Correspondence

Riccati Equation and EM Algorithm Convergence for Inertial Navigation Alignment

Garry A. Einicke, John T. Malos, David C. Reid, and David W. Hainsworth

***Abstract*—This correspondence investigates the convergence of a Kalman filter-based expectation-maximization (EM) algorithm for estimating variances. It is shown that if the variance estimates and the error covariances are initialized appropriately, the underlying Riccati equation solution and the sequence of iterations will be monotonically nonincreasing. Further, the process noise variance estimates converge to the actual values when the measurement noise becomes negligibly small. Conversely, when the process noise variance becomes negligible, the measurement noise variance estimates asymptotically approach the true values. An inertial navigation application is discussed in which performance depends on accurately estimating the process variances.**

***Index Terms*—Inertial navigation, Kalman filtering, parameter estimation, stationary alignment.**

I. Introduction

Inertial navigation systems (see [1]–[5]) typically use optimal minimum-variance filters to track platform trajectories. However, attaining good tracking performance requires precise knowledge of the underlying state-space model parameters and noise statistics. An iterative technique for estimating these unknowns is the expectation-maximization (EM) algorithm which is described in [6]–[9].

The EM algorithm for parameter estimation was first proposed by Dempster, Laird, and Rubin [6]. The procedure consists of iterating two steps: an expectation step and a maximization step. The expectation step of [6] involves least squares calculations on the incomplete observations using the current parameter iterations to estimate the underlying states. The maximization step involves re-estimating the parameters by maximizing a joint log likelihood function using state estimates from the previous expectation step. This sequence is repeated for either a finite number of iterations or until the estimates and the log likelihood function are stable. The paper [6] established parameter map conditions for the convergence of the algorithm, namely that the incomplete data log likelihood function is monotonically nonincreasing. Wu [7] subsequently noted an equivalence between the conditions for a map to be closed and the continuity of a function. In particular, if the likelihood function satisfies certain modality, continuity and differentiability conditions, the parameter sequence converges to some stationary value. In [8], a Kalman filter is used within the expectation step to recover the states. A multiparameter estimation problem is decoupled into separate maximum-likelihood estimations (MLEs) within the EM algorithm of [9]. Applications of the EM algorithm include equalization [10], speech model parameter identification [11], economic forecasting [8] and tomography [12].

This correspondence addresses the problem of estimating the variances from incomplete observations which has been previously studied in [13]–[15]. It is noted in [6] that the likelihood functions for variance estimation do not exist in explicit closed form. This precludes straightforward calculation of the Hessians required in [7] to assert convergence. Therefore, an alternative analysis is presented to establish the monotonicity of variance iterations. Here, the expectation step employs an approach introduced in [8] that involves calculating optimal state estimates which relies on solving Riccati equations. The maximization step involves the calculation of decoupled MLEs similarly to [9]. As is the case in [7], it is shown under prescribed conditions that the estimate sequences will be monotonic nonincreasing and asymptotically approach the exact values.

The correspondence is organized as follows. The monotonicity properties of Riccati difference equation (RDE) solutions are discussed in Section II. Conditions for the monotonicity and convergence of measurement and process noise variance estimates are set out in Section III. It is shown that if the solution to the design Riccati equation is monotonically nonincreasing, and if the estimate sequences are suitably initialized, they will also be monotonically nonincreasing. Further, as the measurement noise becomes negligible and the states are reconstructed exactly, the process noise variance iterations asymptotically converge to the actual values. Conversely, when the process noise variance becomes negligible, the measurement noise variance estimates asymptotically approach the true values. The identification of process noise variances for the stationary alignment of inertial navigation equations is demonstrated in Section IV.

II. Some Properties of Riccati Difference Equations

Consider a linear system having the state-space realization

$$x_{k+1} = Ax_k + w_k \tag{1}$$
$$z_k = Cx_k + v_k \tag{2}$$

where $A \in \mathbb{R}^{nxn}$, $C \in \mathbb{R}^{pxn}$, $x_k \in \mathbb{R}^n$, $w_k \in \mathbb{R}^n$, $v_k \in \mathbb{R}^p$. It is assumed that the process noise w_k and measurement noise v_k are independent, zero mean, stationary, white processes, with actual covariances $E\{w_k w_k^T\} = Q$ and $E\{v_k v_k^T\} = R$, respectively. The optimal Kalman filter [16] which estimates the states x_k from the measurements z_k is given by

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + L_k(z_k - C\hat{x}_{k/k-1}) \tag{3}$$
$$\hat{x}_{k+1/k} = A\hat{x}_{k/k} \tag{4}$$

where $L_k = P_{k/k-1}C^T(CP_{k/k-1}C^T + R)^{-1}$ is the filter gain and $P_{k/k-1} \in \mathbb{R}^{n\times n}$ is the solution of the RDE

$$P_{k+1/k} = (A - K_kC)P_{k/k-1}(A - K_kC)^T + K_kRK_k^T + Q \tag{5}$$

in which $K_k = AL_k$ is the predictor gain.

Manuscript received November 21, 2007; revised August 28, 2008. First published October 31, 2008; current version published January 06, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. James Lam.

The authors are with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Technology Court, Pullenvale QLD 4069, Australia (e-mail: garry.einicke@csiro.au; john.malos@csiro.au; david.reid@csiro.au; david.hainsworth@csiro.au).

Color versions of one or more of the figures in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2008.2007090

1053-587X/$25.00 © 2008 Australian Crown Copyright

The optimality of (3)–(5) is reliant on A, C, Q and R being known precisely. Iterative procedures will be described subsequently for estimating the noise covariances. Let Q_i and R_i denote the $i\text{th}$ estimates of Q and R respectively. A design RDE is then given by

$$\begin{aligned} P_{i,k+1/k} &= (A - K_{i,k}C)P_{i,k/k-1}(A - K_{i,k}C)^T \\ &\quad + K_{i,k}RK_{i,k}^T + Q_i \qquad (6) \\ &= (A - K_{i,k}C)P_{i,k/k-1}(A - K_{i,k}C)^T \\ &\quad \times K_{i,k}RK_{i,k}^T + Q + \delta_{i,k} \qquad (7) \end{aligned}$$

where $K_{i,k} = AP_{i,k/k-1}C^T(CP_{k/k-1}C^T + R_i)^{-1}$ and $\delta_{i,k} = Q_i - Q + K_{i,k}(R_i - R)K_{i,k}^T$. Suppose that a Kalman filter is designed with (6), using the estimates Q_i and R_i. Let $\tilde{x}_{i,k+1/k} = x_{k+1} - \hat{x}_{i,k+1/k}$ denote the predicted state error at iteration i and time k. Subtracting (4) from (1) yields

$$\begin{aligned} \tilde{x}_{i,k+1/k} &= Ax_k + w_k - A\hat{x}_{i,k/k-1} \\ &\quad - K_{i,k}(C\hat{x}_{i,k/k-1} - v_k) \\ &= (A - K_{i,k}C)\tilde{x}_{i,k/k-1} - K_{i,k}v_k + w_k. \qquad (8) \end{aligned}$$

Similarly, the recursion for the corrected state error is

$$\tilde{x}_{i,k/k} = (I - L_{i,k}C)\tilde{x}_{i,k/k-1} - L_{i,k}v_k \qquad (9)$$

where $\tilde{x}_{i,k/k} = x_k - \hat{x}_{i,k/k}$ and $L_{i,k} = P_{i,k/k-1}C^T(CP_{i,k/k-1}C^T + R)^{-1}$ is the filter gain at iteration i. The observed corrected error covariance is calculated from $\Sigma_{i,k/k} = E\{\tilde{x}_{i,k/k}, \tilde{x}_{i,k/k}^T\}$ and (9) as

$$\begin{aligned} \Sigma_{i,k/k} &= (I - L_{i,k}C)\Sigma_{i,k/k-1}(I - L_{i,k}C)^T + L_{i,k}RL_{i,k}^T \\ &= \Sigma_{i,k/k-1} - \Sigma_{i,k/k-1}C^T \\ &\quad \times (C\Sigma_{i,k/k-1}C^T + R)^{-1}C\Sigma_{i,k/k-1}. \qquad (10) \end{aligned}$$

Note that (8) can be written as $\tilde{x}_{i,k+1/k} = A\tilde{x}_{i,k/k} + w_k$, so the observed predicted error covariance $\Sigma_{i,k/k} = E\{\tilde{x}_{i,k/k}, \hat{x}_{i,k/k}^T\}$ is given by

$$\Sigma_{i,k+1/k} = A\Sigma_{i,k/k}A^T + Q. \qquad (11)$$

In the ensuing discussion, the matrix inequality $X \geq Y$ means $X - Y \geq 0$, i.e., the matrix $X - Y$ is positive semi-definite and has all its eigenvalues greater than or equal to zero. The solutions of (6) are monotonically dependent on $J\Gamma_i$ where

$$\Gamma_i = \begin{bmatrix} A & -C^TR_i^{-1}C \\ -Q_i & -A^T \end{bmatrix}$$

is the Hamiltonian matrix corresponding to (6) and

$$J = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}$$

in which I is the identity matrix (see [17]–[21]). Let Γ_{i+1} be associated with a second design RDE

$$\begin{aligned} P_{i+1,k+1/k} &= (A - K_{i+1,k}C)P_{i+1,k/k-1} \\ &\quad \times (A - K_{i+1,k}C)^T \\ &\quad + K_{i+1,k}R_{i+1}K_{i+1,k}^T + Q_{i+1} \qquad (12) \end{aligned}$$

in which Q_{i+1} and R_{i+1} denote the $(i+1)\text{th}$ estimates of Q and R respectively. From the Riccati comparison results in [17]–[21], if $J\Gamma_i \geq J\Gamma_{i+1}$, i.e., if

$$\begin{bmatrix} Q_i & A^T \\ A & -C^TR_i^{-1}C \end{bmatrix} \geq \begin{bmatrix} Q_{i+1} & A^T \\ A & -C^TR_i^{-1}C \end{bmatrix}$$

and the design RDE (6) is suitably initialized then its solutions will be monotonic. Conditions for the monotonicity of the Kalman filter design error covariance and observed error covariance are now specified formally below. These conditions are used to establish the convergence of the EM algorithms described in Section III.

Lemma 2.1: Suppose the following:

i) the data z_k has been generated by the model (1)–(2) in which A is known and its eigenvalues are inside the unit circle, C is known and the pair $(A,\ C)$ is observable;

ii) there exist estimates satisfying $R_i \geq R$ and $Q_i \geq Q$ for $i \geq 1$.

Then:

i) $P_{i,k+1/k} \geq \Sigma_{i,k+1/k}$;

ii) $P_{i,k/k} \geq \Sigma_{i,k/k}$;

iii) $R_{i+1} \geq R_i$, $Q_{i+1} \geq Q_i$, $P_{i+1,1/0} \geq P_{i,1/0} => P_{i+1,k+1/k} \geq P_{i,k+1/k}$ (and equivalently $R_i \geq R_{i+1}$, $Q_i \geq Q_{i+1}$, $P_{i,1} \geq P_{i+1,1} => P_{i,k} \geq P_{i+1,k}$);

$\forall i \geq 1$.

Proof: See Appendix I.

Thus, the sequence of observed prediction and correction error covariances is bounded above by the design prediction and correction error covariances, which depend monotonically on δ_i. Next, it is argued that the sequence of the observed prediction and correction error covariances also depend monotonically on δ_i.

Lemma 2.2: Under the conditions of Lemma 2.1:

i) $R_{i+1} \geq R_i, Q_{i+1} \geq Q_i => \Sigma_{i+1,k+1/k} \geq \Sigma_{i,k+1/k}$ (and equivalently $R_i \geq R_{i+1}$, $Q_i \geq Q_{i+1} => \Sigma_{i,k+1/k} \geq \Sigma_{i+1,k+1/k}$);

ii) $R_{i+1} \geq R_i, Q_{i+1} \geq Q_i => \Sigma_{i+1,k/k} \geq \Sigma_{i,k/k}$ (and equivalently $R_i \geq R_{i+1}, Q_i \geq Q_{i+1} => \Sigma_{i,k/k} \geq \Sigma_{i+1,k/k}$).

Proof: See Appendix II.

III. Iterative Parameter Estimation

A. Estimation of Measurement Noise Variances

This section describes the application of an EM algorithm (see [6]–[12]) to iteratively estimate the measurement noise variances. In respect of (1)– (2), assume that $v_k \in \mathbb{R}^p$ consists of independent, zero-mean, white Gaussian, measurement noise sequences. Then (2) may be written as

$$\begin{bmatrix} z_{1,k} \\ z_{2,k} \\ \vdots \\ z_{p,k} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} x_k + \begin{bmatrix} v_{1,k} \\ v_{2,k} \\ \vdots \\ v_{p,k} \end{bmatrix}$$

where c_j, $j = 1, \ldots, p$, refers to the $j\text{th}$ row of C. Denote the actual measurement noise covariance by

$$R = \begin{bmatrix} r_1 & 0 & \ldots & 0 \\ 0 & r_2 & \ldots & 0 \\ 0 & \vdots & \ddots & 0 \\ 0 & 0 & 0 & r_p \end{bmatrix}$$

where $r_j = E\{v_jv_j^T\} \in \mathbb{R}$. From the approach of [22], it is assumed that $z_{j,k} \sim N(c_jx_k, r_j)$, i.e., the probability density function of $z_{i,j}$ is $p(z_{j,k}) = (1/(2\pi r_j)^{N/2})\exp\{-(1/2r_j)\sum_{k=1}^{N}(z_{j,k} - c_jx_k)^2\}$. By setting $\partial \log_e p(z_{j,k})/\partial r_j = 0$, it is straightforward to show that an unbiased MLE for the $j\text{th}$ measurement noise variance is given by

$$r_j = \frac{1}{N-1}\sum_{k=1}^{N}(z_{j,k} - c_jx_k)^2. \qquad (13)$$

Denote the estimated measurement noise covariance by

$$R_i = \begin{bmatrix} r_{i,1} & 0 & \dots & 0 \\ 0 & r_{i,2} & \dots & 0 \\ 0 & \vdots & \ddots & 0 \\ 0 & 0 & 0 & r_{i,p} \end{bmatrix}$$

where $r_{i,j}$ is the $i\text{th}$ estimate of r_j. Suppose that a Kalman filter designed with R_i has produced corrected state estimates, which are denoted by $\hat{x}_{i,k/k}$. Let $\hat{x}_{i,j,k/k}$ denote the $j\text{th}$ component of $\hat{x}_{i,k/k}$. An EM algorithm for iteratively re-estimating R_i arises by a finite repetition of the following two-step procedure.

Step 1) Use a Kalman filter designed with R_i to calculate corrected state estimates $\hat{x}_{i,k/k}$.

Step 2) For $j = 1, \dots, p$, use $\hat{x}_{i,j,k/k}$ within (13) to obtain R_{i+1}.

It is shown below that if the error covariance and measurement noise variance estimates are initialized appropriately then the sequence of subsequent estimates will be monotonically nonincreasing.

Lemma 3.1: In respect of the above EM algorithm for estimating R, suppose for $j = 1, \dots, p$ and an $i = 1$ the following:

i) A is known and its eigenvalues are inside the unit circle;
ii) C is known and the pair $(A,\ C)$ is observable;
iii) a $Q_i \geq Q$ has been selected;
iv) some $r_{i,j} \geq r_j$ have been selected.

Then:

i) $P_{i+1,k+1/k} \leq P_{i,k+1/k}$;
ii) $R_{i+1} \leq R_i$;

$\forall i \geq 1$.

Proof: See Appendix III.

It is known (e.g., see [11]) that when the estimation problem is dominated by measurement noise, that is, when the ratio of the measurement noise to the process noise intensities is large, the measurement noise variance iterations converge to the actual value.

Lemma 3.2: Under the conditions of Lemma 3.1, additionally suppose that C is diagonal, Q and R^{-1} approach the zero matrix, then

$$\underset{Q\to 0, R_i^{-1}\to 0, i\to\infty}{Lim} R_i = R. \tag{14}$$

Proof: See Appendix IV.

B. Estimation of Process Noise Variances

In respect of (1) and (2), assume that $w_k \in \mathbb{R}^n$ consists of independent, zero-mean, white Gaussian, measurement noise sequences. Let $x_{j,k+1}$ denote the $j\text{th}$ row of x_{k+1}, then (1) may be written as

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ \vdots \\ x_{n,k+1} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} x_k + \begin{bmatrix} w_{1,k} \\ w_{2,k} \\ \vdots \\ w_{n,k} \end{bmatrix}$$

where a_j, $j = 1, \dots, n$ refers to the $j\text{th}$ row of A. Denote the actual process noise covariance by

$$Q = \begin{bmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ 0 & \vdots & \ddots & 0 \\ 0 & 0 & 0 & q_n \end{bmatrix}$$

where $q_j = E\{w_j w_j^T\} \in \mathbb{R}$. From the approach of [22], it is assumed that $x_{j,k+1} \sim N(a_j x_k, q_j)$, i.e., the probability density function of $x_{j,k+1}$ is $p(x_{j,k+1}) = (1/(2\pi q_j)^{N/2}) \exp\{-(1/2q_j)\sum_{k=1}^{N-1}(x_{j,k+1} - a_j x_k)^2\}$. By setting $(\partial p(x_{j,k+1})/\partial q_j) = 0$, it is straightforward to show that an unbiased MLE for the $j\text{th}$ process noise variance is given by

$$q_j = \frac{1}{N-2}\sum_{k=1}^{N-1}(x_{j,k+1} - a_j x_k)^2. \tag{15}$$

Denote the estimated process noise covariance by

$$Q_i = \begin{bmatrix} q_{i,1} & 0 & \dots & 0 \\ 0 & q_{i,2} & \dots & 0 \\ 0 & \vdots & \ddots & 0 \\ 0 & 0 & 0 & q_{i,p} \end{bmatrix}$$

where $q_{i,j}$ is the $i\text{th}$ estimate of q_j. Suppose at iteration i that a Kalman filter designed with Q_i has produced corrected state estimates, which are denoted by $\hat{x}_{i,k/k}$. Let $\hat{x}_{i,j,k/k}$ denote the $j\text{th}$ row of $\hat{x}_{i,k/k}$. An EM algorithm for iteratively estimating Q arises by repeating the following two-step procedure.

Step 1) Use a Kalman filter designed with Q_i to calculate corrected state estimates $\hat{x}_{i,k/k}$.

Step 2) For $j = 1, \dots, n$, use $\hat{x}_{i,j,k/k}$ and $\hat{x}_{i,j,k+1/k+1}$ within (15) to obtain Q_{i+1}.

It is shown below that if the process noise variance estimate is initialized appropriately then the sequence of subsequent estimates will be monotonically nonincreasing.

Lemma 3.3: In respect of the above EM algorithm for estimating Q, suppose for $j = 1, \dots, n$ and an $i = 1$ the following:

i) A is known and its eigenvalues are inside the unit circle;
ii) C is known and the pair $(A,\ C)$ is observable;
iii) an $R_i \geq R$ has been selected;
iv) some $q_{i,j} \geq q_j$ have been selected.

Then:

i) $P_{i+1,k+1/k} \leq P_{i,k+1/k}$;
ii) $Q_{i+1} \leq Q_i$;

$\forall i \geq 1$.

Proof: See Appendix V.

It is known that when the ratio of the process noise to the measurement noise intensities is large, the states are reconstructed exactly [11], in which case the process noise variance iterations converge to the actual value.

Lemma 3.4: Under the conditions of Lemma 3.3, additionally suppose that C is diagonal and R approaches the zero matrix, then

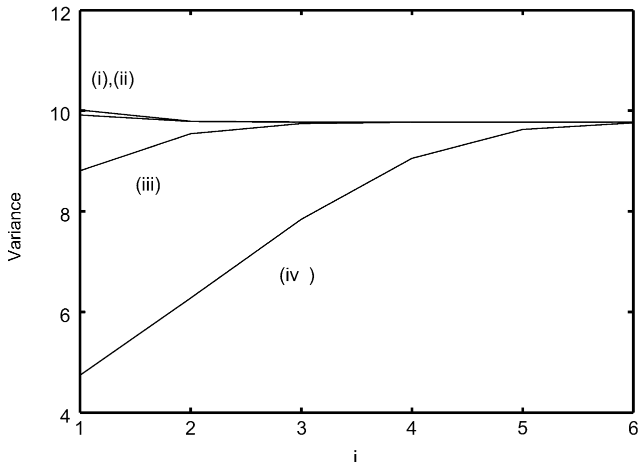
$$\underset{R\to 0, Q^{-1}\to 0, i\to\infty}{Lim} Q_i = Q. \tag{16}$$

Proof: See Appendix VI.

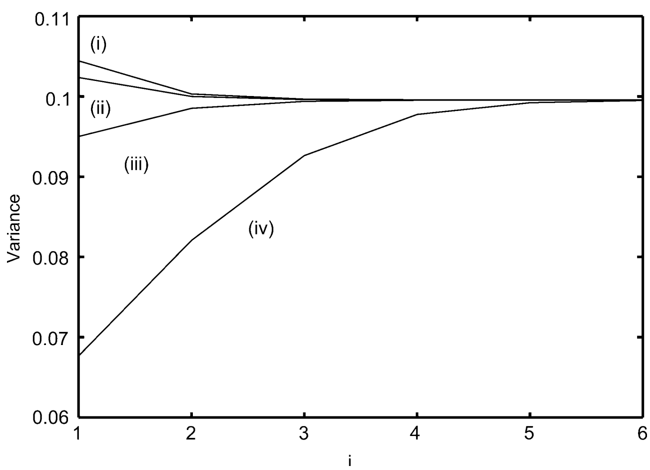
C. Example

In respect of the model (1) and (2), assume that $a = 0.9$, $b = 1$, $c = 1$ and $\sigma_w^2 = 0.1$ are known. Suppose that $\sigma_v^2 = 10$ but is unknown. The EM algorithm of Section III-A was used to jointly estimate the states within (1) and the unknown variance. Some calculated EM algorithm variance iterations, initialized with $\hat{\sigma}_v^2 = 14$ and 12, are indicated by traces (i) and (ii) of Fig. 1(a), respectively. It can be seen for this example that the sequence of variance estimates are monotonically nonincreasing and converge to $\hat{\sigma}_v^2 = 9.6$ by the second iteration.

An alternative to the EM algorithm involves calculating maximum likelihood estimates within Newton-Raphson iterations [14]. The calculated Newton–Raphson variance iterations, initialized with $\hat{\sigma}_v^2 = 14$ and 12, are indicated by traces (iii) and (iv) of Fig. 1(a), respectively. It can be seen for this example that the Newton–Raphson estimates converge by the sixth iteration.



(a)



(b)

Fig. 1. (a). Measurement noise variance estimate versus iteration number: (i) EM algorithm initialized with 14, (ii) EM algorithm initialized with 12, (iii) Newton–Raphson method initialized with 14, and (iv) Newton–Raphson method initialized with 12. (b). Process noise variance estimate versus iteration number: (i) EM algorithm initialized with 0.14, (ii) EM algorithm initialized with 0.12, (iii) Newton–Raphson method initialized with 0.14, and (iv) Newton–Raphson method initialized with 0.12.

Suppose instead that $\sigma_v^2 = 0.01$ is known, $\sigma_w^2 = 0.1$ but is unknown. The EM algorithm of Section III-B and the Newton–Raphson method [14] were used to jointly estimate the states and the unknown variance. Some example variance iterations initialized with $\hat{\sigma}_w^2 = 0.14$ and 0.12 are shown in Fig. 1(b). The figure demonstrates that the sequence of EM algorithm estimates are monotonically nonincreasing and converge to $\hat{\sigma}_w^2 = 0.098$ by the second iteration, whereas the Newton–Raphson method exhibits a comparatively lower convergence rate.

IV. Inertial Navigation Application

Our team is engaged in developing inertial navigation systems to automate longwall shearers (see Fig. 2) within underground coal mines. Inertial navigation systems are used to measure the coal face profile in three dimensional space. This information is used to keep the face straight, on track and in the seam. Strapdown inertial navigation systems (see [1]–[5]) possess three axis accelerometer and gyro sensor assemblies. The sensor data is used to calculate estimates of the instantaneous orientation, velocity and position of a mobile platform. The modeling of orientation can be undertaken either by direction cosine matrices [1]–[4] or by quaternions [5]. Our development is based on the approach of [1] and employs direction cosine matrices and a tilt vector.

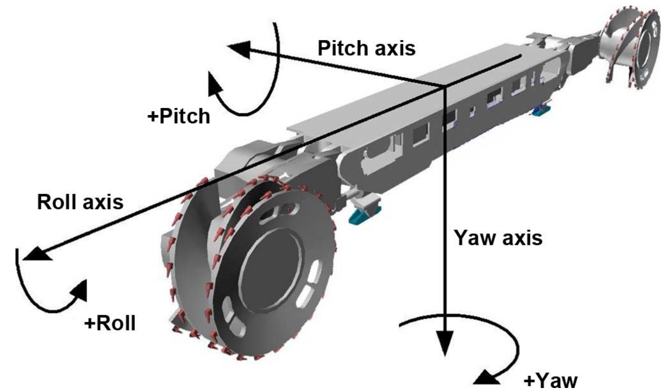


Fig. 2. Underground coal mine longwall shearer Euler angles.

A rotation of a body in three dimensional space can be represented by a simple rotation matrix for each Euler angle, namely yaw, pitch and roll, which are shown in Fig. 2. A direction cosine matrix is the product of these three rotation matrices. A standard calculation can be applied to transform the direction cosine matrix into a three dimensional tilt vector which is also known as the orientation vector (see [23]–[25]).

Alignment is the process of estimating the Earth rotation rate and rotating the attitude direction cosine matrix, so that it transforms the body-frame sensor signals to a locally level frame, wherein certain components of accelerations and velocities approach zero when the platform is stationary. This can be achieved via an alignment Kalman filter using the model

$$x_{k+1} = Ax_k + u_k \tag{17}$$

where, $x_k^T = [\delta\omega_{X,k}, \gamma_{X,k}, \delta v_{X,k}, \delta r_{X,k}]^T$, $\delta\omega_{X,k}\gamma_{X,k}$, $\delta v_{X,k}$ and $\delta r_{X,k} \in \mathbb{R}$ are the x components of the error in earth rotation rate, tilt, velocity and position vectors respectively, and $\mu_k \in \mathbb{R}^4$ is a deterministic signal which is a nonlinear function of the states (see [1]). The state transition matrix is given by $A = I + \Phi T_s + 1/2!(\Phi T_s)^2 + 1/3!(\Phi T_s)^3$, where T_s is the sampling period and

$$\Phi = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & g & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

is the continuous-time transition matrix, in which g is the universal gravitational constant. The output mapping within (2) is $C = [0 \quad 0 \quad 0 \quad 1]$. It is demonstrated below that the EM algorithm described in Section III-B can be used to estimate the unknown Q from measured data.

Raw three axis accelerometer and gyro data was recorded from a stationary Litton LN270 Inertial Navigation System at a 500-Hz data rate. From the conditions of Lemma 3.3, the initial parameter estimates and RDE solution need to be larger than the steady state values. That is, selecting arbitrarily large initial values will suffice. However, in order to generate a compact plot, the initial estimates were selected to be ten times the steady-state values. The diagonal components of Q, normalized by their value after ten iterations, are shown in Fig. 3. The figure demonstrates that approximate MLEs (15) can approach steady-state values from above, which is consistent with Lemma 3.3.

The estimated Earth rotation rate magnitude versus time is shown in Fig. 4(a). At 100 s, the estimated magnitude of the of the Earth rate is 72.53 μrad/s, that is, one revolution every 24.06 h. This estimated

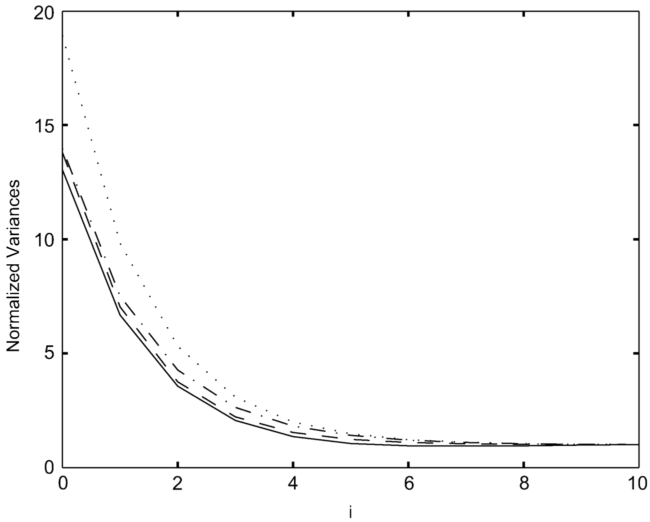
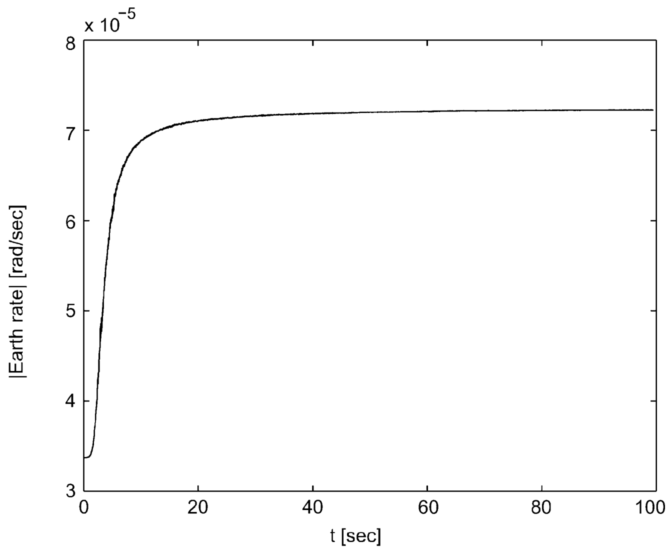
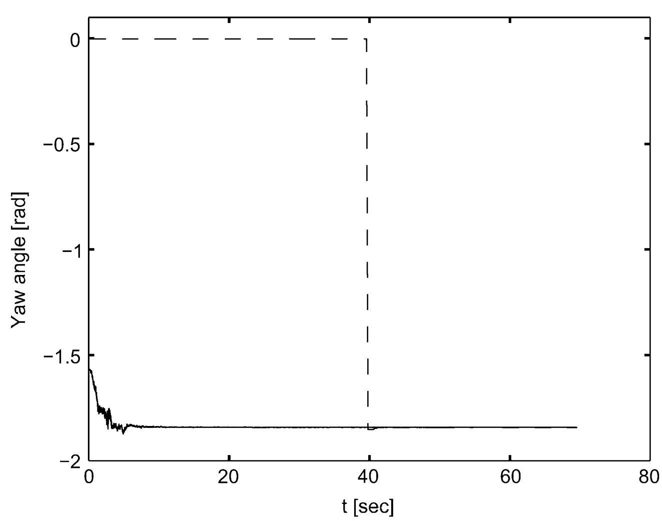


Fig. 3. Normalized (1,1) component of Q_i (solid line), normalized (2,2) component of Q_i (dashed line), normalized (3,3) component of Q_i (dot-dashed line) and normalized (3,3) component of Q_i (dotted line).



(a)



(b)

Fig. 4. (a). Estimated magnitude of Earth rotation rate. (b). Estimated yaw angle (solid line) and LN270 reported yaw angle (dashed line).

Earth rate is about 0.5% in error compared with the mean sidereal day of 23.93 h [26].

A comparison of the calculated yaw angle and that reported by the LN270 is shown in Fig. 4(b). It can be seen that the estimated yaw angle (indicated by the solid line) agrees with the yaw angle reported by the LN270 after 40 seconds (indicated by the dashed line). Since the estimated Earth rate and yaw angle are in reasonable agreement, it is suggested that the MLEs for the unknown Q are satisfactory.

V. Conclusion

This convergence of variance MLEs within an EM algorithm is investigated. The following is established:

i) the sequence of observed error covariances depend monotonically on the maximum-likelihood variance estimates;
ii) the maximum-likelihood variance estimates depend monotonically on the observed error covariances;
iii) when the process noise becomes negligible, the MLEs of the measurement noise variances asymptotically approach the actual values;
iv) when the measurement noise becomes negligible, the MLEs of the process noise variances asymptotically approach the actual values.

An illustration is provided by an inertial navigation application, in which performance is reliant on accurate variance estimates.

Appendix I
Proof of Lemma 2.1

i) If $R_i = R$ and $Q_i = Q$ for $i \geq 1$ so that $\delta_{1,k} = 0$ then $P_{i,k/k-1} = \Sigma_{i,k/k-1} \; \forall \, k \geq 0$. Subtracting (11) from (7) yields

$$\begin{aligned} &P_{i,k+1/k} - \Sigma_{i,k+1/k} \\ &\quad = (A - K_{i,k}C)(P_{i,k/k-1} - \Sigma_{i,k/k-1}) \\ &\qquad \times (A - K_{i,k}C)^T + \delta_{i,k}. \end{aligned} \tag{18}$$

Since $\delta_{i,k} = Q_i - Q + K_{i,k}(R_i - R)K_{i,k}^T \geq 0$, it follows by induction that $P_{i,k+1/k} - \Sigma_{i,k+1/k} \geq 0$, provided that $P_{i,k/k-1} - \Sigma_{i,k/k-1} \geq 0$.

ii) The result is immediate by considering $A = I$ within the proof for i).

iii) Conditions i) and ii) satisfy $J\Gamma_i \geq J\Gamma_{i+1}$ and the result follows from [1, Cor. 2.1]. ■

Appendix II
Proof of Lemma 2.2

i) To establish that the solution of $\Sigma_{i,k+1/k} = (A - K_{i,k}C)\Sigma_{i,k/k-1}(A - K_{i,k}C)^T + K_{i,k}RK_{i,k}^T + Q$ is monotonically nondecreasing, from [21, Cor. 2.1], it is required to show that

$$\begin{bmatrix} Q + K_{i+1,k}RK_{i+1,k}^T & A - K_{i+1,k}C \\ (A - K_{i+1,k}C)^T & 0 \end{bmatrix} \geq \begin{bmatrix} Q + K_{i,k}RK_{i,k}^T & A - K_{i,k}C \\ (A - K_{i,k}C)^T & 0 \end{bmatrix}.$$

Since A, Q and R are assumed to be time-invariant, it suffices to show that

$$\begin{bmatrix} L_{i+1,k}L_{i+1,k}^T & I - L_{i+1,k}C \\ (I - L_{i+1,k}C)^T & 0 \end{bmatrix} \geq \begin{bmatrix} L_{i,k}L_{i,k}^T & I - L_{i,k}C \\ (I - L_{i,k}C)^T & 0 \end{bmatrix}. \tag{19}$$

Note for an X and Y satisfying $1 \geq Y \geq X \geq 0$ that $YY^T - XX^T \geq (1-Y)(1-Y)^T - (1-X)(1-X)^T$. Therefore, $P_{i+1,k+1/k} \geq P_{i,k+1/k}$ (from Lemma 2.1), together with $R_{i+1} \geq R_i$ imply that $1 \geq L_iC \geq L_{i+1}C$ and the result (19) follows.

ii) The result is immediate by considering $A = I$ within the proof for i). ∎

Appendix III
Proof of Lemma 3.1

i) The result follows immediately from iii) of Lemma 2.1, which is a consequence of $J\Gamma_i \geq J\Gamma_{i+1}$ (see [17]–[21]).

ii) Let $\tilde{x}_{i,k} = x_k - \hat{x}_{i,k/k}$ denote the corrected state error at iteration i and time k, which is orthogonal to v_k. Then Step i) results in $r_{i+1,j} = c_j\Sigma_{i,k/k}c_j^T + r_{i,j}$, which implies $R_{i+1} = C\Sigma_{i,k/k}C^T + R$, where $\Sigma_{i,k/k} = E\{\tilde{x}_{i,k/k}, \tilde{x}_{i,k/k}^T\}$. Since R_{i+1} is affine to $\Sigma_{i,k/k}$, which (from Lemma 2.2) is monotonically nonincreasing, then $R_{i+1} \leq R_i$. ∎

Appendix IV
Proof of Lemma 3.2

From the proof of [11, Prop. 1], it is straightforward to show that $\underset{Q_i \to 0, R_i^{-1} \to 0}{Lim} L_iC^T = 0$, which together with the observation $\underset{Q \to 0, R^{-1} \to 0}{Lim} E\{zz^T\} = R$ yield (14) since the variance MLE (13) is unbiased. ∎

Appendix V
Proof of Lemma 2.3

i) The result follows immediately from Lemma 2.1.

ii) Define $\tilde{x}_{i,k+1/k} = x_k - \hat{x}_{i,k/k+1}$ to be the state prediction error at iteration i, which is orthogonal to v_k. Denote the $j\text{th}$ component of $\tilde{x}_{i,k+1/k}$ by $\tilde{x}_{i,j,k+1/k}$. Let $L_{i,j,k+1}$ denote the $j\text{th}$ row of the filter gain at iteration i and time $k+1$. Substituting $\hat{x}_{i,j,k+1/k+1} - a_{j,k}\hat{x}_{i,j,k/k} = L_{i,j,k+1}(z_{j,k+1} - c_j\hat{x}_{i,j,k+1/k}) = L_{i,j,k+1}(c_j\tilde{x}_{i,j,k+1/k} + v_{j,k+1})$ into (15) leads to $Q_{i+1} = L_{i,k+1}(C\Sigma_{i,k+1/k}C^T + R)L_{i,k+1}^T$, where $\Sigma_{i,k+1/k} = E\{\tilde{x}_{i,k+1/k}, \tilde{x}_{i,k+1/k}^T\}$. Since Q_{i+1} is affine to $\Sigma_{i,k+1/k}$, which (from Lemma 2.2) is monotonically nonincreasing, then $Q_{i+1} \leq Q_i$. ∎

Appendix VI
Proof of Lemma 3.4

From the proof of [11, Prop. 1], it is straightforward to show that $\underset{R_i \to 0, Q_i^{-1} \to 0}{Lim} L_{i,k}C^T = I$ and the ∞ -norm of the output estimation error is given by the trace of R, i.e., the filtered states approach the actual states. In this case, the result (16) follows since the variance MLE (15) is unbiased. ∎

References

[1] R. P. Savage, *Strapdown Analytics*. Maple Plain, MN: Strapdown Associates, 2000, vol. 2, ch. 15, pp. 15.1–15.142.

[2] M. S. Grewal, V. D. Henderson, and R. S. Miyasako, "Application of Kalman filtering to the calibration and alignment of inertial navigation systems," *IEEE Trans. Autom. Control*, vol. 36, no. 1, pp. 4–13, Jan. 1991.

[3] Z. Chuanbin, T. Weifeng, and J. Zhihua, "A novel method improving the alignment accuracy of a strapdown inertial navigation system on a stationary base," *Meas. Sci. Technol.*, vol. 15, pp. 765–769, 2004.

[4] J. Zhou, J. Yuan, X. Yue, and J. Luo, "A new approach for sins stationary self-alignment," *Coordinates*, pp. 24–27, Feb. 2007.

[5] Y. Wu, X. Hu, M. Wu, and D. Hu, "Strapdown inertial navigation using dual quaternion algebra: Error analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 259–266, Jan. 2006.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[7] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, Mar. 1983.

[8] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Series Anal.*, vol. 3, no. 4, pp. 253–264, 1982.

[9] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Signal Process.*, vol. 36, no. 4, pp. 477–489, Apr. 1988.

[10] T. Y. Al-Naffouri, "An EM-based forward-backward Kalman filter for the estimation of time-variant channels in OFDM," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3924–3930, Jul. 2007.

[11] G. A. Einicke, "Optimal and robust noncausal filter formulations," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1069–1077, Mar. 2006.

[12] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.

[13] R. K. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Autom. Control*, vol. 15, pp. 175–184, Apr. 1970.

[14] R. K. Mehra, "Approaches to adaptive filtering," *IEEE Trans. Autom. Control*, vol. 17, pp. 693–698, Oct. 1972.

[15] B. Carew and P. R. Belanger, "Identification of optimal filter steady-sate gain for systems with unknown noise covariance," *IEEE Trans. Autom. Control*, vol. 18, pp. 582–587, Dec. 1973.

[16] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[17] H. K. Wimmer, "Monotonicity and maximality of solutions of discrete-time algebraic Riccati equations," *J. Math. Syst. Estim. Control* , vol. 2, no. 2, pp. 219–235, 1992.

[18] H. K. Wimmer and M. Pavon, "A comparison theorem for matrix Riccati difference equations," *Syst. Control Lett.*, vol. 19, pp. 233–239, 1992.

[19] G. Freiling, G. Jank, and H. Abou-Kandil, "Generalized Riccati difference and differential equations," *Lin. Algebra Its Appl.*, pp. 291–303, 1996.

[20] G. Freiling and G. Jank, "Existence and comparison theorems for algebraic Riccati equations and Riccati differential and difference equations," *J. Dynam. Control Syst.*, vol. 2, no. 4, pp. 529–547, 1996.

[21] G. A. Einicke, "Asymptotic optimality of the minimum-variance fixed-interval smoother," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1543–1547, Apr. 2007.

[22] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993, pp. 157–204.

[23] J. E. Bortz, "A new mathematical formulation for strapdown inertial navigation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 7, no. 1, pp. 61–66, Jan. 1971.

[24] Y. F. Jiang and Y. P. Lin, "On the rotation vector differential equation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, no. 1, pp. 181–183, Jan. 1991.

[25] M. D. Shuster, "The kinematic equation for the rotation vector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 263–267, Jan. 1993.

[26] *Explanatory Supplement to the Astronomical Almanac*, P. K. Seidelmann, Ed. Mill Valley, CA: Univ. Science Books, 1992, pp. 52–698.