Kalman Filtering with Newton's Method

JEFFREY HUMPHERYS and JEREMY WEST

The Kalman filter is arguably one of the most notable algorithms of the 20th century [1]. In this article, we derive the Kalman filter using Newton's method for root finding. We show that the one-step Kalman filter is given by a single iteration of Newton's method on the gradient of a quadratic objective function, and with a judiciously chosen initial guess. This derivation is different from those found in standard texts [2]–[6], since it provides a more general framework for recursive state estimation. Although not presented here, this approach can also be used to derive the extended Kalman filter for nonlinear systems.

BACKGROUND

Linear Estimation

Suppose that data are generated by the linear model

$$b = Ax + \varepsilon, \tag{1}$$

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 $Q = \mathbb{E}[\varepsilon\varepsilon^T]$, denoted $Q > 0$, and $b \in \mathbb{R}^m$ represents known, but inexact, measurements with errors given by ε. The vector $x \in \mathbb{R}^n$ is the set of parameters to be estimated.

The estimator $\hat{x}$ is *linear* if $\hat{x} = Kb$ for some $n \times m$ matrix K. We say that the estimator $\hat{x}$ is unbiased if $\mathbb{E}[\hat{x}] = x$. Since

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

it follows that the linear estimator $\hat{x} = Kb$ is unbiased if and only if $KA = I$. When $m > n$, many choices of K could be made to satisfy this condition. We want to find the mean-squared error-minimizing linear unbiased estimator, or, in other words, the minimizer of the quantity $\mathbb{E}[\|\hat{x} - x\|^2]$ over all matrices K that satisfy the constraint $KA = I$. In "What Is the Best Linear Unbiased Estimator," we prove the Gauss-Markov theorem, which states that the optimal K is given by $K = (A^TQ^{-1}A)^{-1}A^TQ^{-1}$, and thus the corresponding estimator is given by

$$\hat{x} = (A^TQ^{-1}A)^{-1}A^TQ^{-1}b, \tag{2}$$

whose covariance is given by

$$P = \mathbb{E}[(\hat{x} - x)(\hat{x} - x)^T] = (A^TQ^{-1}A)^{-1}. \tag{3}$$

In addition, the Gauss-Markov theorem states that every other linear unbiased estimator of x has larger covariance than (3). Thus we call (2) the best *linear unbiased estimator* of (1).

Weighted Least Squares

It is not a coincidence that the estimator in (2) has the same form as the solution of the weighted least squares problem

$$\hat{x} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \frac{1}{2}\|b - Ax\|_W^2, \tag{4}$$

where $W > 0$ is a problem-specific weighting matrix and $\|\cdot\|_W$ describes a weighted 2-norm defined as $\|x\|_W^2 := x^TWx$. The factor $1/2$ is not necessary but makes calculations involving the derivative simpler. Scaling the objective function by a constant does not affect the minimizer. Therefore, consider the positive-definite quadratic objective function

$$\begin{aligned}J(x) &= \frac{1}{2}\|b - Ax\|_W^2 \\ &= \frac{1}{2}x^TA^TWAx - x^TA^TWb + \frac{1}{2}b^TWb.\end{aligned} \tag{5}$$

Its minimizer $\hat{x}$ satisfies $\nabla J(\hat{x}) = 0$, which yields the *normal equations*

$$A^TWA\hat{x} = A^TWb. \tag{6}$$

Since A is assumed to have full column rank, the weighted Gramian A^TWA is nonsingular, and thus (6) has the solution

$$\hat{x} = (A^TWA)^{-1}A^TWb. \tag{7}$$

Hence, in comparing (2) with (7), we see that the best linear unbiased estimator of (1) is found by solving the weighted least squares problem (4) with weight

Digital Object Identifier 10.1109/MCS.2010.938485

1066-033X/10/$26.00©2010IEEE

What Is the Best Linear Unbiased Estimator?

Suppose that data are generated by the linear model

$$b = Ax + \varepsilon,$$

where A is a known $m \times n$ matrix of rank n, and ε is an m-dimensional random variable with zero mean and covariance $Q > 0$.

Recall that a linear estimator $\hat{x} = Kb$ of x is unbiased if and only if $KA = I$. If $n < m$, then there are many choices of K that are possible. We want the matrix K that minimizes the mean-squared error $\mathbb{E}[\|\hat{x} - x\|^2]$. Note, however, that

$$\begin{aligned}\|\hat{x} - x\|^2 &= \|Kb - x\|^2\\ &= \|KAx + K\varepsilon - x\|^2\\ &= \|K\varepsilon\|^2\\ &= \varepsilon^T K^T K \varepsilon.\end{aligned}$$

Moreover, since $\varepsilon^T K^T K \varepsilon$ is a scalar, we have

$$\varepsilon^T K^T K \varepsilon = \operatorname{tr}(\varepsilon^T K^T K \varepsilon) = \operatorname{tr}(K \varepsilon \varepsilon^T K^T).$$

It follows, by the linearity of expectation, that

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

Thus the mean-squared error minimizing linear unbiased estimator is found by minimizing $\operatorname{tr}(KQK^T)$ subject to the constraint $KA = I$, where $K \in \mathbb{R}^{n \times m}$. This is a convex optimization problem, that is, the objective is a convex function and the feasible set is a convex set. Hence, to find the unique minimizer, it suffices to find the unique critical point of the Lagrangian

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

where the matrix Lagrange multiplier $\lambda \in \mathbb{R}^{n \times n}$ corresponds to the n^2 constraints $KA = I$. Thus the minimum occurs when

$$\begin{aligned}0 &= \nabla_K \mathcal{L}(K, \lambda)\\ &= \frac{\partial}{\partial K} \operatorname{tr}(KQK^T - \lambda^T (KA - I))\\ &= KQ^T + KQ - \lambda A^T\\ &= 2KQ - \lambda A^T,\end{aligned}$$

that is, when

$$K = \frac{1}{2} \lambda A^T Q^{-1}.$$

Since $KA = I$, we have that $\lambda = 2(A^T Q^{-1} A)^{-1}$. Therefore, $K = (A^T Q^{-1} A)^{-1} A^T Q^{-1}$, and the optimal estimator is given by

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

To compute the covariance of the estimator we expand $\hat{x}$ as

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

Thus the covariance is given by

$$\begin{aligned}P &= \mathbb{E}[(\hat{x} - x)(\hat{x} - x)^T]\\ &= (A^T Q^{-1} A)^{-1} A^T Q^{-1} \mathbb{E}[\varepsilon \varepsilon^T] Q^{-1} A (A^T Q^{-1} A)^{-1}\\ &= (A^T Q^{-1} A)^{-1}.\end{aligned}$$

We now show that every linear unbiased estimator other than $\hat{x}$ produces a larger covariance. If $\hat{x}_L = Lb$ is a linear unbiased estimator of the linear model, then since $KA = I$, it follows that there exists a matrix D such that $DA = 0$ and $L = K + D$. The covariance of $\hat{x}_L$ is given by

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

Note, however, that $KQD^T = 0$. Thus

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

Therefore, $\hat{x}$ is the best linear unbiased estimator of x.

$W = Q^{-1} > 0$. Also from (3) it can be noted that the inverse weighted-Gramian matrix $P = (A^T Q^{-1} A)^{-1}$ is the covariance of this estimator.

Newton's Method

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a smooth function. Assume that $f(\hat{x}) = 0$ and that the Jacobian matrix $Df(x)$ is nonsingular in a neighborhood of $\hat{x}$. If the initial guess x_0 is sufficiently close to $\hat{x}$, then Newton's method, which is given by

$$x_{k+1} = x_k - Df(x_k)^{-1} f(x_k), \tag{8}$$

produces a sequence $\{x_k\}_{k=0}^{\infty}$ that converges quadratically to $\hat{x}$, that is, there exists a constant $c > 0$ such that, for every positive integer k,

$$\|x_{k+1} - \hat{x}\| \leq c \|x_k - \hat{x}\|^2.$$

Newton's method is widely used in optimization problems since the local extrema of an objective function $J(x)$ can be obtained by finding roots of its gradient $f(x) := \nabla J(x)$. Hence, from (8) we have

$$x_{k+1} = x_k - D^2 J(x_k)^{-1} \nabla J(x_k), \tag{9}$$

where $D^2 J(x_k)$ denotes the Hessian of J evaluated at x_k. This iterative process converges, likewise at a quadratic rate, to the isolated local minimizer $\hat{x}$ of J whenever the starting point x_0 is sufficiently close to $\hat{x}$ and the Hessian $D^2 J(x)$ is positive definite in a neighborhood of $\hat{x}$. In practice, we do not invert the Hessian $D^2 J(x_k)$ when computing (9). Indeed, by a factor of roughly two, we can

In this article, we derive the Kalman filter using Newton's method for root finding.

more efficiently compute the Newton update by solving the linear system $D^2J(x_k)y_k = -\nabla J(x_k)$, for example by Gaussian elimination, and then setting $x_{k+1} = x_k + y_k$; see [7] for details.

To find the minimizer of a positive-definite quadratic form (5), it is not necessary to use an iterative scheme, such as Newton's method, since the gradient is affine in x and the unique minimizer can be found by solving the normal equations (6) directly. However, positive-definite quadratic forms have the special property that Newton's method (9) converges in a single step for each starting point $x \in \mathbb{R}^n$, that is, for all $x \in \mathbb{R}^n$, the minimizer $\hat{x}$ satisfies

$$\hat{x} = x - D^2J(x)^{-1}\nabla J(x). \tag{10}$$

This observation provides a key insight used to derive the Kalman filter below. Note also that the Hessian of the quadratic form (5) is $D^2J(x) = A^TWA$, which is constant in x. Thus, throughout we denote the Hessian as D^2J, dropping the explicit dependence on x.

Recursive Least Squares

We now consider the least squares problem (4) in the case of recursive implementation. At each time k, we seek the best linear unbiased estimator $\hat{x}_k$ of x for the linear model

$$\beta_k = \mathcal{A}_k x + \varepsilon_k, \tag{11}$$

where

$$\beta_k = \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}, \quad \mathcal{A}_k = \begin{bmatrix} A_1 \\ \vdots \\ A_k \end{bmatrix}, \quad \text{and} \quad \varepsilon_k = \begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix}.$$

We assume, for each $j = 1, \ldots, k$, that the noise terms v_j have zero mean and are uncorrelated with $\mathbb{E}[v_j v_j^T] = R_j > 0$. We also assume that A_1 has full column rank, and thus each $\mathcal{A}_k$ has full column rank. The estimate is found by solving the weighted least squares problem

$$\hat{x}_k = \operatorname*{argmin}_x \|\beta - \mathcal{A}_k x\|^2_{\mathcal{W}_k}, \tag{12}$$

where the weight is given by the inverse covariance $\mathcal{W}_k = \mathbb{E}[\varepsilon_k \varepsilon_k^T]^{-1} = \operatorname{diag}(R_1^{-1}, \ldots, R_k^{-1})$.

As time marches forward, the number of rows of the linear system increases, thus altering the least squares solution. We now show that it is possible to use the least squares solution $\hat{x}_{k-1}$ at time $k-1$ to efficiently compute the least squares solution $\hat{x}_k$ at time k. This process is the *recursive least squares* algorithm.

We begin by rewriting (12) as

$$J_k(x) = \frac{1}{2}\sum_{i=1}^{k} \|b_i - A_i x\|^2_{R_i^{-1}}. \tag{13}$$

The positive-definite quadratic form (13) can be expressed recursively as

$$J_k(x) = J_{k-1}(x) + \frac{1}{2}\|b_k - A_k x\|^2_{R_k^{-1}}.$$

The gradient and Hessian of J_k are given by

$$\nabla J_k(x) = \nabla J_{k-1}(x) + A_k^T R_k^{-1}(A_k x - b_k)$$

and

$$D^2J_k = D^2J_{k-1} + A_k^T R_k^{-1} A_k, \tag{14}$$

respectively. Since A_1 has full column rank, we see that $D^2J_1 > 0$. From (14), it follows that $D^2J_k > 0$ for every positive integer k, and hence from (10) the minimizer of (13) becomes

$$\hat{x}_k = x - (D^2J_k)^{-1}(\nabla J_{k-1}(x) + A_k^T R_k^{-1}(A_k x - b_k)), \tag{15}$$

where the starting point $x \in \mathbb{R}^n$ can be arbitrarily chosen. Since $\nabla J_{k-1}(\hat{x}_{k-1}) = 0$, we set $x = \hat{x}_{k-1}$ in (15). Thus (15) becomes

$$\hat{x}_k = \hat{x}_{k-1} - K_k A_k^T R_k^{-1}(A_k \hat{x}_{k-1} - b_k),$$

where $K_k := (D^2J_k)^{-1}$ is the inverse of the Hessian of J, and from (3) represents the covariance of the estimate. Observing from (14) that $K_k^{-1} = K_{k-1}^{-1} + A_k^T R_k^{-1} A_k$ and using Lemma 1 in "Inversion Lemmata," we have

$$\begin{aligned} K_k &= (K_{k-1}^{-1} + A_k^T R_k^{-1} A_k)^{-1} \\ &= K_{k-1} - K_{k-1} A_k^T (R_k + A_k K_{k-1} A_k^T)^{-1} A_k K_{k-1}. \end{aligned}$$

Thus, to summarize, the recursive least squares method is

$$\begin{aligned} K_k &= K_{k-1} - K_{k-1} A_k^T (R_k + A_k K_{k-1} A_k^T)^{-1} A_k K_{k-1} \\ \hat{x}_k &= \hat{x}_{k-1} - K_k A_k^T R_k^{-1}(A_k \hat{x}_{k-1} - b_k). \end{aligned}$$

At each time k, this algorithm gives the best linear unbiased estimator of (11), along with its covariance.

The Kalman filter is arguably one of the most notable algorithms of the 20th century.

The recursive least squares algorithm allows for efficient updating of the weighted least squares solution. Indeed, if p rows are added during the kth step, then the expression $R_k + A_k K_{k-1} A_k^T$ is a $p \times p$ matrix, which is small in size compared to the matrix $\mathcal{A}_k \mathcal{W}_k \mathcal{A}_k^T$, which is needed, say, if we compute the solution directly using (7).

STATE-SPACE MODELS AND STATE ESTIMATION

Consider the stochastic discrete-time linear dynamic system

$$x_{k+1} = F_k x_k + G_k u_k + w_k, \tag{16}$$

$$y_k = H_k x_k + v_k, \tag{17}$$

where the variables represent the state $x_k \in \mathbb{R}^n$, the input $u_k \in \mathbb{R}^p$, and the output $y_k \in \mathbb{R}^q$. The terms w_k and v_k are noise processes, which are assumed to have mean zero and be mutually uncorrelated, with known covariances $Q_k > 0$ and $R_k > 0$, respectively. We assume that the system has the stochastic initial state $x_0 = \mu_0 + w_{-1}$, where μ_0 is the mean and w_{-1} is a zero-mean random variable with covariance $\mathbb{E}[w_{-1} w_{-1}^T] = Q_{-1} > 0$.

State Estimation

We formulate the state estimation problem, given m known observations $y_1, \ldots, y_m$ and k known inputs $u_0, \ldots, u_{k-1}$, where $k \geq m$. To find the best linear unbiased estimator of the states $x_1, \ldots, x_k$, we begin by writing (16)–(17) as the large linear estimation problem

$$\begin{array}{ccccc}
\mu_0 & = & x_0 & - & w_{-1}, \\
G_0 u_0 & = & x_1 - F_0 x_0 & - & w_0, \\
y_1 & = & H_1 x_1 & + & v_1, \\
 & \vdots & & \vdots & \\
G_{m-1} u_{m-1} & = & x_m - F_{m-1} x_{m-1} & - & w_{m-1}, \\
y_m & = & H_m x_m & + & v_m, \\
G_m u_m & = & x_{m+1} - F_m x_m & - & w_m, \\
 & \vdots & & \vdots & \\
G_{k-1} u_{k-1} & = & x_k - F_{k-1} x_{k-1} & - & w_{k-1}.
\end{array}$$

Note that the known measurements, namely, the inputs and outputs, are on the left, whereas the unknown states are on the right, together with the noise terms. The parameters to be estimated in this linear estimation problem are the states $x_0, x_1, \ldots, x_k$, which we write as

$$z_k = [x_0 \ \cdots \ x_k]^T \in \mathbb{R}^{(k+1)n}.$$

Then the linear system takes the form

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

where

$$\varepsilon_{k|m} = [w_{-1}^T \quad w_0^T \quad v_1^T \quad \cdots \quad w_{m-1}^T \quad v_m^T \quad w_m^T \quad \cdots \quad w_{k-1}^T]^T$$

is a zero-mean random variable whose inverse covariance is the positive-definite block-diagonal matrix

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

Inversion Lemmata

We provide some technical lemmata that are used throughout the article. The first lemma is the Sherman–Morrison–Woodbury formula, which shows how to invert an additively updated matrix when we know the inverse of the original matrix. The second inversion lemma tells us how to invert block matrices. These statements can be verified by direct computation; see also [8, p. 18–19].

LEMMA 1 (SHERMAN-MORRISON-WOODBURY)

Let A and C be square matrices and B and D be given so that the sum $A + BCD$ is nonsingular. If A, C and $C^{-1} + DA^{-1}B$ are also nonsingular, then

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}.$$

LEMMA 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

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

The second observation is that the least-squares estimation problem can be solved by minimizing a positive-definite quadratic functional.

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

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

Since the state estimation problem is cast as a linear model, we can equivalently solve the estimation problem by minimizing the positive-definite objective function

$$J_{k|m}(z_k) = \frac{1}{2}\|A_{k|m} z_k - b_{k|m}\|_{W_{k|m}}^2. \tag{18}$$

We can find its minimizer $\hat{z}_{k|m}$ by performing a single iteration of Newton's method given by

$$\hat{z}_{k|m} = z - (A_{k|m}^T W_{k|m} A_{k|m})^{-1} A_{k|m}^T W_{k|m} (A_{k|m} z - b_{k|m}), \tag{19}$$

where $z \in \mathbb{R}^{(k+1)n}$ is arbitrary. In the following, we set $m = k$ and choose a canonical z that simplifies (19), thus leaving us with the Kalman filter.

Kalman Derivation with Newton's Method

Consider the state estimation problem in the case $m = k$, that is, the number of observations equals the number of inputs. It is of particular interest in applications to determine the current state x_k given the observations $y_1, \ldots, y_k$ and inputs $u_0, \ldots, u_{k-1}$. The Kalman filter gives the best linear unbiased estimator $\hat{x}_{k|k}$ of x_k in terms of the previous state estimator $\hat{x}_{k-1|k-1}$ and the latest data u_{k-1} and y_k up to that point in time.

We begin by rewriting (18) as

$$\begin{aligned} J_{k|m}(z_k) = {} & \frac{1}{2}\|x_0 - \mu_0\|_{Q_1^{-1}}^2 + \frac{1}{2}\sum_{i=1}^{m} \|y_i - H_i x_i\|_{R_i^{-1}}^2 \\ & + \frac{1}{2}\sum_{i=1}^{k} \|x_i - F_{i-1} x_{i-1} - G_{i-1} u_{i-1}\|_{Q_{i-1}^{-1}}^2. \end{aligned} \tag{20}$$

For convenience, 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. For $m = k$, (20) can be expressed recursively as

$$\begin{aligned} J_k(z_k) = {} & J_{k-1}(z_{k-1}) + \frac{1}{2}\|y_k - H_k x_k\|_{R_k^{-1}}^2 \\ & + \frac{1}{2}\|x_k - \mathcal{F}_{k-1} z_{k-1} - G_{k-1} u_{k-1}\|_{Q_{k-1}^{-1}}^2, \end{aligned}$$

where $\mathcal{F}_k = [0 \ \cdots \ 0 \ \ F_k]$. The gradient and Hessian of J_k are given by

$$\nabla J_k(z_k) = \begin{bmatrix} \nabla J_{k-1}(z_{k-1}) + \mathcal{F}_{k-1}^T Q_{k-1}^{-1}(F_{k-1} z_{k-1} - x_k + G_{k-1} u_{k-1}) \\ -Q_{k-1}^{-1}(F_{k-1} z_{k-1} - x_k + G_{k-1} u_{k-1}) + H_k^T R_k^{-1}(H_k x_k - y_k) \end{bmatrix}$$

and

$$D^2 J_k = \begin{bmatrix} D^2 J_{k-1} + \mathcal{F}_{k-1}^T Q_{k-1}^{-1} \mathcal{F}_{k-1} & -F_{k-1}^T Q_{k-1}^{-1} \\ -Q_{k-1}^{-1} F_{k-1} & Q_{k-1}^{-1} + H_k^T R_k^{-1} H_k \end{bmatrix}, \tag{21}$$

respectively. Since $D^2 J_0 = Q_1^{-1} > 0$, it follows inductively that $D^2 J_k > 0$ for every positive integer k. The proof follows from the observation that

$$\begin{aligned} z_k^T D^2 J_k z_k = {} & z_{k-1}^T D^2 J_{k-1} z_{k-1} \\ & + \|\mathcal{F}_{k-1} z_{k-1} - x_k\|_{Q_{k-1}^{-1}}^2 + \|H_k x_k\|_{R_k^{-1}}^2, \end{aligned}$$

and thus the right-hand side is nonnegative, being zero only if $z_{k-1} = 0$ and $x_k = 0$, or, equivalently, $z_k = 0$.

From one iteration of Newton's method, we have that

$$\hat{z}_k = z_k - (D^2 J_k)^{-1} \nabla J_k(z_k) \tag{22}$$

for all $z_k \in \mathbb{R}^{(k+1)n}$. Since $\nabla J_{k-1}(\hat{z}_{k-1}) = 0$ and $\mathcal{F}_{k-1}\hat{z}_{k-1} = F_{k-1}\hat{x}_{k-1}$, we set

$$z_k = \begin{bmatrix} \hat{z}_{k-1} \\ F_{k-1}\hat{x}_{k-1} + G_{k-1} u_{k-1}. \end{bmatrix}.$$

Thus,

$$\nabla J_k(z_k) = \begin{bmatrix} 0 \\ H_k^T R_k^{-1}[H_k(F_{k-1}\hat{x}_{k-1} + G_{k-1} u_{k-1}) - y_k] \end{bmatrix},$$

and the bottom row of (22) becomes

$$\begin{aligned} \hat{x}_k = {} & F_{k-1}\hat{x}_{k-1} + G_{k-1} u_{k-1} - P_k H_k^T R_k^{-1} \\ & \times [H_k(F_{k-1}\hat{x}_{k-1} + G_{k-1} u_{k-1}) - y_k], \end{aligned}$$

where P_k is the bottom-right block of the inverse Hessian $(D^2 J_k)^{-1}$, which from Lemma 2 in "Inversion Lemmata," is given by

$$\begin{aligned} P_k &= (Q_{k-1}^{-1} + H_k^T R_k^{-1} H_k - Q_{k-1}^{-1}\mathcal{F}_{k-1} \\ &\quad \times (D^2 J_{k-1} + \mathcal{F}_{k-1}^T Q_{k-1}^{-1} F_{k-1})^{-1} F_{k-1}^T Q_{k-1}^{-1})^{-1} \\ &= [(Q_{k-1} + \mathcal{F}_{k-1}(D^2 J_{k-1})^{-1} F_{k-1}^T)^{-1} + H_k^T R_k^{-1} H_k]^{-1} \\ &= [(Q_{k-1} + F_{k-1} P_{k-1} F_{k-1}^T)^{-1} + H_k^T R_k^{-1} H_k]^{-1}. \end{aligned}$$

Note that P_k is the covariance of the estimator $\hat{x}_k$.

To summarize, we have the recursive estimator for x_k given by

$$P_k = [(Q_{k-1} + F_{k-1}P_{k-1}F_{k-1}^T)^{-1} + H_k^T R_k^{-1} H_k]^{-1}$$
$$\hat{x}_k = F_{k-1}\hat{x}_{k-1} + G_{k-1}u_{k-1} - P_k H_k^T R_k^{-1} \times [H_k(F_{k-1}\hat{x}_{k-1} + G_{k-1}u_{k-1}) - y_k],$$

where $\hat{x}_0 = \mu_0$, $P_0 = Q_{-1}$. This recursive estimator is the *one-step Kalman filter*.

CONCLUSIONS

There are a few key observations in the Newton method derivation of the Kalman filter. First is that the state estimation problem for linear systems (16)–(17) is a potentially large linear least squares estimation problem. The second observation is that the least-squares estimation problem can be solved by minimizing a positive-definite quadratic functional. The third point is that a positive-definite quadratic functional can be minimized with a single iteration of Newton's method for any starting point $x \in \mathbb{R}^n$. The final observation is that by judiciously choosing the starting point to be the solution of the previous state estimate, several terms cancel and the Newton update simplifies to the one-step Kalman filter. We remark that this approach to state estimation can be generalized to include predictive estimates on future states and smoothed estimates on previous states. We can also use Newton's method to derive the extended Kalman filter.

ACKNOWLEDGMENTS

The authors thank Randall W. Beard, Magnus Egerstedt, C. Shane Reese, and H. Dennis Tolley for useful conversations. This work was supported in part by the National Science Foundation, Division of Mathematical Sciences, grant numbers CAREER 0847074 and CSUMS 0639328.

AUTHOR INFORMATION

Jeffrey Humpherys received the Ph.D. in mathematics from Indiana University in 2002 and was an Arnold Ross Assistant Professor at The Ohio State University from 2002-2005. Since 2005, he has been at Brigham Young University, where he is currently an associate professor of mathematics. His interests are in applied mathematical analysis and numerical computation.

Jeremy West received B.S. degrees in mathematics and computer science and the M.S. in mathematics in 2009, all from Brigham Young University. He is currently a Ph.D. student in mathematics at the University of Michigan. His interests are in discrete mathematics.

REFERENCES

[1] J. L. Casti, *Five More Golden Rules: Knots, Codes, Chaos, and other Great Theories of 20th-Century Mathematics*. New York: Wiley, 2000.

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

[3] D. Simon, *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*. New York: Wiley-Interscience, 2006.

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

[5] R. Brown and P. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*. New York: Wiley, 1997.

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

[7] C. T. Kelley, *Solving Nonlinear Equations with Newton's Method*. Philadelphia, PA: SIAM, 2003.

[8] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.