A Matrix Theoretic Derivation of the Kalman Filter*

Johnathan M. Bardsley†

4 September 2008

Abstract

We present a matrix theoretic derivation of the Kalman filter—motivated by the statistical technique of minimum variance estimation—in order to make its theoretical underpinnings accessible to a broader audience. Standard derivations of the filter utilize probabilistic arguments that are less familiar to the matrix analyst and computational mathematician.

1 Introduction

Of key interest to us are stochastic linear systems of the form

$$\mathbf{d} = \mathbf{G}\boldsymbol{\phi} + \mathbf{e}, \tag{1}$$

where $\mathbf{d} \in \mathbb{R}^m$ is measured data; $\mathbf{G} \in \mathbb{R}^{m\times n}$ is a known model matrix; $\boldsymbol{\phi} \in \mathbb{R}^n$ is the unknown parameter vector to be estimated; $\mathbf{e} \in \mathbb{R}^m$ is a zero-mean Gaussian random vector; and $\boldsymbol{\phi} \in \mathbb{R}^n$ is a zero-mean Gaussian random vector. In this case, the minimum variance estimator of $\boldsymbol{\phi}$ [4] can be expressed as the solution of a $n \times n$ linear system.

Our derivation of the Kalman filter [3, 5] arises from an application of minimum variation estimation to the sequentially coupled system of linear equations

$$\boldsymbol{\phi}_k = \mathbf{M}_k\boldsymbol{\phi}_{k-1} + \mathbf{E}_k, \tag{2}$$
$$\mathbf{d}_k = \mathbf{G}_k\boldsymbol{\phi}_k + \mathbf{e}_k, \tag{3}$$

where equation (3) is analogous to (1); and in (2), $\mathbf{M}_k \in \mathbb{R}^n$ is the known evolution matrix, $\boldsymbol{\phi}_{k-1}$ is a Gaussian random vector, and $\mathbf{E}_k \in \mathbb{R}^n$ is a zero-mean Gaussian random vector.

Our discussion requires a knowledge of some basic statistical definitions and results.

*Mathematics Subject Classifications: 15A99, 65C60.

†Department of Mathematical Sciences, University of Montana, Missoula, MT 59812. Email: bardsleyj@mso.umt.edu

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

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

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 $[\mathrm{cov}(\mathbf{x})]_{ij} = 0$.

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

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

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

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

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 example of primary interest to us in this paper, the Gaussian distribution. If $\mathbf{d}$ is an $n \times 1$ Gaussian random vector, then its probability density function has the form

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

where $\boldsymbol{\mu} \in \mathbb{R}^n$, $\mathbf{C}$ is an $n \times n$ symmetric positive definite matrix, and $\det(\cdot)$ denotes matrix determinant. Then $E(\mathbf{d}) = \boldsymbol{\mu}$ and $\mathrm{cov}(\mathbf{d}) = \mathbf{C}$. We will use the notation $\mathbf{d} \sim N(\boldsymbol{\mu}, \mathbf{C})$ in this case.

For more details on introductory mathematical statistics, see one of many introductory mathematics statistics texts, including [2].

3 Minimum Variance Estimation

We consider linear system (1) and assume that $\boldsymbol{\phi} \sim N(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\phi}})$ with $\boldsymbol{C}_{\boldsymbol{\phi}}$ a symmetric positive definite matrix. We assume, furthermore, that $\boldsymbol{\phi}$ and $\mathbf{e}$ are independent random variables. The problem of estimating $\boldsymbol{\phi}$ is statistical, however as we will see, the minimum variance estimator can be expressed as the solution of a linear system. We can now define the minimum variance estimator of $\boldsymbol{\phi}$.

Definition 1. *Suppose $\boldsymbol{\phi}$ and $\mathbf{d}$ are jointly distributed random vectors whose components have finite expected squares. The* minimum variance estimator *of $\boldsymbol{\phi}$ from $\mathbf{d}$ is given by*

$$\boldsymbol{\phi}^{est} = \hat{\mathbf{B}}\mathbf{d},$$

where

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

We now state and prove a theorem that gives us the minimum variance estimator in an elegant closed form.

Theorem 1. *If* $\boldsymbol{\Gamma}_{\mathbf{dd}}$ *(defined in (4)) is invertible, then the minimum variance estimator of* $\boldsymbol{\phi}$ *from* $\mathbf{d}$ *is given by*

$$\boldsymbol{\phi}^{est} = (\boldsymbol{\Gamma}_{\boldsymbol{\phi}\mathbf{d}}\boldsymbol{\Gamma}_{\mathbf{dd}}{}^{-1})\mathbf{d}.$$

Proof. First we note, via properties of the trace function, that

$$\begin{aligned} E(\|\mathbf{B}\mathbf{d} - \boldsymbol{\phi}\|^2) &= \text{trace}(E[(\mathbf{B}\mathbf{d} - \boldsymbol{\phi})(\mathbf{B}\mathbf{d} - \boldsymbol{\phi})^T]), \\ &= \text{trace}\left(\mathbf{B}E[\mathbf{d}\mathbf{d}^T]\mathbf{B}^T - \mathbf{B}E[\mathbf{d}\boldsymbol{\phi}^T] - E[\boldsymbol{\phi}\mathbf{d}^T]\mathbf{B}^T + E[\boldsymbol{\phi}\boldsymbol{\phi}^T]\right). \end{aligned}$$

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

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

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

$$\mathbf{B} = \boldsymbol{\Gamma}_{\boldsymbol{\phi}\mathbf{d}}\boldsymbol{\Gamma}_{\mathbf{dd}}{}^{-1},$$

where $\boldsymbol{\Gamma}_{\boldsymbol{\phi}\mathbf{d}}$ and $\boldsymbol{\Gamma}_{\mathbf{dd}}$ are defined in (4). □

Given our assumptions above – recall that $\boldsymbol{\phi}$ and $\mathbf{e}$ are independent zero mean random vectors – we have from (4) and (5)

$$\begin{aligned} \boldsymbol{\Gamma}_{\boldsymbol{\phi}\mathbf{d}} &= \boldsymbol{\Gamma}_{\boldsymbol{\phi}\boldsymbol{\phi}}\mathbf{G}^T + \boldsymbol{\Gamma}_{\boldsymbol{\phi}\mathbf{e}} = \boldsymbol{C}_{\boldsymbol{\phi}}\mathbf{G}^T, \\ \boldsymbol{\Gamma}_{\mathbf{dd}} &= \mathbf{G}\boldsymbol{\Gamma}_{\boldsymbol{\phi}\boldsymbol{\phi}}\mathbf{G}^T + \Gamma_{\mathbf{ee}} = \mathbf{G}\boldsymbol{C}_{\boldsymbol{\phi}}\mathbf{G}^T + \mathbf{C}_{\mathbf{e}}. \end{aligned}$$

Thus the minimum variance estimator is given by

$$\begin{aligned} \boldsymbol{\phi}^{est} &= \boldsymbol{C}_{\boldsymbol{\phi}}\mathbf{G}^T(\mathbf{G}\boldsymbol{C}_{\boldsymbol{\phi}}\mathbf{G}^T + \mathbf{C}_{\mathbf{e}})^{-1}\mathbf{d}, \\ &= (\mathbf{G}^T\mathbf{C}_{\mathbf{e}}^{-1}\mathbf{G} + \boldsymbol{C}_{\boldsymbol{\phi}}{}^{-1})^{-1}\mathbf{G}^T\mathbf{C}_{\mathbf{e}}^{-1}\mathbf{d}. \end{aligned} \tag{8}$$

The second equality is valid since $\boldsymbol{C}_{\boldsymbol{\phi}}$ is assumed to be nonsingular (a nice exercise), and it shows us that in this case $\boldsymbol{\phi}^{est}$ can also be expressed

$$\boldsymbol{\phi}^{est} = \arg\min_{\boldsymbol{\phi}} \|\mathbf{G}\boldsymbol{\phi} - \mathbf{d}\|^2_{\mathbf{C}_{\mathbf{e}}^{-1}} + \|\boldsymbol{\phi}\|^2_{\mathbf{C}_{\boldsymbol{\phi}}^{-1}}. \tag{9}$$

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 $\boldsymbol{C}_{\boldsymbol{\phi}} = \sigma_2^2\mathbf{I}$, the minimum variance estimate can be written

$$\boldsymbol{\phi}^{est} = \arg\min_{\boldsymbol{\phi}} \|\mathbf{G}\boldsymbol{\phi} - \mathbf{d}\|_2^2 + (\sigma_1^2/\sigma_2^2)\|\boldsymbol{\phi}\|_2^2,$$

which has classical Tikhonov form.

4 The Kalman Filter

Up to this point, we have considered stationary linear models, but suppose that what we wish to estimate ($\boldsymbol{\phi}$ above) changes in time. More specifically, suppose our model now has the form (1), (2). Equation (1) is the equation of evolution for $\boldsymbol{\phi}_k$ with $\mathbf{M}_k$ the $n \times n$ linear evolution matrix, and $\mathbf{E}_k \sim N(\mathbf{0}, \mathbf{C}_{\mathbf{E}_k})$. In equation (2), $\mathbf{d}_k$ denotes the $m \times 1$ observed data, $\mathbf{G}_k$ the $m \times n$ linear observation matrix, and $\mathbf{e}_k \sim N(\mathbf{0}, \mathbf{C}_{\mathbf{e}_k})$. In both equations, k denotes the time index.

The problem is to estimate $\boldsymbol{\phi}_k$ at time k from $\mathbf{d}_k$ and an estimate $\boldsymbol{\phi}_{k-1}^{est}$ of the state at time $k-1$. We assume $\boldsymbol{\phi}_{k-1}^{est} \sim N(\boldsymbol{\phi}_{k-1}, \mathbf{C}_{k-1}^{est})$.

To facilitate a more straightforward application of the result of Theorem 1, we rewrite (1), (2). First, define

$$\begin{aligned} \boldsymbol{\phi}_k^a &= \mathbf{M}_k \boldsymbol{\phi}_{k-1}^{est} &(10)\\ \mathbf{z}_k &= \boldsymbol{\phi}_k - \boldsymbol{\phi}_k^a, &(11)\\ \mathbf{r}_k &= \mathbf{d}_k - \mathbf{G}_k \boldsymbol{\phi}_k^a. &(12) \end{aligned}$$

Then, subtracting (10) from (1) and $\mathbf{G}_k \boldsymbol{\phi}_k^a$ from both sides of (2), and dropping the k dependence for notational simplicity, we obtain the stochastic linear equations

$$\begin{aligned} \mathbf{z} &= \mathbf{M}(\boldsymbol{\phi} - \boldsymbol{\phi}^{est}) + \mathbf{E}, &(13)\\ \mathbf{r} &= \mathbf{G}\mathbf{z} + \mathbf{e}. &(14) \end{aligned}$$

The minimum variance estimator of $\mathbf{z}$ from $\mathbf{r}$ given (13), (14) is then given, via Theorem 1, by

$$\mathbf{z}^{est} = \boldsymbol{\Gamma}_{\mathbf{zr}} \boldsymbol{\Gamma}_{\mathbf{rr}}{}^{-1} \mathbf{r}.$$

We assume that $\boldsymbol{\phi} - \boldsymbol{\phi}^{est}$ is independent of $\mathbf{E}$, and that $\mathbf{z} = \boldsymbol{\phi} - \boldsymbol{\phi}^a$ is independent of $\mathbf{e}$. We note, furthermore, that by assumption both of these random vectors have zero mean. Then, from (4), (5), (6), (13) and (14), we obtain

$$\begin{aligned} \boldsymbol{\Gamma}_{\mathbf{zz}} &= \mathbf{M}\mathbf{C}^{est}\mathbf{M}^T + \mathbf{C}_{\mathbf{E}} \stackrel{\text{def}}{=} \mathbf{C}^a, &(15)\\ \boldsymbol{\Gamma}_{\mathbf{zr}} &= \mathbf{C}^a \mathbf{G}^T, \\ \boldsymbol{\Gamma}_{\mathbf{rr}} &= \mathbf{G}\mathbf{C}^a\mathbf{G}^T + \mathbf{C}_{\mathbf{e}}. \end{aligned}$$

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

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

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

$$\boldsymbol{\phi}_+^{est} = \boldsymbol{\phi}^a + \mathbf{H}(\mathbf{d} - \mathbf{G}\boldsymbol{\phi}^a). \quad (17)$$

where

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

is known as the Kalman Gain.

Finally, in order to compute the covariance of $\boldsymbol{\phi}_+^{est}$, we note that by (17) and (2),

$$\boldsymbol{\phi}_+^{est} = (\mathbf{I} - \mathbf{HG})\boldsymbol{\phi}^a + \mathbf{He} + \mathbf{HG}\boldsymbol{\phi},$$

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

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

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

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

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

The Kalman Filter Algorithm

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

Step 1: Compute the evolution model estimate and covariance:
A. Compute $\boldsymbol{\phi}_k^a = \mathbf{M}_k\boldsymbol{\phi}_{k-1}^{est}$;
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 $\mathbf{H}_k = \mathbf{C}_k^a\mathbf{G}_k^T(\mathbf{G}_k\mathbf{C}_k^a\mathbf{G}_k^T + \mathbf{C}_\mathbf{e})^{-1}$;
B. Compute the estimate $\boldsymbol{\phi}_k^{est} = \boldsymbol{\phi}_k^a + \mathbf{H}_k(\mathbf{d}_k - \mathbf{G}_k\boldsymbol{\phi}_k^a)$;
C. Compute the estimate covariance $\mathbf{C}_k^{est} = \mathbf{C}_k^a - \mathbf{H}_k\mathbf{G}_k\mathbf{C}_k^a$.

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)), we can rewrite equation (16) in the form

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

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

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

It can be shown using a Taylor series argument that

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

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

$$\begin{aligned}\nabla\ell(\boldsymbol{\phi}) &= \mathbf{G}^T\mathbf{C}_\mathbf{e}^{-1}(\mathbf{d} - \mathbf{G}\boldsymbol{\phi}) + (\mathbf{C}^a)^{-1}(\boldsymbol{\phi} - \boldsymbol{\phi}^a), \\ \nabla^2\ell(\boldsymbol{\phi}) &= \mathbf{G}^T\mathbf{C}_\mathbf{e}^{-1}\mathbf{G} + (\mathbf{C}^a)^{-1}.\end{aligned}$$

By the matrix inversion lemma, we have

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

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

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

This allows us to define an iteration analogous to that given above.

The Variational Kalman Filter Algorithm

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

Step 1: Compute the evolution model estimate and covariance:
A. Compute $\boldsymbol{\phi}_k^a = \mathbf{M}_k\boldsymbol{\phi}_{k-1}^{est}$;
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 $\boldsymbol{\phi}_k^{est} = \arg\min_{\boldsymbol{\phi}} \ell(\boldsymbol{\phi})$;
C. Compute the estimate covariance $\mathbf{C}_k^{est} = \nabla^2\ell(\boldsymbol{\phi})^{-1}$.

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 [1], can be computed approximately using an iterative minimization method. 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 (EKF) is the extension of the Kalman filter when (1), (2) are replaced by

$$\boldsymbol{\phi}_k = \mathcal{M}(\boldsymbol{\phi}_{k-1}) + \mathbf{E}_k, \tag{22}$$
$$\mathbf{d}_k = \mathcal{G}(\boldsymbol{\phi}_k) + \mathbf{e}_k, \tag{23}$$

where $\mathcal{M}$ and $\mathcal{G}$ are (possibly) nonlinear functions. EKF is obtained by the following simple modification of either of the above algorithms: in Step 1, A use, instead, $\boldsymbol{\phi}_k^a = \mathcal{M}(\boldsymbol{\phi}_k^{est})$, and define

$$\mathbf{M}_k = \frac{\partial\mathcal{M}(\boldsymbol{\phi}_{k-1}^{est})}{\partial\boldsymbol{\phi}}, \quad \text{and} \quad \mathbf{G}_k = \frac{\partial\mathcal{G}(\boldsymbol{\phi}_k^a)}{\partial\boldsymbol{\phi}}, \tag{24}$$

where $\frac{\partial}{\partial\boldsymbol{\phi}}$ denotes the Jacobian.

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 of the Kalman filter, as well as the extended Kalman filter for nonlinear problems.

References

[1] A. Auvinen, J. Bardsley, H. Haario, and T. Kauranne, "Large-Scale Kalman Filtering Using the Limited Memory BFGS Method", submitted.

[2] Peter Bickel and Kjell Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics, Vol. 1*, Prentice Hall, New Jersey, 2001.

[3] Rudolph 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.

[4] Curtis R. Vogel, *Computational Methods for Inverse Problems*, SIAM 2002.

[5] Greg Welch and Gary Bishop, *An Introduction to the Kalman Filter*, UNC Chapel Hill, Dept. of Computer Science Tech. Report, TR 95-041.