A Matrix Theoretic Derivation of the Kalman Filter

Johnathan M. Bardsley

Abstract The Kalman filter is a data analysis method used in a wide range of engineering and applied mathematics problems. This paper presents a matrix-theoretic derivation of the method in the linear model, Gaussian measurement error case. Standard derivations of the Kalman filter make use of probabilistic notation and arguments, whereas we make use, primarily, of methods from numerical linear algebra. In addition to the standard Kalman filter, we derive an equivalent variational (optimization-based) formulation, as well as the extended Kalman filter for nonlinear problems.

1 Introduction

We start with the standard linear model with Gaussian measurment error:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{b} \in \mathbb{R}^m$ is measured data; $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a known observation matrix; $\mathbf{x} \in \mathbb{R}^n$ is the unknown parameter vector to be estimated; $\mathbf{e} \in \mathbb{R}^m$ is a zero-mean Gaussian random vector with covariance matrix $\mathbf{C_e}$, which we denote by $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{C_e})$; and $\mathbf{x} \in \mathbb{R}^n$ is the unknown vector that is to be estimated.

The standard technique for estimating $\mathbf{x}$, known as *least squares estimation*, was developed by Gauss in his study of planetary motion [1]. The extension of least squares estimation to the case when the unknown $\mathbf{x}$ is also assumed to be a Gaussian random vector, which will be the case for us, is known as *minimum variance estimation* [4].

In the study of time varying phenomena, it is natural to generalize (1) as follows:

Department of Mathematical Sciences, University of Montana, Missoula, MT 59812. e-mail: bardsleyj@mso.umt.edu

$$\mathbf{x}_k = \mathbf{M}_k \mathbf{x}_{k-1} + \mathbf{E}_k, \tag{2}$$
$$\mathbf{b}_k = \mathbf{A}_k \mathbf{x}_k + \mathbf{e}_k, \tag{3}$$

where equation (3) is defined analogous to (1) for each k; and in (2), $\mathbf{M}_k \in \mathbb{R}^{n\times n}$ is the known evolution matrix, $\mathbf{x}_{k-1}$ is a Gaussian random vector, and $\mathbf{E}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{E}_k})$. The Kalman filter [2, 5] is the extension of minimum variance (and hence least squares) estimation to the problem of sequentially estimating $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ given data $\{\mathbf{b}_1, \mathbf{b}_2, \ldots\}$ arising from the model in (2), (3). For the interested reader, a discussion of the progression of ideas from Gauss to Kalman is the subject of the excellent paper [3].

This paper is organized as follows. First, in Section 2, we present the basic statistical definitions and results that we will need in our later discussion. In Section 3, we define the minimum variance estimator, which we then apply to (2), (3) to derive the Kalman filter in Section 4. Finally, we present an equivalent formulation of the Kalman filter, which we call the variational Kalman filter, as well as the extended Kalman filter for the case when (2), (3) contain nonlinear evolution and/or observation operators.

2 Statistical Preliminaries

Let $\mathbf{x} = (x_1, \ldots, x_n)^T$ be a random vector with $E(x_i)$ the mean of x_i and $E((x_i - \mu_i)^2)$, where $\mu_i = E(x_i)$, its variance. The mean of $\mathbf{x}$ is then defined $E(\mathbf{x}) = (E(x_1), \ldots, E(x_n))^T$, while the $n \times n$ covariance matrix of $\mathbf{x}$ is defined

$$[\mathrm{cov}(\mathbf{x})]_{ij} = E((x_i - \mu_i)(x_j - \mu_j)), \quad 1 \le i, j \le n.$$

Note that the diagonal of $\mathrm{cov}(\mathbf{x})$ contains the variances of $x_1, \ldots, x_n$, while the off diagonal elements contain the covariance values. Thus if x_i and x_j are independent $[\mathrm{cov}(\mathbf{x})]_{ij} = 0$ for $i \ne j$.

The $n \times m$ cross correlation matrix of the random n-vector $\mathbf{x}$ and m-vector $\mathbf{y}$, which we will denote $\Gamma_{\mathbf{xy}}$, is defined

$$\Gamma_{\mathbf{xy}} = E(\mathbf{x}\mathbf{y}^T), \tag{4}$$

where $[E(\mathbf{x}\mathbf{y}^T)]_{ij} = E(x_i y_j)$. If $\mathbf{x}$ and $\mathbf{y}$ are independent, then $\Gamma_{\mathbf{xy}}$ is the zero matrix. Furthermore,

$$E(\mathbf{x}) = \mathbf{0} \quad \text{implies} \quad \Gamma_{\mathbf{xx}} = \mathrm{cov}(\mathbf{x}). \tag{5}$$

Finally, given an $m \times n$ matrix $\mathbf{A}$ and a random n-vector $\mathbf{x}$, it is not difficult to show that

$$\mathrm{cov}(\mathbf{A}\mathbf{x}) = \mathbf{A}\mathrm{cov}(\mathbf{x})\mathbf{A}^T. \tag{6}$$

We end these preliminary comments with the probability density function of primary interest to us in this paper, the Gaussian distribution. If $\mathbf{b}$ is an $n \times 1$ Gaussian random vector, then its probability density function has the form

$$p_{\mathbf{b}}(\mathbf{b};\mu,\mathbf{C}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C})}} \exp\left(-\frac{1}{2}(\mathbf{b}-\mu)^T \mathbf{C}^{-1}(\mathbf{b}-\mu)\right), \tag{7}$$

where $\mu \in \mathbb{R}^n$ is the mean of $\mathbf{b}$; $\mathbf{C}$ is an $n \times n$ symmetric positive definite covariance matrix of $\mathbf{b}$; and $\det(\cdot)$ denotes matrix determinant. As above, we will use the notation $\mathbf{b} \sim \mathcal{N}(\mu, \mathbf{C})$ in this case. For more details on introductory mathematical statistics, see one of many introductory mathematics statistics texts.

3 Minimum Variance Estimation

First, we consider model (1). When $\mathbf{x}$ is assumed to be deterministic, it is a standard exercise to show that if $\mathbf{A}$ has full column rank, the least squares estimator is given by

$$\mathbf{x}^{ls} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}.$$

However, we are interested in the case in which $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{x}})$. We assume, furthermore, that $\mathbf{x}$ and $\mathbf{e}$ are independent random variables. We now define the minimum variance estimator of $\mathbf{x}$.

Definition 1. Suppose $\mathbf{b}$ is defined as in (1), $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{x}})$, and $\mathbf{e}$ and $\mathbf{x}$ independent random vectors. Then the *minimum variance estimator* of $\mathbf{x}$ given $\mathbf{b}$ has the form

$$\mathbf{x}^{est} = \hat{\mathbf{B}}\mathbf{b},$$

where $\hat{\mathbf{B}} \in \mathbb{R}^{n\times m}$ solves the optimization problem

$$\hat{\mathbf{B}} = \arg\min_{\mathbf{B}\in\mathbb{R}^{n\times m}} E\left(\|\mathbf{B}\mathbf{b}-\mathbf{x}\|\|_2^2\right).$$

Because our model (1) is a linear model with Gaussian measurment error, $\hat{\mathbf{B}}$ has an elegant closed form, as described in the following theorem.

Theorem 1. *If $\Gamma_{\mathbf{bb}}$ is invertible, then the minimum variance estimator of $\mathbf{x}$ from $\mathbf{b}$ is given by*

$$\mathbf{x}^{est} = (\Gamma_{\mathbf{xb}}\Gamma_{\mathbf{bb}}{}^{-1})\mathbf{b}.$$

Proof. First, we note that

$$\begin{aligned} E(\|\mathbf{B}\mathbf{b}-\mathbf{x}\|_2^2) &= \operatorname{trace}\left(E[(\mathbf{B}\mathbf{b}-\mathbf{x})(\mathbf{B}\mathbf{b}-\mathbf{x})^T]\right), \\ &= \operatorname{trace}\left(\mathbf{B}E[\mathbf{b}\mathbf{b}^T]\mathbf{B}^T - \mathbf{B}E[\mathbf{b}\mathbf{x}^T] - E[\mathbf{x}\mathbf{b}^T]\mathbf{B}^T + E[\mathbf{x}\mathbf{x}^T]\right). \end{aligned}$$

Then, using the distributive property of the trace function and the identity

$$\frac{d}{d\mathbf{B}}\operatorname{trace}(\mathbf{B}^T\mathbf{C}) = \left(\frac{d}{d\mathbf{B}}\operatorname{trace}(\mathbf{B}\mathbf{C})\right)^T = \mathbf{C},$$

we see that $dE(\|\mathbf{B}\mathbf{b}-\mathbf{x}\|^2)/d\mathbf{B} = \mathbf{0}$ when

$$\hat{\mathbf{B}} = \Gamma_{\mathbf{xb}}\Gamma_{\mathbf{bb}}{}^{-1},$$

which establishes the result.

In the context of (1), and given our assumptions stated above, we can obtain a more concrete form for the minimum variance estimator. In particular, we note that since $\mathbf{x}$ and $\mathbf{e}$ are assumed to be independent, $\Gamma_{\mathbf{xe}} = \Gamma_{\mathbf{ex}} = \mathbf{0}$. Hence, using (1), we obtain

$$\begin{aligned} \Gamma_{\mathbf{xb}} &= E[\mathbf{x}(\mathbf{Ax}+\mathbf{e})^T], \\ &= \Gamma_{\mathbf{xx}}\mathbf{A}^T. \end{aligned}$$

Similarly,

$$\begin{aligned} \Gamma_{\mathbf{bb}} &= E[(\mathbf{Ax}+\mathbf{e})(\mathbf{Ax}+\mathbf{e})^T], \\ &= \mathbf{A}\Gamma_{\mathbf{xx}}\mathbf{A}^T + \Gamma_{\mathbf{ee}}. \end{aligned}$$

Thus, since $\Gamma_{\mathbf{xx}} = \mathbf{C}_{\mathbf{x}}$ and $\Gamma_{\mathbf{ee}} = \mathbf{C}_{\mathbf{e}}$, the minimum variance estimator has the form

$$\begin{aligned} \mathbf{x}^{est} &= \hat{\mathbf{B}}\mathbf{b} \\ &= \mathbf{C}_{\mathbf{x}}\mathbf{A}^T(\mathbf{A}\mathbf{C}_{\mathbf{x}}\mathbf{A}^T + \mathbf{C}_{\mathbf{e}})^{-1}\mathbf{b}, \\ &= (\mathbf{A}^T\mathbf{C}_{\mathbf{e}}^{-1}\mathbf{A} + \mathbf{C}_{\mathbf{x}}^{-1})^{-1}\mathbf{A}^T\mathbf{C}_{\mathbf{e}}^{-1}\mathbf{b}. \end{aligned} \tag{8}$$

We note, in passing, that (8) can also be expressed as

$$\mathbf{x}^{est} = \arg\min_{\mathbf{x}} \left\{ \|\mathbf{Ax}-\mathbf{b}\|^2_{\mathbf{C}_{\mathbf{e}}^{-1}} + \|\mathbf{x}\|^2_{\mathbf{C}_{\mathbf{x}}^{-1}} \right\}, \tag{9}$$

where "arg min" denotes "argument of the minimum" and $\|\mathbf{x}\|^2_{\mathbf{C}} \stackrel{\text{def}}{=} \mathbf{x}^T\mathbf{C}\mathbf{x}$. This establishes a clear connection between minimum variance estimation and generalized Tikhonov regularization [4]. Note in particular that if $\mathbf{C}_{\mathbf{e}} = \sigma_1^2\mathbf{I}$ and $\mathbf{C}_{\mathbf{x}} = \sigma_2^2\mathbf{I}$, problem (9) can be equivalently expressed as

$$\mathbf{x}^{est} = \arg\min_{\mathbf{x}} \left\{ \|\mathbf{Ax}-\mathbf{b}\|_2^2 + (\sigma_1^2/\sigma_2^2)\|\mathbf{x}\|_2^2 \right\},$$

which has classical Tikhonov form. This formulation is also equivalent to maximum a posteriori (MAP) estimation.

4 The Kalman Filter

In the previous section, we considered the stationary linear model (1), but suppose our model now has the form (2), (3). Equation (2) is the equation of evolution for $\mathbf{x}_k$ with $\mathbf{M}_k$ the $n \times n$ linear evolution matrix, and $\mathbf{E}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{E}_k})$. In equation (3), $\mathbf{b}_k$ denotes the $m \times 1$ observed data, $\mathbf{A}_k$ the $m \times n$ linear observation matrix, and $\mathbf{e}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{e}_k})$. In both equations, k denotes the time index.

The problem is to estimate $\mathbf{x}_k$ at time k from $\mathbf{b}_k$ and an estimate $\mathbf{x}_{k-1}^{est}$ of the state at time $k-1$. We assume $\mathbf{x}_{k-1}^{est} \sim \mathcal{N}(\mathbf{x}_{k-1}, \mathbf{C}_{k-1}^{est})$. To facilitate a more straightforward application of the result of Theorem 1, we rewrite (2), (3). First, define

$$\mathbf{x}_k^a = \mathbf{M}_k \mathbf{x}_{k-1}^{est} \tag{10}$$

$$\mathbf{z}_k = \mathbf{x}_k - \mathbf{x}_k^a, \tag{11}$$

$$\mathbf{r}_k = \mathbf{b}_k - \mathbf{A}_k \mathbf{x}_k^a. \tag{12}$$

Then, subtracting (10) from (2) and $\mathbf{A}_k \mathbf{x}_k^a$ from both sides of (3), and dropping the k dependence for notational simplicity, we obtain the stochastic linear equations

$$\mathbf{z} = \mathbf{M}(\mathbf{x} - \mathbf{x}^{est}) + \mathbf{E}, \tag{13}$$

$$\mathbf{r} = \mathbf{A}\mathbf{z} + \mathbf{e}. \tag{14}$$

The minimum variance estimator of $\mathbf{z}$ from $\mathbf{r}$ given (13), (14) is then given, via Theorem 1 (note that $\mathbf{z}$ is a zero mean Gaussian random vector), by

$$\mathbf{z}^{est} = \Gamma_{\mathbf{zr}} \Gamma_{\mathbf{rr}}^{-1} \mathbf{r}.$$

We assume that $\mathbf{x} - \mathbf{x}^{est}$ is independent of $\mathbf{E}$, and that $\mathbf{z} = \mathbf{x} - \mathbf{x}^a$ is independent of $\mathbf{e}$. Then, from (4), (5), (6), (13) and (14), we obtain

$$\Gamma_{\mathbf{zz}} = \mathbf{M}\mathbf{C}^{est}\mathbf{M}^T + \mathbf{C}_{\mathbf{E}} \stackrel{\text{def}}{=} \mathbf{C}^a, \tag{15}$$

$$\Gamma_{\mathbf{zr}} = \mathbf{C}^a \mathbf{A}^T,$$

$$\Gamma_{\mathbf{rr}} = \mathbf{A}\mathbf{C}^a\mathbf{A}^T + \mathbf{C}_{\mathbf{e}}.$$

where $\mathbf{C}^{est}$ and $\mathbf{C}^a$ are the covariance matrices for $\mathbf{x}^{est}$ and $\mathbf{x}^a$, respectively. Thus, finally, the minimum variance estimator of $\mathbf{z}$ is given by

$$\mathbf{z}^{est} = \mathbf{C}^a\mathbf{A}^T(\mathbf{A}\mathbf{C}^a\mathbf{A}^T + \mathbf{C}_{\mathbf{e}})^{-1}\mathbf{r}, \tag{16}$$

From (16) and (11) we then immediately obtain the Kalman Filter estimate of $\mathbf{x}$ given by

$$\mathbf{x}_+^{est} = \mathbf{x}^a + \mathbf{H}(\mathbf{b} - \mathbf{A}\mathbf{x}^a), \tag{17}$$

where

$$\mathbf{H} = \mathbf{C}^a\mathbf{A}^T(\mathbf{A}\mathbf{C}^a\mathbf{A}^T + \mathbf{C}_{\mathbf{e}})^{-1} \tag{18}$$

is known as the Kalman Gain matrix.

Finally, in order to compute the covariance of $\mathbf{x}_+^{est}$, we note that by (17) and (3),

$$\mathbf{x}_+^{est} = (\mathbf{I} - \mathbf{H}\mathbf{A})\mathbf{x}^a + \mathbf{H}\mathbf{e} + \mathbf{H}\mathbf{A}\mathbf{x},$$

where $\mathbf{x}$ is the true state. Given our assumptions and using (6), the covariance then takes the form

$$\mathbf{C}_{+}^{est} = (\mathbf{I}-\mathbf{HA})\mathbf{C}^{a}(\mathbf{I}-\mathbf{HA})^{T} + \mathbf{H}\mathbf{C}_{\mathbf{e}}\mathbf{H}^{T},$$

which can be rewritten, using the identity $\mathbf{H}\mathbf{C}_{\mathbf{e}}\mathbf{H}^{T} = (\mathbf{I}-\mathbf{HA})\mathbf{C}^{a}\mathbf{A}^{T}\mathbf{H}^{T}$, in the simplified form

$$\mathbf{C}_{+}^{est} = \mathbf{C}^{a} - \mathbf{HAC}^{a}. \tag{19}$$

Incorporating the k dependence again leads directly to the Kalman filter iteration.

The Kalman Filter Algorithm

Step 0: Select initial guess $\mathbf{x}_0^{est}$ and covariance $\mathbf{C}_0^{est}$, and set $k = 1$.

Step 1: Compute the evolution model estimate and covariance:
A. Compute $\mathbf{x}_k^{a} = \mathbf{M}_k\mathbf{x}_{k-1}^{est}$;
B. Compute $\mathbf{C}_k^{a} = \mathbf{M}_k\mathbf{C}_{k-1}^{est}\mathbf{M}_k^{T} + \mathbf{C}_{\mathbf{E}_k} := \mathbf{C}_k^{a}$.

Step 2: Compute the Kalman filter estimate and covariance:
A. Compute the Kalman Gain $\mathbf{H}_k = \mathbf{C}_k^{a}\mathbf{A}_k^{T}(\mathbf{A}_k\mathbf{C}_k^{a}\mathbf{A}_k^{T} + \mathbf{C}_{\mathbf{e}})^{-1}$;
B. Compute the estimate $\mathbf{x}_k^{est} = \mathbf{x}_k^{a} + \mathbf{H}_k(\mathbf{b}_k - \mathbf{A}_k\mathbf{x}_k^{a})$;
C. Compute the estimate covariance $\mathbf{C}_k^{est} = \mathbf{C}_k^{a} - \mathbf{H}_k\mathbf{A}_k\mathbf{C}_k^{a}$.

Step 3: Update $k := k+1$ and return to Step 1.

4.1 A Variational Formulation of the Kalman Filter

As in the stationary case (see (8), (9)), we can rewrite equation (16) in the form

$$\mathbf{z}^{est} = (\mathbf{A}^{T}\mathbf{C}_{\mathbf{e}}^{-1}\mathbf{A} + (\mathbf{C}^{a})^{-1})^{-1}\mathbf{A}^{T}\mathbf{C}_{\mathbf{e}}^{-1}\mathbf{r},$$

which, yields, using (11), the Kalman filter estimate

$$\begin{aligned}\mathbf{x}_{+}^{est} &= \mathbf{x}^{a} + [\mathbf{A}^{T}\mathbf{C}_{\mathbf{e}}^{-1}\mathbf{A} + (\mathbf{C}^{a})^{-1}]^{-1}\mathbf{A}^{T}\mathbf{C}_{\mathbf{e}}^{-1}(\mathbf{b}-\mathbf{A}\mathbf{x}^{a}),\\ &= \arg\min_{\mathbf{x}}\left\{\ell(\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{2}(\mathbf{b}-\mathbf{A}\mathbf{x})^{T}\mathbf{C}_{\mathbf{e}}^{-1}(\mathbf{b}-\mathbf{A}\mathbf{x}) + \frac{1}{2}(\mathbf{x}-\mathbf{x}^{a})^{T}(\mathbf{C}^{a})^{-1}(\mathbf{x}-\mathbf{x}^{a})\right\}.\end{aligned}$$

It can be shown using a Taylor series argument that

$$\mathbf{x}_{+}^{est} = \mathbf{x}^{a} - \nabla^{2}\ell(\mathbf{x}^{a})^{-1}\nabla\ell(\mathbf{x}^{a}), \tag{20}$$

where $\nabla\ell$ and $\nabla^{2}\ell$ denote the gradient and Hessian of ℓ respectively, and are given by

$$\begin{aligned}\nabla\ell(\mathbf{x}) &= \mathbf{A}^{T}\mathbf{C}_{\mathbf{e}}^{-1}(\mathbf{b}-\mathbf{A}\mathbf{x}) + (\mathbf{C}^{a})^{-1}(\mathbf{x}-\mathbf{x}^{a}),\\ \nabla^{2}\ell(\mathbf{x}) &= \mathbf{A}^{T}\mathbf{C}_{\mathbf{e}}^{-1}\mathbf{A} + (\mathbf{C}^{a})^{-1}.\end{aligned}$$

By the matrix inversion lemma, we have

$$(\mathbf{A}^T\mathbf{C}_\mathbf{e}^{-1}\mathbf{A}+(\mathbf{C}^a)^{-1})^{-1}=\mathbf{C}^a-\mathbf{C}^a\mathbf{A}^T(\mathbf{A}\mathbf{C}^a\mathbf{A}^T+\mathbf{C}_e)^{-1}\mathbf{A}\mathbf{C}^a.$$

Then from equations (18) and (19), we obtain the interesting fact that

$$\mathbf{C}_+^{est}=\nabla^2\ell(\mathbf{x})^{-1}. \tag{21}$$

This allows us to define the following equivalent formulation of the Kalman filter, which we call the variational Kalman filter.

The Variational Kalman Filter Algorithm

Step 0: Select initial guess $\mathbf{x}_0^{est}$ and covariance $\mathbf{C}_0^{est}$, and set $k=1$.

Step 1: Compute the evolution model estimate and covariance:
A. Compute $\mathbf{x}_k^a=\mathbf{M}_k\mathbf{x}_{k-1}^{est}$;
B. Compute $\mathbf{C}_k^a=\mathbf{M}_k\mathbf{C}_k^{est}\mathbf{M}_k^T+\mathbf{C}_{\mathbf{E}_k}:=\mathbf{C}_k^a$.

Step 2: Compute the Kalman filter estimate and covariance:
A. Compute the estimate $\mathbf{x}_k^{est}=\arg\min_\mathbf{x}\ell(\mathbf{x})$;
C. Compute the estimate covariance $\mathbf{C}_k^{est}=\nabla^2\ell(\mathbf{x})^{-1}$.

Step 3: Update $k:=k+1$ and return to Step 1.

A natural question is, what is the use of this equivalent formulation of the Kalman filter? Theoretically there is no benefit gained in using the variational Kalman filter if the estimate and its covariance are computed exactly. However, with the variational approach, the filter estimate, and even its covariance, can be computed approximately using an iterative minimization method, such as conjugate gradient. This is particularly important for large-scale problems where the exact Kalman filter is prohibitively expensive to compute.

4.2 The Extended Kalman Filter

The extended Kalman filter is the extension of the Kalman filter when (2), (3) are replaced by

$$\mathbf{x}_k=\mathscr{M}(\mathbf{x}_{k-1})+\mathbf{E}_k, \tag{22}$$
$$\mathbf{b}_k=\mathscr{A}(\mathbf{x}_k)+\mathbf{e}_k, \tag{23}$$

where $\mathscr{M}$ and $\mathscr{A}$ are (possibly) nonlinear functions. The extended Kalman filter is obtained by the following simple modification of either of the above algorithms: in Step 1, A use, instead, $\mathbf{x}_k^a=\mathscr{M}(\mathbf{x}_k^{est})$, and define

$$\mathbf{M}_k = \frac{\partial \mathscr{M}(\mathbf{x}_{k-1}^{est})}{\partial \mathbf{x}}, \quad \text{and} \quad \mathbf{A}_k = \frac{\partial \mathscr{A}(\mathbf{x}_k^a)}{\partial \mathbf{x}}, \tag{24}$$

where $\frac{\partial f}{\partial \mathbf{x}}$ denotes the Jacobian of f.

5 Conclusions

We have presented a derivation of the Kalman filter that utilizes matrix analysis techniques as well as the Bayesian statistical approach of minimum variance estimation. In addition, we presented an equivalent variational formulation, which we call the variational Kalman filter, as well as the extended Kalman filter for nonlinear problems.

References

1. K. F. Gauss, *Theory of Motion of the Heavenly Bodies*, New York: Dover, 1963.
2. R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME–Journal of Basic Engineering, **82**, Series D, pp. 35–45, 1960.
3. H. W. Sorenson, *Least Squares Estimation: From Gauss to Kalman*, IEEE Spectrum, vol. 7, pp. 63-68, July 1970.
4. C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM 2002.
5. G. Welch and G. Bishop, *An Introduction to the Kalman Filter*, UNC Chapel Hill, Dept. of Computer Science Tech. Report, TR 95-041.