Q is symmetric, this reduces to

$$0 = \nabla_K \mathcal{L}(K, \lambda) = 2KQ - \lambda A^T.$$

By multiplying on the right by $Q^{-1}A$ and using the constraint $KA = I$, we have that $\lambda = 2(A^T Q^{-1} A)^{-1}$ and $K = (A^T Q^{-1} A)^{-1} A^T Q^{-1}$. This yields (2.2).

We can verify that (2.2) is unbiased by noting that

$$\hat{x} = (A^T Q^{-1} A)^{-1} A^T Q^{-1}(Ax + \varepsilon) = x + (A^T Q^{-1} A)^{-1} A^T Q^{-1} \varepsilon.$$

Hence $\mathbb{E}[\hat{x}] = x$. Moreover the covariance of the estimate (2.2) is given by

$$\mathbb{E}[(\hat{x} - x)(\hat{x} - x)^T] = (A^T Q^{-1} A)^{-1} A^T Q^{-1} \mathbb{E}[\varepsilon\varepsilon^T] Q^{-1} A (A^T Q^{-1} A)^{-1},$$

which reduces to (2.3).

If $\hat{x}_L = Lb$ is a linear unbiased estimator of (2.1), then $L = K + D$ for some matrix D satisfying $DA = 0$. The variance of $\hat{x}_L$ is given by

$$\begin{aligned}\mathbb{E}[(\hat{x}_L - x)(\hat{x}_L - x)^T] &= \mathbb{E}[(K + D)\varepsilon\varepsilon^T(K^T + D^T)]\\ &= (K + D)Q(K^T + D^T)\\ &= KQK^T + DQD^T + KQD^T + (KQD^T)^T.\end{aligned}$$

Since $DA = 0$, we have that

$$KQD^T = (A^TQ^{-1}A)^{-1}A^TQ^{-1}QD^T = (A^TQ^{-1}A)^{-1}(DA)^T = 0.$$

Therefore,

$$\mathbb{E}[(\hat{x}_L - x)(\hat{x}_L - x)^T] = KQK^T + DQD^T \geq KQK^T,$$

and so (2.2) is indeed the minimum variance linear unbiased estimator of (2.1), and this completes the proof.

Appendix B. Lemmas from Linear Algebra. We provide some technical results that are used throughout the paper. The first lemma shows that the Hessian matrix in (3.7) is positive definite. The second result tells us how to invert block matrices. The third lemma, namely, the Sherman–Morrison–Woodbury formula, shows how to invert an additively updated matrix when we know the inverse of the original; see [17] for a thorough historical treatise on these inversion results. The last lemma provides some matrix manipulations needed to compute various forms of the covariance matrices in the paper.

Lemma B.1. *Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{m\times m}$, and $D \in \mathbb{R}^{m\times m}$, with $A, C > 0$ and $D \geq 0$. Then*

$$\begin{bmatrix} A + B^TCB & -B^TC \\ -CB & C + D \end{bmatrix} > 0. \tag{B.1}$$

As a result, the matrix (3.7) *is positive definite.*

Proof. Note that

$$\begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} A + B^TCB & -B^TC \\ -CB & C + D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x^TAx + y^TDy + \|Bx - y\|_C^2 \geq 0,$$

with equality only if each term on the right is zero. Since $A > 0$, we get equality only if $x = 0$, which also implies that $y = 0$. Thus (B.1) is positive definite. ∎

Lemma B.2 (Schur). *Let M be a square matrix with block form*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

If A, D, $A - BD^{-1}C$, and $D - CA^{-1}B$ are nonsingular, then we have two equivalent descriptions of the inverse of M, that is,

$$M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} \tag{B.2}$$

and

$$M^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}. \tag{B.3}$$

The matrices $A - BD^{-1}C$ *and* $D - CA^{-1}B$ *are the* Schur complements *of* A *and* D*, respectively.*

Proof. These identities can be verified by inspection. ∎

Lemma B.3 (Sherman–Morrison–Woodbury). *Let* A *and* D *be invertible matrices and* B *and* D *be given so that the sum* $A + BD^{-1}C$ *is nonsingular. If* $D - CA^{-1}B$ *is also nonsingular, then*

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}.$$

Proof. This follows by equating the upper-left blocks of (B.2) and (B.3). ∎

Lemma B.4. *Assume that* $A, C > 0$ *and, for notational convenience, denote*

$$S = BAB^T + C^{-1} \quad \textit{and} \quad G = AB^TS^{-1}.$$

Then

$$(A^{-1} + B^TCB)^{-1} = (I - GB)A \tag{B.4}$$
$$= (I - GB)A(I - B^TG^T) + GC^{-1}G^T \tag{B.5}$$

and

$$G = (A^{-1} + B^TCB)^{-1}B^TC. \tag{B.6}$$

Proof. We obtain (B.4) as follows:

$$\begin{aligned}(A^{-1} + B^TCB)^{-1} &= A - AB^T(C^{-1} + BAB^T)^{-1}BA \\ &= A - AB^TS^{-1}BA \\ &= (I - GB)A.\end{aligned}$$

Similarly we have

$$\begin{aligned}(A^{-1} + B^TCB)^{-1} &= A - AB^TS^{-1}BA \\ &= A - 2AB^TS^{-1}BA + AB^TS^{-1}SS^{-1}BA \\ &= A - AB^TS^{-1}BA - AB^TS^{-1}BA \\ &\quad + AB^TS^{-1}BAB^TS^{-1}BA + AB^TS^{-1}C^{-1}S^{-1}BA \\ &= A - GBA - AB^TG^T + GBAB^TG^T + GC^{-1}G^T \\ &= (I - GB)A(I - B^TG^T) + GC^{-1}G^T.\end{aligned}$$

Finally,

$$\begin{aligned}G &= AB^TS^{-1} \\ &= AB^T(BAB^T + C^{-1})^{-1} \\ &= (A^{-1} + B^TCB)^{-1}(A^{-1} + B^TCB)AB^T(BAB^T + C^{-1})^{-1} \\ &= (A^{-1} + B^TCB)^{-1}B^TC(C^{-1} + BAB^T)(BAB^T + C^{-1})^{-1} \\ &= (A^{-1} + B^TCB)^{-1}B^TC. \qquad ∎\end{aligned}$$

Appendix C. Classical Kalman Filter Derivation. In this section, we present a classical derivation of the Kalman filter, which follows the derivation given by Kalman

[22]. While many of the ideas presented by Kalman work in more generality, his derivation assumes that the noise is Gaussian in order to produce explicit equations for the filter. Therefore, we now assume w_k and v_k are independent and normally distributed with covariance $Q_k, R_k > 0$. We also assume that x_0 is normally distributed with mean μ_0 and covariance Q_0 and that w_k and v_k are independent of x_j and y_j for $j \leq k$.

For the normally distributed random variable x_0 it is clear that the optimal estimate is the mean $\hat{x}_0 = \mu_0$. In fact, the assumption that x_0, w_k, and v_k are normally distributed is enough to guarantee that x_k is normally distributed for all k, a fact that we prove in this section. Therefore, to obtain the optimal estimate $\hat{x}_{k|m}$ it is sufficient to determine the mean of the random variable x_k, so we will use $\hat{x}_{k|m}$ to denote both the optimal estimate and the mean of x_k given the observations $y_1, \ldots, y_m$.

LEMMA C.1. *If the distribution of x_{k-1}, given the first $k-1$ observations, is normal with mean $\hat{x}_{k-1}$ and covariance P_{k-1}, then the random vector $\begin{bmatrix} x_k & y_k \end{bmatrix}^T$, without additional observations, is normally distributed with mean*

$$\begin{bmatrix} \hat{x}_{k|k-1} \\ \hat{y}_{k|k-1} \end{bmatrix} = \begin{bmatrix} F_k \hat{x}_{k-1} + G_k u_k \\ H_k \hat{x}_{k|k-1} \end{bmatrix}$$

and covariance

$$\operatorname{Cov} \begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} F_k P_{k-1} F_k^T + Q_k & P_{k|k-1} H_k^T \\ H_k P_{k|k-1} & H_k P_{k|k-1} H_k^T + R_k \end{bmatrix}.$$

Proof. Assuming $k-1$ observations, the mean is

$$\mathbb{E} \begin{bmatrix} x_k \\ y_k \end{bmatrix} = \mathbb{E} \begin{bmatrix} F_k x_{k-1} + G_k u_k + w_k \\ H_k x_k + v_k \end{bmatrix} = \begin{bmatrix} F_k \hat{x}_{k-1} + G_k u_k \\ H_k \hat{x}_{k|k-1} \end{bmatrix}.$$

For the covariance, we employ the notation $s(\ldots)^T = ss^T$ where convenient. The covariance matrix is

$$\operatorname{Cov} \begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} \mathbb{E}[(x_k - \hat{x}_{k|k-1})(\ldots)^T] & \mathbb{E}[(x_k - \hat{x}_{k|k-1})(y_k - \hat{y}_{k|k-1})^T] \\ \mathbb{E}[(y_k - \hat{y}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T] & \mathbb{E}[(y_k - \hat{y}_{k|k-1})(\ldots)^T] \end{bmatrix}.$$

Observing that $\mathbb{E}[(x_k - \hat{x}_{k|k-1})(y_k - \hat{y}_{k|k-1})^T] = \mathbb{E}[(y_k - \hat{y}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T]^T$, we compute

$$\begin{aligned} \mathbb{E}[(x_k - \hat{x}_{k|k-1})(\ldots)^T] &= \mathbb{E}[(F_k(x_{k-1} - \hat{x}_{k-1}) + w_k)(\ldots)^T] \\ &= F_k \mathbb{E}[(x_{k-1} - \hat{x}_{k-1})(\ldots)^T] F_k^T + \mathbb{E}[w_k w_k^T] \\ &= F_k P_{k-1} F_k^T + Q_k. \end{aligned}$$

The cross-terms of x_{k-1} and $\hat{x}_{k-1}$ with w_k are zero since x_{k-1} is independent of w_k and w_k has zero mean. Similarly,

$$\mathbb{E}[(y_k - \hat{y}_{k|k-1})(\ldots)^T] = \mathbb{E}[(H_k(x_k - \hat{x}_{k|k-1}) + v_k)(\ldots)^T] = H_k P_{k|k-1} H_k^T + R_k$$

and

$$\begin{aligned} \mathbb{E}[(x_k - \hat{x}_{k|k-1})(y_k - \hat{y}_{k|k-1})^T] &= \mathbb{E}[(x_k - \hat{x}_{k|k-1})(H_k(x_k - \hat{x}_{k|k-1}) + v_k)^T] \\ &= P_{k|k-1} H_k^T. \qquad \square \end{aligned}$$

The previous lemma establishes the prediction step for the Kalman filter. In order to determine the update from an observation, we require the following general result about joint normal distributions.

THEOREM C.2. *Let x and y be random n- and m-vectors, respectively, and let the $(n+m)$-vector $z = \begin{bmatrix} x & y \end{bmatrix}^T$ be normally distributed with mean and positive-definite covariance*

$$\hat{z} = \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} \quad \text{and} \quad P_z = \begin{bmatrix} P_x & P_{xy} \\ P_{yx} & P_y \end{bmatrix}.$$

Then the distribution of x given y is normal with mean $\hat{x} + P_{xy}P_y^{-1}(y - \hat{y})$ and covariance $P_x - P_{xy}P_y^{-1}P_{yx}$.

Proof. The random vector z has density function

$$f(x,y) = f(z) = \frac{1}{(2\pi)^{(n+m)/2}(\det P_z)^{1/2}} \exp\left[-\frac{1}{2}(z-\hat{z})^T P_z^{-1}(z-\hat{z})\right],$$

with a similar expression for the density $g(y)$ of y. The density of the conditional random vector $(x|y)$ is

$$\frac{f(x,y)}{g(y)} = \frac{(\det P_y)^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}(\det P_z)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left[(z-\hat{z})^T P_z^{-1}(z-\hat{z}) - (y-\hat{y})^T P_y^{-1}(y-\hat{y})\right]\right].$$

By Lemma B.2,

$$P_z^{-1} = \begin{bmatrix} (P_x - P_{xy}P_y^{-1}P_{yx})^{-1} & -(P_x - P_{xy}P_y^{-1}P_{yx})^{-1}P_{xy}P_y^{-1} \\ -P_y^{-1}P_{yx}(P_x - P_{xy}P_y^{-1}P_{yx})^{-1} & (P_y - P_{yx}^T P_x^{-1}P_{xy})^{-1} \end{bmatrix}.$$

Therefore,

$$\begin{aligned}&(z-\hat{z})^T P_z^{-1}(z-\hat{z}) - (y-\hat{y})^T P_y^{-1}(y-\hat{y}) \\ &\quad = [x - \hat{x} + P_{xy}P_y^{-1}(y-\hat{y})]^T (P_x - P_{xy}P_y^{-1}P_{yx})^{-1}[x - \hat{x} + P_{xy}P_y^{-1}(y-\hat{y})].\end{aligned}$$

Furthermore, P_z satisfies

$$\begin{bmatrix} I & -P_{xy}P_y^{-1} \\ 0 & I \end{bmatrix} P_z = \begin{bmatrix} P_x - P_{xy}P_y^{-1}P_{yx} & 0 \\ P_{yx} & P_y \end{bmatrix},$$

so it follows that

$$\det P_z = \det(P_x - P_{xy}P_y^{-1}P_{yx}) \det P_y.$$

Inserting these simplifications into the density function for $h(x|y) = f(x,y)/g(y)$, we see that it is the density of a normal distribution with mean $\hat{x} + P_{xy}P_y^{-1}(y-\hat{y})$ and covariance $P_x - P_{xy}P_y^{-1}P_{yx}$, which establishes the result. ∎

Returning to the context of the filtering problem, we immediately obtain the following update equations for $\hat{x}_k$ given y_k from the previous theorem.

COROLLARY C.3. *If $\begin{bmatrix} x_k & y_k \end{bmatrix}^T$ is distributed as in Lemma C.1, then the distribution of x_k, given the observation y_k, is normal with mean*

$$\hat{x}_k = \hat{x}_{k|k-1} + P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1}(y_k - H_k\hat{x}_{k|k-1})$$

and covariance

$$P_k = P_{k|k-1} - P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1}H_kP_{k|k-1}.$$

In summary, from Lemma C.1 we obtain the standard a priori estimates (4.4a) and (4.4b) and from Corollary C.3 we use the observation y_k to obtain the standard a posteriori estimates (3.11a) and (3.11b) for the two-step Kalman filter.

Having obtained the two-step Kalman filter using the above statistical approach, it is a straightforward algebraic problem to produce equations for a one-step filter by manipulating the covariance matrix using the lemmas from Appendix B.

Acknowledgments. The authors express their thanks to Randal W. Beard, Magnus Egerstedt, Gregory Lyng, Bryce B. Ready, C. Shane Reese, Gilbert Strang, H. Dennis Tolley, and Divakar Viswanath for useful suggestions and conversations. They also thank the referees for their excellent comments and suggestions.

REFERENCES

[1] B. M. Bell, *The iterated Kalman smoother as a Gauss–Newton method*, SIAM J. Optim., 4 (1994), pp. 626–636.

[2] B. M. Bell, J. V. Burke, and G. Pillonetto, *An inequality constrained nonlinear Kalman-Bucy smoother by interior point likelihood maximization*, Automatica J. IFAC, 45 (2009), pp. 25–33.

[3] B. M. Bell and F. W. Cathey, *The iterated Kalman filter update as a Gauss-Newton method*, IEEE Trans. Automat. Control, 38 (1993), pp. 294–297.

[4] D. P. Bertsekas, *Incremental least squares methods and the extended Kalman filter*, SIAM J. Optim., 6 (1996), pp. 807–822.

[5] E. J. Bomhoff, *Four Econometric Fashions and the Kalman Filter Alternative*, Working Papers 324, Rochester Center for Economic Research (RCER), University of Rochester, 1992; http://ideas.repec.org/p/roc/rocher/324.html.

[6] G. E. P. Box and N. R. Draper, *Empirical Model-Building and Response Surfaces*, Wiley Ser. Probab. Math. Statist. Appl. Probab. Statist., John Wiley, New York, 1987.

[7] R. G. Brown and P. H. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley, New York, 1997.

[8] J. Chandrasekar, D. S. Bernstein, O. Barrero, and B. L. R. De Moor, *Kalman filtering with constrained output injection*, Internat. J. Control, 80 (2007), pp. 1863–1879.

[9] A. Doucet and A. M. Johansen, *A tutorial on particle filtering and smoothing: Fifteen years later*, in Handbook of Nonlinear Filtering, D. Crisan and B. Rozovsky, eds., Oxford University Press, 2011, pp. 656–704.

[10] S. Ertřk, *Real-time digital image stabilization using Kalman filters*, Real-Time Imaging, 8 (2002), pp. 317–328.

[11] S. Gannot, D. Burshtein, and E. Weinstein, *Iterative and sequential Kalman filter-based speech enhancement algorithms*, IEEE Trans. Speech and Audio Processing, 6 (1998), pp. 373–385.

[12] A. Gelb, *Applied Optimal Estimation*, MIT Press, Cambridge, MA, 1974.

[13] F. A. Graybill, *Theory and Application of the Linear Model*, Duxbury Press, North Scituate, MA, 1976.

[14] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB*, 2nd ed., John Wiley, New York, 2001.

[15] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, UK, 1991.

[16] S. S. Haykin, *Kalman Filtering and Neural Networks*, Wiley-Interscience, New York, 2001.

[17] H. V. Henderson and S. R. Searle, *On deriving the inverse of a sum of matrices*, SIAM Rev., 23 (1981), pp. 53–60.

[18] J. Humpherys and J. West, *Kalman filtering with Newton's method*, IEEE Control Syst. Mag., 30 (2010), pp. 101–106.

[19] R. Isermann, *Process fault detection based on modeling and estimation methods—a survey*, Automatica, 20 (1984), pp. 387–404.

[20] S. J. Julier and J. K. Uhlmann, *A new extension of the Kalman filter to nonlinear systems*, in International Symposium on Aerospace/Defense Sensing, Simulation and Controls, Vol. 3, 1997, p. 26.

[21] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*, Prentice-Hall, Englewood Cliffs, NJ, 2000.

[22] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Trans. ASME J. Basic Engrg., 82 (1960), pp. 35–45.
[23] R. E. Kalman, *Mathematical description of linear dynamical systems*, J. SIAM Control Ser. A, 1 (1963), pp. 152–192.
[24] R. E. Kalman and R. S. Bucy, *New results in linear filtering and prediction theory*, Trans. ASME J. Basic Engrg., 83 (1961), pp. 95–108.
[25] E. Kaplan and C. Hegarty, *Understanding GPS: Principles and Applications*, 2nd ed., Artech House, 2006.
[26] C. T. Kelley, *Solving Nonlinear Equations with Newton's Method*, SIAM, Philadelphia, 2003.
[27] U. Kiencke and L. Nielsen, *Automotive control systems: For engine, driveline, and vehicle*, Measurement Sci. Technol., 11 (2000), p. 1828.
[28] S. L. Lauritzen, *Time series analysis in* 1880: *A discussion of contributions made by T. N. Thiele*, Internat. Statist. Rev., 49 (1981), pp. 319–331.
[29] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, *Moving target classification and tracking from real-time video*, in IEEE Workshop on Applications of Computer Vision, Vol. 14, 1998, p. 3.
[30] L. Ljung, *System Identification: Theory for the User*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1998.
[31] M. Manoliu and S. Tompaidis, *Energy futures prices: Term structure models with Kalman filter estimation*, Appl. Math. Finance, 9 (2002), pp. 21–43.
[32] L. A. McGee and S. F. Schmidt, *Discovery of the Kalman Filter as a Practical Tool for Aerospace and Industry*, Technical Memorandum 86847, National Aeronautics and Space Administration, 1985.
[33] J. B. Pearson and E. B. Stear, *Kalman filter applications in airborne radar tracking*, IEEE Trans. Aerospace and Electronic Systems, 10 (1974), pp. 319–329.
[34] C. Ridder, O. Munkelt, and H. Kirchner, *Adaptive background estimation and foreground detection using Kalman-filtering*, in Proceedings of International Conference on Recent Advances in Mechatronics, Istanbul, Turkey, 1995, pp. 193–199.
[35] S. I. Roumeliotis and G. A. Bekey, *Bayesian estimation and Kalman filtering: A unified framework for mobile robot localization*, in IEEE International Conference on Robotics and Automation, Vol. 3, 2000, pp. 2985–2992.
[36] L. L. Scharf and C. Demeure, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, Reading, MA, 1991.
[37] S. F. Schmidt, *Application of state-space methods to navigation problems*, Adv. Control Systems, 3 (1966), pp. 293–340.
[38] D. Simon, *Optimal State Estimation: Kalman, H–Infinity, and Nonlinear Approaches*, Wiley-Interscience, New York, 2006.
[39] G. Strang and K. Borre, *Linear Algebra, Geodesy, and GPS*, Wellesley–Cambridge Press, Wellesley, MA, 1997.
[40] R. L. Stratonovich, *Optimum nonlinear systems which bring about a separation of a signal with constant parameters from noise*, Radiofizika, 2 (1959), pp. 892–901.
[41] R. L. Stratonovich, *Application of the theory of Markov processes for optimum filtration of signals*, Radio Eng. Electron. Phys, 1 (1960), pp. 1–19.
[42] P. Swerling, *Note on the Minimum Variance of Unbiased Estimates of Doppler Shift*, Tech. report, RAND, 1958.
[43] P. Swerling, *Parameter estimation for waveforms in additive Gaussian noise*, J. Soc. Indust. Appl. Math., 7 (1959), pp. 152–166.
[44] T. N. Thiele, *Om anvendelse af mindste Kvadraters Methode i nogle tilfaelde, hvor en Komplikation af visse slags uensartede tilfaeldige Fejkilder giver Fejlene en "systematisk" Karakter*, Luno, 1880.
[45] T. N. Thiele, *Sur la compensation de quelques erreurs quasi-systématiques par la méthode des moindres carrés*, CA Reitzel, 1880.
[46] E. A. Wan and R. Van Der Merwe, *The unscented Kalman filter for nonlinear estimation*, in the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC 2000), 2002, pp. 153–158.
[47] M. A. Woodbury, *The Stability of Out-Input Matrices*, Chicago, IL, 1949.
[48] M. A. Woodbury, *Inverting Modified Matrices*, Statistical Research Group, Memo. Report 42, Princeton University, Princeton, NJ, 1950.
[49] P. Zarchan, *Tactical and Strategic Missile Guidance*, 5th ed., American Institute of Aeronautics and Astronautics, 2007.