AAS 08-219

AN EXPECTATION-MAXIMIZATION APPROACH TO ATTITUDE SENSOR CALIBRATION

Yang Cheng* **and John L. Crassidis**†

An Expectation-Maximization approach to sensor calibration is presented and applied to three-axis-magnetometer calibration. This approach is different from the existing attitude-independent approaches mainly in how the attitude parameters in the attitude sensor measurement model are handled. The attitude-independent approaches rely on a conversion of the body and reference representations of the Earth's magnetic field vector into an attitude-independent scalar observation based on scalar checking. The Expectation-Maximization (EM) algorithm essentially maximizes the expectation of the complete likelihood function with respect to the probability distribution of the attitude parameters. Comparisons of the EM algorithm and the attitude-independent approach using simulated magnetometer data are presented.

Spacecraft attitude estimation is the process of inferring the orientation of the spacecraft with respect to a reference frame from attitude sensor measurements. Typical spacecraft attitude sensors include three-axis magnetometers, sun sensors, and star trackers. Gyros are also widely used in spacecraft attitude determination systems to provide information about the inertial angular velocity. The attitude sensor measurements are related to the spacecraft attitude by the sensor measurement model. In order to achieve high-precision attitude estimation, it is of paramount importance to make the sensor measurement model consistent with the actual sensor measurement data by sensor calibration. The structure of the sensor measurement model is usually assumed to be fixed and known. So, what needs to be calibrated is a set of sensor parameters of the model. The optimal values of these parameters are usually obtained by solving a parameter estimation problem. One of the most widely used methods for parameter estimation is maximum likelihood estimation. A common issue in sensor calibration is that the attitude parameters in the sensor measurement model may be unknown. Calibration methods differ from one another in how the attitude parameters are treated and how much attitude knowledge is utilized.

Three-axis magnetometers are widely used for onboard operations of low Earth orbiting spacecraft because they provide both the direction and magnitude of the Earth's magnetic field, are lightweight and reliable, have low-power requirements, and have no moving parts. The calibration parameters of three-axis magnetometers include biases, scale factors, and nonorthogonality corrections. Scale factors and nonorthogonality corrections occur because the individual magnetometer axes are not orthonormal, typically due to thermal gradients within the magnetometer or to mechanical stress from the spacecraft. The problem of magnetometer calibration has been studied in Refs. **?**, **?**, **?**, **?** and the references therein. Except for Ref. ?, all of them are iterative batch methods. All are attitude-independent methods based on a scalar checking process, in which the original

*Research Assistant Professor, Department of Mechanical and Aerospace Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260.

†Professor, Department of Mechanical and Aerospace Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260.

vector measurement model is replaced by an attitude-independent scalar measurement model. The parameter estimation criterion used is maximum likelihood estimation.

The main objective of this paper is to present two expectation-maximization (EM) algorithms for magnetometer calibration and compare them with the existing attitude-independent scalar-checking methods. The EM algorithm is an iterative algorithm used to learn parameters of statistical models in the presence of incomplete data or hidden variables.[?] A classic treatment of the EM algorithm is shown in Ref. **?**. In the proposed EM algorithms for sensor calibration, the attitude matrix is treated as hidden variables, estimated using a smoother, and integrated out in the complete likelihood function.

The organization of the remainder of the paper proceeds as follows. First, the magnetometer calibration model is presented. Then, the attitude-independent approach based on scalar checking is reviewed. Next, the basics of the EM algorithm and two EM algorithms for magnetometer calibration are presented. Finally, numerical results are presented.

MAGNETOMETER CALIBRATION MODEL

A general magnetometer model is given by

$$\mathbf{B}_k = T^T(A'_k\mathbf{H}_k + \mathbf{b}' + \boldsymbol{\epsilon}'_k) \tag{1}$$

where $\mathbf{B}_k$ is the 3×1 measurement of geomagnetic field vector by the magnetometer, $\mathbf{H}_k$ is the representation of the geomagnetic field vector with respect to an Earth-fixed reference frame, computed from the spacecraft position and the Earth magnetic field model, A'_k is the attitude matrix relating the Earth-fixed reference frame and the sensor frame, $\mathbf{b}'$ is an unknown constant bias vector in the (ideal) sensor frame, $\boldsymbol{\epsilon}'_k$ is the measurement noise vector in the sensor frame that is assumed to a zero-mean Gaussian white sequence with covariance Σ, and T is a 3×3 unknown non-singular matrix of nine degrees of freedom accounting for scale factors and nonorthogonality of the magnetometer sensing axes. The polar decomposition of T,

$$T = \mathcal{O}(I_{3\times3} + D)^{-1} \tag{2}$$

where $\mathcal{O}$ is an orthogonal matrix and $(I_{3\times3} + D)^{-1}$ is a symmetric positive-definite matrix, leads to

$$\mathbf{B}_k = (I_{3\times3} + D)^{-1}\mathcal{O}(A'_k\mathbf{H}_k + \mathbf{b}' + \boldsymbol{\epsilon}'_k) \tag{3}$$

In the above equation, the unknown calibration parameters are D, $\mathcal{O}$, and $\mathbf{b}'$. However, unless we have independent knowledge about A'_k, we cannot estimate both $\mathcal{O}$ and A'_k without ambiguity. Therefore, we rewrite the above equation as

$$\mathbf{B}_k = (I_{3\times3} + D)^{-1}(A_k\mathbf{H}_k + \mathbf{b} + \boldsymbol{\epsilon}_k) \tag{4}$$

with $A_k \triangleq \mathcal{O}A'_k$, $\mathbf{b} \triangleq \mathcal{O}\mathbf{b}'$, and $\boldsymbol{\epsilon} \triangleq \mathcal{O}\boldsymbol{\epsilon}'$, and limit our objective to estimating $\mathbf{b}$ and D, where

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \quad D = \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{12} & D_{22} & D_{23} \\ D_{13} & D_{23} & D_{33} \end{bmatrix} \tag{5}$$

Because D is symmetric, it only has six independent components, and so $\mathbf{b}$ and D have nine independent parameters in total. Let us define the 9×1 calibration vector as

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_b^T & \boldsymbol{\theta}_D^T \end{bmatrix}^T = \begin{bmatrix} b_1 & b_2 & b_3 & D_{11} & D_{22} & D_{33} & D_{12} & D_{13} & D_{23} \end{bmatrix}^T \tag{6}$$

with obvious definition of $\boldsymbol{\theta}_b$ and $\boldsymbol{\theta}_D$. The mapping between $\boldsymbol{\theta}$ and $(\mathbf{b}, D)$ is one to one. A useful relation is

$$D\mathbf{B}_k = B(\mathbf{B}_k)\boldsymbol{\theta}_D \tag{7}$$

with

$$B(\mathbf{B}_k) = \begin{bmatrix} B_{1k} & 0 & 0 & B_{2k} & B_{3k} & 0 \\ 0 & B_{2k} & 0 & B_{1k} & 0 & B_{3k} \\ 0 & 0 & B_{3k} & 0 & B_{1k} & B_{2k} \end{bmatrix} \tag{8}$$

Given the magnetometer measurement model, the likelihood function of A_k and $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta}, A_k) \triangleq p(\mathbf{B}_k | A_k, \boldsymbol{\theta}) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-J_k\right) \tag{9}$$

with

$$J_k(\boldsymbol{\theta}, A_k) = \frac{1}{2}\left[(I_{3\times 3} + D)\mathbf{B}_k - A_k\mathbf{H}_k - \mathbf{b}\right]^T \Sigma^{-1} \left[(I_{3\times 3} + D)\mathbf{B}_k - A_k\mathbf{H}_k - \mathbf{b}\right] \tag{10}$$

Throughout this paper, we assume

$$\Sigma = \sigma^2 I_{3\times 3} \tag{11}$$

If the attitude A_k is perfectly known, the maximum likelihood estimate of $\boldsymbol{\theta}$ is the least-square solution of the following linear system of equations:

$$\mathbf{z}_k = \mathcal{H}_k\boldsymbol{\theta} + \boldsymbol{\epsilon}_k \tag{12}$$

with

$$\mathcal{H}_k = \begin{bmatrix} I_{3\times 3} & -B(\mathbf{B}_k) \end{bmatrix} \tag{13}$$

$$\mathbf{z}_k = \mathbf{B}_k - A_k\mathbf{H}_k \tag{14}$$

The maximum likelihood estimation of $\boldsymbol{\theta}$ based on magnetometer measurements becomes complicated when the attitude is unknown or partially known.

ATTITUDE-INDEPENDENT APPROACH

The attitude-independent approach estimates the calibration parameters without use of any attitude knowledge. The first step of the attitude-independent approach is to eliminate the attitude matrix and introduce an effective scalar observation based on the fact that the attitude matrix preserves the norm of a vector. Equation (4) is rewritten as

$$(I_{3\times 3} + D)\mathbf{B}_k - \mathbf{b} - \boldsymbol{\epsilon}_k = A_k\mathbf{H}_k \tag{15}$$

Taking the norms of the both sides and using $\|A_k\mathbf{H}_k\| = \|\mathbf{H}_k\| = \sqrt{\mathbf{H}_k^T\mathbf{H}_k}$ leads to

$$z_k = \left[(I_{3\times 3} + D)\mathbf{B}_k - \mathbf{b}\right]^T \left[(I_{3\times 3} + D)\mathbf{B}_k - \mathbf{b}\right] - \mathbf{H}_k^T\mathbf{H}_k = \upsilon_k \tag{16}$$

where z_k is the scalar observation and υ_k is the effective noise, defined by

$$\upsilon_k \triangleq 2\left[(I_{3\times 3} + D)\mathbf{B}_k - \mathbf{b}\right]^T \boldsymbol{\epsilon}_k - \boldsymbol{\epsilon}_k^T\boldsymbol{\epsilon}_k \tag{17}$$

whose mean is

$$\mu_z = -\operatorname{tr}(\Sigma) \tag{18}$$

and whose variance is

$$\sigma_{zk}^2 = 4\left[(I_{3\times 3} + D)\mathbf{B}_k - \mathbf{b}\right]^T \Sigma \left[(I_{3\times 3} + D)\mathbf{B}_k - \mathbf{b}\right] + 2\operatorname{tr}(\Sigma^2) \tag{19}$$

The operator tr returns the trace of a matrix. Note that σ_{zk}^2 is a function of $\boldsymbol{\theta}$. Approximating $p(z_k|\boldsymbol{\theta})$ as a Gaussian distribution, we have

$$p(z_k|\boldsymbol{\theta}) \approx \frac{1}{\sqrt{2\pi\sigma_{zk}^2}} \exp\left(-\frac{(z_k - \mu_z)^2}{2\sigma_{zk}^2}\right) \tag{20}$$

It follows that the likelihood function of $\boldsymbol{\theta}$ is given by

$$L(\boldsymbol{\theta}) \approx \prod_{k=1}^{N} p(z_k|\boldsymbol{\theta}) \tag{21}$$

Subsequently, the maximum likelihood estimate of $\boldsymbol{\theta}$ in the attitude-independent approach minimizes the following cost function:

$$J^{\text{scalar}'}(\boldsymbol{\theta}) = \sum_{k=1}^{N} \left[\frac{(z_k - \mu_z)^2}{2\sigma_{zk}^2} + \frac{\log \sigma_{zk}^2}{2}\right] \tag{22}$$

of which an approximation is

$$J^{\text{scalar}}(\boldsymbol{\theta}) = \sum_{k=1}^{N} \frac{(z_k - \mu_z)^2}{2\sigma_{zk}^2} \tag{23}$$

EXPECTATION-MAXIMIZATION ALGORITHM

In this section, the basic theory of the EM algorithm is briefly reviewed first. Then, two EM algorithms for magnetometer calibration are presented. Both EM algorithms involve the estimation of the attitude. One EM algorithm estimates the attitude from the magnetometer measurements with the aid of the gyro data. The other "estimates" from the magnetometer measurements the attitude under the assumption that the attitude is totally unknown and uncorrelated at different times. It is an attitude-independent approach in essence.

Basic Theory

Suppose in a typical parameter estimation problem, $\boldsymbol{\theta}$ is the parameter of interest and $\mathcal{Y}$ is the observation data. The likelihood function of $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta}) = p(\mathcal{Y}|\boldsymbol{\theta}) \tag{24}$$

and the maximum likelihood estimate of $\boldsymbol{\theta}$ maximizes Eq. (24). The EM algorithm is a solution of maximum likelihood estimation in the presence of missing data or hidden variables. Let us assume that what is immediately available is not $p(\mathcal{Y}|\boldsymbol{\theta})$, but

$$p(\mathcal{X}, \mathcal{Y}|\boldsymbol{\theta}) = p(\mathcal{Y}|\mathcal{X}, \boldsymbol{\theta})p(\mathcal{X}) \tag{25}$$

with $\mathcal{X}$ the hidden variables. In terms of $p(\mathcal{X}, \mathcal{Y}|\boldsymbol{\theta})$, the likelihood function is defined by an integral

$$L(\boldsymbol{\theta}) = \int p(\mathcal{X}, \mathcal{Y}|\boldsymbol{\theta}) d\mathcal{X} \tag{26}$$

where the integration is taken over $\mathcal{X}$.

For any probability density $q(\mathcal{X})$, it holds that

$$\log L(\boldsymbol{\theta}) = \log \int p(\mathcal{X}, \mathcal{Y}|\boldsymbol{\theta}) d\mathcal{X} \geq \int q(\mathcal{X}) \log \frac{p(\mathcal{X}, \mathcal{Y}|\boldsymbol{\theta})}{q(\mathcal{X})} d\mathcal{X} \tag{27}$$

This inequality relates the logarithm of an integral to the integral of a logarithm, with the latter easier to handle. The theory of the EM algorithm is based on the inequality.[?] The EM algorithm is an iterative algorithm, alternating between maximizing $\int q(\mathcal{X}) \log p(\mathcal{X}, \mathcal{Y}|\boldsymbol{\theta})/q(\mathcal{X}) d\mathcal{X}$ with respect to $q(\mathcal{X})$ and with respect to $\boldsymbol{\theta}$. The density maximizing $\int q(\mathcal{X}) \log p(\mathcal{X}, \mathcal{Y}|\boldsymbol{\theta})/q(\mathcal{X}) d\mathcal{X}$ is the smoothing density of $\mathcal{X}$,[?] $q^*(\mathcal{X}) = p(\mathcal{X}|\mathcal{Y}, \boldsymbol{\theta})$.

Suppose the current estimate of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}}^{(l)}$. An iteration of the EM algorithms consists of an Expectation-Step (E-Step) and an Maximization-Step (M-Step):

1. E-Step: calculate the smoothing density of $\mathcal{X}$ given $\mathcal{Y}$ and $\hat{\boldsymbol{\theta}}^{(l)}$:

$$q^{*(l)}(\mathcal{X}) = p(\mathcal{X}|\mathcal{Y}, \hat{\boldsymbol{\theta}}^{(l)}) \tag{28}$$

2. M-Step: find out $\hat{\boldsymbol{\theta}}^{(l+1)}$ that maximizes the expectation of $\log p(\mathcal{X}, \mathcal{Y}|\boldsymbol{\theta})$ with respect to $p(\mathcal{X}|\mathcal{Y}, \hat{\boldsymbol{\theta}}^{(l)})$:

$$\hat{\boldsymbol{\theta}}^{(l+1)} = \arg\max_{\boldsymbol{\theta}} E_{p(\mathcal{X}|\mathcal{Y}, \hat{\boldsymbol{\theta}}^{(l)})}[\log p(\mathcal{X}, \mathcal{Y}|\boldsymbol{\theta})] = \arg\max_{\boldsymbol{\theta}} \int \log p(\mathcal{X}, \mathcal{Y}|\boldsymbol{\theta}) p(\mathcal{X}|\mathcal{Y}, \hat{\boldsymbol{\theta}}^{(l)}) d\mathcal{X} \tag{29}$$

Note that in the above integrand, $\hat{\boldsymbol{\theta}}^{(l)}$ is the current estimate and $\boldsymbol{\theta}$ is the maximizing argument to be sought in the M-Step. The likelihood function $L(\boldsymbol{\theta})$ increases after each iteration of the EM algorithm. Therefore, it is guaranteed that the EM algorithm will eventually reach a local minimum of the likelihood function $L(\boldsymbol{\theta})$.

Linear Gaussian Dynamic System

In this section, it is shown how to estimate constant parameters of a linear Gaussian dynamic system using the EM algorithm. More details can be found in Ref. **?**. This provides the necessary theoretical foundation for designing the EM algorithms for magnetometer calibration. A general linear dynamic system is given by

$$\mathbf{x}_{k+1} = \Phi \mathbf{x}_k + \mathbf{w}_k + \mathbf{u}_k \tag{30}$$

$$\mathbf{y}_k = H \mathbf{x}_k + \mathbf{v}_k \tag{31}$$

where $\mathbf{u}_k$ is the deterministic exogenous input and $\mathbf{w}_k$ and $\mathbf{v}_k$ are zero-mean Gaussian white noise sequences with covariances Q and R. The objective is to estimate Φ, H, Q, and R from the measurements. These matrices are assumed to be constant and characterized by $\boldsymbol{\theta}$. For this linear dynamic system, the state vectors (hidden variables) and observation vectors are denoted by

$$\mathcal{X}_N = \begin{bmatrix} \mathbf{x}_1^T & \ldots & \mathbf{x}_N^T \end{bmatrix}^T \tag{32}$$

and

$$\mathcal{Y}_N = \begin{bmatrix} \mathbf{y}_1^T & \ldots & \mathbf{y}_N^T \end{bmatrix}^T \tag{33}$$

respectively. The optimal estimate of $\boldsymbol{\theta}$ maximizes $p(\mathcal{Y}_N|\boldsymbol{\theta})$. Recall that the EM algorithm does not maximizes $p(\mathcal{Y}_N|\boldsymbol{\theta})$ directly. The two main quantities involved in the EM algorithm are $p(\mathcal{X}_N|\mathcal{Y}_N, \boldsymbol{\theta})$ and $p(\mathcal{X}_N, \mathcal{Y}_N|\boldsymbol{\theta})$, respectively. For the linear Gaussian system, $p(\mathcal{X}_N|\mathcal{Y}_N, \boldsymbol{\theta})$ is a Gaussian distribution. More importantly, its statistics (mean and covariance) are given by the Kalman smoother. The form of $\log p(\mathcal{X}_N, \mathcal{Y}_N|\boldsymbol{\theta})$ as well as $p(\mathcal{X}_N, \mathcal{Y}_N|\boldsymbol{\theta})$ is also simple, as will be shown soon.

In the ensuing, the Kalman smoother will be reviewed. In the Kalman smoother, the state estimates before measurement update are denoted by

$$\hat{\mathcal{X}}_N^- = \begin{bmatrix} (\hat{\mathbf{x}}_1^-)^T & \ldots & (\hat{\mathbf{x}}_N^-)^T \end{bmatrix}^T \tag{34}$$

the states estimates after measurement update are denoted by

$$\hat{\mathcal{X}}_N^+ = \begin{bmatrix} (\hat{\mathbf{x}}_1^+)^T & \ldots & (\hat{\mathbf{x}}_N^+)^T \end{bmatrix}^T \tag{35}$$

and the smoothed state estimates are denoted by

$$\hat{\mathcal{X}}_N^S = \begin{bmatrix} (\hat{\mathbf{x}}_1^S)^T & \ldots & (\hat{\mathbf{x}}_N^S)^T \end{bmatrix}^T \tag{36}$$

The same superscripts "$-$", "$+$", and "S" are used for the associated state error covariance matrices. The Kalman smoother, i.e. the E-Step of the EM algorithm, consists of forward recursions and backward recursions. The forward recursions are given by the Kalman filter.

- Forward Recursions

$$\hat{\mathbf{x}}_{k+1}^- = \Phi \hat{\mathbf{x}}_k^+ \tag{37}$$

$$P_{k+1}^- = \Phi P_k^+ \Phi^T + Q \tag{38}$$

$$\boldsymbol{\nu}_{k+1} = \mathbf{y}_{k+1} - H\hat{\mathbf{x}}_{k+1}^- \tag{39}$$

$$S_{k+1} = H_{k+1} P_{k+1}^- H_{k+1}^T + \Sigma \tag{40}$$

$$K_{k+1} = P_{k+1}^- H^T S_{k+1}^{-1} \tag{41}$$

$$\hat{\mathbf{x}}_{k+1}^+ = \hat{\mathbf{x}}_{k+1}^- + K_{k+1}\boldsymbol{\nu}_{k+1} \tag{42}$$

$$P_{k+1}^+ = \left[I_{n\times n} - K_{k+1}H\right] P_{k+1}^- \tag{43}$$

- Backward Recursions

$$\hat{\mathbf{x}}_N^S = \hat{\mathbf{x}}_N^+ \tag{44}$$

$$P_N^S = P_N^+ \tag{45}$$

$$\mathcal{K}_k = P_k^+ \Phi^T \left(P_{k+1}^-\right)^{-1} \tag{46}$$

$$\hat{\mathbf{x}}_k^S = \hat{\mathbf{x}}_k^+ + \mathcal{K}_k \left[\hat{\mathbf{x}}_{k+1}^S - \hat{\mathbf{x}}_{k+1}^-\right] \tag{47}$$

$$P_k^S = P_k^+ + \mathcal{K}_k \left[P_{k+1}^S - P_{k+1}^-\right] \mathcal{K}_k^T \tag{48}$$

The matrices Φ, H, Q, and R in the formulae correspond to the present estimate $\hat{\boldsymbol{\theta}}$.

For the linear Gaussian system, $\log p(\mathcal{X}_N, \mathcal{Y}_N|\boldsymbol{\theta})$ takes a simple form, given by

$$\begin{aligned}
&\log p(\mathcal{X}_N, \mathcal{Y}_N|\boldsymbol{\theta}) \\
=& -\frac{1}{2}\sum_{k=1}^{N-1}\left[\log|Q| + (\mathbf{x}_{k+1} - \Phi\mathbf{x}_k)^T Q^{-1}(\mathbf{x}_{k+1} - \Phi\mathbf{x}_k)\right] \\
& -\frac{1}{2}\sum_{k=1}^{N}\left[\log|R| + (\mathbf{y}_k - H\mathbf{x}_k)^T R^{-1}(\mathbf{y}_k - H\mathbf{x}_k)\right]
\end{aligned} \tag{49}$$

where $|\cdot|$ denotes the determinant of a matrix. Given $p(\mathcal{X}_N|\mathcal{Y}_N, \hat{\boldsymbol{\theta}})$, closed-form solutions for $\hat{\Phi}$, $\hat{H}$, $\hat{Q}$, and $\hat{R}$, which maximizes the expectation of $\log p(\mathcal{X}_N, \mathcal{Y}_N|\boldsymbol{\theta})$ with respect to $p(\mathcal{X}_N|\mathcal{Y}_N, \hat{\boldsymbol{\theta}})$ are obtained by solving

$$\frac{\partial\left[\int \log p(\mathcal{X}_N, \mathcal{Y}_N|\boldsymbol{\theta}) p(\mathcal{X}_N|\mathcal{Y}_N, \hat{\boldsymbol{\theta}}) d\mathcal{X}_N\right]}{\partial\boldsymbol{\theta}} = \mathbf{0} \tag{50}$$

Finally, a simple example is given to illustrate the process of the EM algorithm. Suppose R is the only unknown parameter in the linear Gaussian system model. The EM algorithm proceeds as follows:

1. Generate an initial guess $\hat{R}$;
2. Run the Kalman smoother assuming $R = \hat{R}$ to obtain $\hat{\mathcal{X}}_N^S$ and their associated error covariance matrices;
3. Update $\hat{R}$ by

$$\hat{R} = \sum_{k=1}^{N}\left(\mathbf{y}_k\mathbf{y}_k^T - H\hat{\mathbf{x}}_k^S\mathbf{y}_k^T - \left(H\hat{\mathbf{x}}_k^S\mathbf{y}_k^T\right)^T + H\widehat{\mathbf{x}_k\mathbf{x}_k^T}^S H^T\right) \tag{51}$$

where

$$\widehat{\mathbf{x}_k\mathbf{x}_k^T}^S = P_k^S + \hat{\mathbf{x}}_k^S(\hat{\mathbf{x}}_k^S)^T \tag{52}$$

4. Repeat steps 2 and 3 until convergence or the maximum number of iterations is reached.

If Q is the only unknown matrix, the corresponding update step is

$$\hat{Q} = \sum_{k=1}^{N-1}\left(\widehat{\mathbf{x}_{k+1}\mathbf{x}_{k+1}^T}^S - \widehat{\mathbf{x}_{k+1}\mathbf{x}_k^T}^S\Phi^T - \Phi\left(\widehat{\mathbf{x}_{k+1}\mathbf{x}_k^T}^S\right)^T + \Phi\widehat{\mathbf{x}_k\mathbf{x}_k^T}^S\Phi^T\right) \tag{53}$$

EM-BASED MAGNETOMETER CALIBRATION ALGORITHMS

Two EM algorithms for magnetometer calibration are presented in this section. The first algorithm uses the measurements of both the magnetometer and the gyro. The second algorithm uses the magnetometer measurements only. In the first algorithm, the gyro measurements are used to relate the attitude matrices at different times, and the extended Kalman smoother (EKS) is used to fuse the gyro measurements and the magnetometer measurements. In the second algorithm, the attitude matrices are assumed to be uncorrelated in time and totally random. The second algorithm is essentially *attitude-independent*. In both algorithms, the E-Step is an attitude smoother given the calibration parameter estimate and the M-Step is a linear least-square estimator for updating the calibration parameter estimate based on the smoothing result.

Magnetometer and Gyro

Gyro Model In order to make use of the gyro measurements, a model relating the gyro measurements to the angular velocity is needed. For the rate-integrating gyro, a widely used three-axis continuous-time model is given by[?]

$$\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega} + \boldsymbol{\beta} + \boldsymbol{\eta}_v \tag{54a}$$

$$\dot{\boldsymbol{\beta}} = \boldsymbol{\eta}_u \tag{54b}$$

where $\tilde{\boldsymbol{\omega}}$ and $\boldsymbol{\omega}$ are the measured and the true angular velocities, $\boldsymbol{\beta}$ is the drift, whose initial value is random and unknown, and $\boldsymbol{\eta}_v$ and $\boldsymbol{\eta}_u$ are independent zero-mean Gaussian white-noise processes with

$$E\left\{\boldsymbol{\eta}_v(t)\boldsymbol{\eta}_v^T(\tau)\right\} = \sigma_v^2\delta(t-\tau)I_{3\times 3} \tag{55a}$$

$$E\left\{\boldsymbol{\eta}_u(t)\boldsymbol{\eta}_u^T(\tau)\right\} = \sigma_u^2\delta(t-\tau)I_{3\times 3} \tag{55b}$$

where $E\{\cdot\}$ denotes expectation, $\delta(t-\tau)$ is the Dirac delta function, and σ_v^2 and σ_u^2 are the spectral densities of the white noise processes.

Attitude Quaternion The attitude quaternion is a widely-used attitude representation in attitude estimation. It is defined by

$$\mathbf{q} = \begin{bmatrix} \boldsymbol{\rho} \\ q_4 \end{bmatrix} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} \tag{56}$$

where $\boldsymbol{\rho}$ is the vector part of the attitude quaternion and q_4 is the scalar part of the attitude quaternion. The attitude quaternion obeys the unity-norm constraint, i.e., $\mathbf{q}^T\mathbf{q} = 1$. The expression for the attitude matrix in terms of the attitude quaternion is given by

$$A(\mathbf{q}) = \Xi^T(\mathbf{q})\Psi(\mathbf{q}) \tag{57}$$

where

$$\Xi(\mathbf{q}) = \begin{bmatrix} q_4 I_{3\times 3} + [\boldsymbol{\rho}\times] \\ -\boldsymbol{\rho}^T \end{bmatrix} \tag{58}$$

$$\Psi(\mathbf{q}) = \begin{bmatrix} q_4 I_{3\times 3} - [\boldsymbol{\rho}\times] \\ -\boldsymbol{\rho}^T \end{bmatrix} \tag{59}$$

with

$$[\boldsymbol{\rho}\times] = \begin{bmatrix} 0 & -q_3 & q_2 \\ q_3 & 0 & -q_1 \\ -q_2 & q_1 & 0 \end{bmatrix} \tag{60}$$

The composition of two quaternions is given by

$$\mathbf{q}_2 \otimes \mathbf{q}_1 = \begin{bmatrix} \Psi(\mathbf{q}_2) & \mathbf{q}_2 \end{bmatrix} \mathbf{q}_1 = \begin{bmatrix} \Sigma(\mathbf{q}_1) & \mathbf{q}_1 \end{bmatrix} \mathbf{q}_2 \tag{61}$$

The inverse of a quaternion is defined by

$$\mathbf{q}^{-1} = \begin{bmatrix} -\boldsymbol{\rho}^T & q_4 \end{bmatrix}^T \tag{62}$$

The quaternion kinematics is

$$\dot{\mathbf{q}} = \frac{1}{2}\Xi(\mathbf{q})\boldsymbol{\omega} \tag{63}$$

E-Step Given $\hat{\boldsymbol{\theta}}$, the states of interest in the E-Step are the attitude and gyro bias. The state dynamics model and observation model are given below.

- Dynamics model

$$\dot{\mathbf{q}} = \frac{1}{2}\Xi(\mathbf{q})(\tilde{\boldsymbol{\omega}} - \boldsymbol{\beta} - \boldsymbol{\eta}_v) \tag{64}$$

$$\dot{\boldsymbol{\beta}} = \boldsymbol{\eta}_u \tag{65}$$

- Observation model

$$(I_{3\times 3} + D)\mathbf{B}_k = A(\mathbf{q}_k)\mathbf{H}_k + \mathbf{b} + \boldsymbol{\epsilon}_k \tag{66}$$

Under the assumption that the attitude error and gyro bias error are small, the joint attitude and gyro bias density is approximately Gaussian. The EKS provides the mean and covariance of the Gaussian attitude density. Like the Kalman smoother, the attitude EKS consists of forward recursions and backward recursions. The forward recursions are given by the standard multiplicative extended Kalman filter (MEKF).[?,?]

- Forward Recursions

$$\hat{\boldsymbol{\omega}}(t) = \tilde{\boldsymbol{\omega}}(t) - \hat{\boldsymbol{\beta}}(t) \tag{67}$$

$$F = \begin{bmatrix} -[\hat{\boldsymbol{\omega}}(t)\times] & I_{3\times 3} \\ 0_{3\times 3} & 0_{3\times 3} \end{bmatrix} \tag{68}$$

$$G = \begin{bmatrix} -I_{3\times 3} & 0_{3\times 3} \\ 0_{3\times 3} & I_{3\times 3} \end{bmatrix} \tag{69}$$

$$\dot{\hat{\mathbf{q}}} = \frac{1}{2}\Xi(\hat{\mathbf{q}})\hat{\boldsymbol{\omega}} \tag{70}$$

$$\dot{P} = FP + PF^T + GQG^T \tag{71}$$

$$\boldsymbol{\nu}_k = (I_{3\times 3} + D)\mathbf{B}_k - A(\hat{\mathbf{q}}_k^-)\mathbf{H}_k - \mathbf{b} \tag{72}$$

$$H_k = \left[\left[(A(\hat{\mathbf{q}}_k^-)\mathbf{H}_k)\times \right] \quad 0_{3\times 3} \right] \tag{73}$$

$$K_k = P_k^- H_k^T \left[H_k P_k^- H_k^T + \Sigma \right]^{-1} \tag{74}$$

$$\hat{\mathbf{x}}_k^+ = \begin{bmatrix} \boldsymbol{\delta}\hat{\boldsymbol{\alpha}}_k^+ \\ \boldsymbol{\Delta}\hat{\boldsymbol{\beta}}_k^+ \end{bmatrix} = K_k \boldsymbol{\nu}_k \tag{75}$$

$$\hat{\mathbf{q}}_k^+ = \hat{\mathbf{q}}_k^- + \frac{1}{2}\Xi(\hat{\mathbf{q}}_k^-)\boldsymbol{\delta\alpha}_k^+, \quad \hat{\mathbf{q}}_k^+ \leftarrow \hat{\mathbf{q}}_k^+ / \|\hat{\mathbf{q}}_k^+\| \tag{76}$$

$$\hat{\boldsymbol{\beta}}_k^+ = \hat{\boldsymbol{\beta}}_k^- + \Delta\hat{\boldsymbol{\beta}}_k^+ \tag{77}$$

$$P_k^+ = \left[I_{6\times 6} - K_k H_k \right] P_k^- \tag{78}$$

- Backward Recursions

$$P_N^S = P_N^+ \tag{79}$$

$$\mathcal{K}_k = P_k^+ \Phi^T \left(P_{k+1}^- \right)^{-1} \tag{80}$$

$$\hat{\mathbf{x}}_k^S = \begin{bmatrix} \boldsymbol{\delta}\hat{\boldsymbol{\alpha}}_k^S \\ \boldsymbol{\Delta}\hat{\boldsymbol{\beta}}_k^S \end{bmatrix} = \mathcal{K}_k \begin{bmatrix} 2\left(\boldsymbol{\rho}(\hat{\mathbf{q}}_{k+1}^S \otimes (\hat{\mathbf{q}}_k^+)^{-1}) - \boldsymbol{\rho}(\hat{\mathbf{q}}_{k+1}^- \otimes (\hat{\mathbf{q}}_k^+)^{-1}) \right) \\ \hat{\boldsymbol{\beta}}_{k+1}^S - \hat{\boldsymbol{\beta}}_{k+1}^- \end{bmatrix} \tag{81}$$

$$\hat{\mathbf{q}}_k^S = \hat{\mathbf{q}}_k^+ + \frac{1}{2}\Xi(\hat{\mathbf{q}}_k^+)\boldsymbol{\delta\alpha}_k^S, \quad \hat{\mathbf{q}}_k^S \leftarrow \hat{\mathbf{q}}_k^S / \|\hat{\mathbf{q}}_k^S\| \tag{82}$$

$$\hat{\boldsymbol{\beta}}_k^S = \hat{\boldsymbol{\beta}}_k^+ + \Delta\hat{\boldsymbol{\beta}}_k^S \tag{83}$$

$$P_k^S = P_k^+ + \mathcal{K}_k \left[P_{k+1}^S - P_{k+1}^- \right] \mathcal{K}_k^T \tag{84}$$

where $\boldsymbol{\rho}(\cdot)$ denotes the vector part of the quaternion.

M-Step Given $\hat{A}_k^S \triangleq A(\hat{\mathbf{q}}_k^S)$, the magnetometer calibration parameters $\boldsymbol{\theta}$ are updated in the M-Step by solving the following least-square estimation problem:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2} \sum_{k=1}^{N} \left[(I_{3\times 3} + D)\mathbf{B}_k - \hat{A}_k^S \mathbf{H}_k - \mathbf{b} \right]^T \Sigma^{-1} \left[(I_{3\times 3} + D)\mathbf{B}_k - \hat{A}_k^S \mathbf{H}_k - \mathbf{b} \right] \tag{85}$$

The solution is given by

$$\hat{\boldsymbol{\theta}} = \left(\sum_{k=1}^{N} \mathcal{H}_k^T \Sigma^{-1} \mathcal{H}_k \right)^{-1} \left(\sum_{k=1}^{N} \mathcal{H}_k^T \Sigma^{-1} \mathbf{z}_k \right) \tag{86}$$

with

$$\mathcal{H}_k = \left[I_{3\times 3} \quad -B(\mathbf{B}_k) \right] \tag{87}$$

$$\mathbf{z}_k = \mathbf{B}_k - \hat{A}_k^S \mathbf{H}_k \tag{88}$$

Note that only the first moment (the mean of the attitude) is needed for this problem.

Magnetometer Only

In the magnetometer only case, we assume that the attitude matrices at different times are uncorrelated. The corresponding dynamics model is

$$A_{k+1} = 0 \cdot A_k + \text{large random noise} \tag{89}$$

The observation model remains

$$(I_{3\times 3} + D)\mathbf{B}_k = A_k \mathbf{H}_k + \mathbf{b} + \boldsymbol{\epsilon}_k \tag{90}$$

It is obvious that three-axis attitude cannot be determined based on the models. However, $(A\mathbf{H})_k \triangleq A_k \mathbf{H}_k$ can be estimated. The EM algorithm in this case estimates $(A\mathbf{H})_k$ in the E-Step and then updates the calibration parameters in the M-Step.

E-Step Because the state transition matrix is the null matrix, there is no difference between a filter or a smoother. We will denote the estimate of $(A\mathbf{H})_k$ by $\widehat{(A\mathbf{H})}_k$:

$$\widehat{(A\mathbf{H})}_k = \frac{(I_{3\times 3} + D)\mathbf{B}_k - \mathbf{b}}{\|(I_{3\times 3} + D)\mathbf{B}_k - \mathbf{b}\|} \cdot \|\mathbf{H}_k\| \tag{91}$$

M-Step The same least-square estimator as in the magnetometer/gyro case is used except that $\hat{A}_k^S \mathbf{H}_k$ is replaced by $\widehat{(A\mathbf{H})}_k$.

$$\hat{\boldsymbol{\theta}} = \left(\sum_{k=1}^{N} \mathcal{H}_k^T \Sigma^{-1} \mathcal{H}_k \right)^{-1} \left(\sum_{k=1}^{N} \mathcal{H}_k^T \Sigma^{-1} \mathbf{z}_k \right) \tag{92}$$

with

$$\mathcal{H}_k = \begin{bmatrix} I_{3\times 3} & -B(\mathbf{B}_k) \end{bmatrix} \tag{93}$$

$$\mathbf{z}_k = \mathbf{B}_k - \widehat{(A\mathbf{H})}_k \tag{94}$$

NUMERICAL RESULTS

In this section, results of the EM algorithms and the scalar checking methods are shown using simulated data. The simulated spacecraft is modeled after the Tropical Rainfall Measurement Mission spacecraft. This is an Earth-pointing spacecraft (rotating about its y axis) in low Earth orbit, with an inclination of 35 deg.[?] The angular velocity of the spacecraft in the body frame is $[0, -\Omega_o, 0]$ with $\Omega_o = 0.001$ rad/s. The geomagnetic field is simulated using a 10th-order International Geomagnetic Reference Field model.[?] The magnetometer-body and geomagnetic-reference vectors for the simulated runs each have a magnitude of about 500 mG. The time history of the magnetic field is plotted in Figure 1. The magnetometer measurement noise is white and Gaussian, and the covariance is isotropic with the standard deviation $\sigma = 0.5$ mG. The measurements are sampled every 10 s over an 8-h span. The true values for the bias $\mathbf{b}$ and elements of the D matrix are:

$$b_1 = 50 \text{ mG}, \quad b_2 = 30 \text{ mG}, \quad b_3 = 60 \text{ mG} \tag{95}$$

$$D_{22} = 0.10, \quad D_{11} = D_{33} = D_{12} = D_{13} = D_{23} = 0.05 \tag{96}$$

The noise parameters of the gyro are given by $\sigma_v = \sqrt{10} \times 10^{-7}rad/s^{1/2}$ and $\sigma_u = \sqrt{10} \times 10^{-10}rad/s^{3/2}$. The initial gyro drift bias is given as 0.1 deg/s. The scalar checking methods take as input the effective scalar measurements and the EM algorithms take as input the original magnetometer measurements. The first scalar checking method minimizes the cost function given by Eq. (22) and the second one minimizes the cost function given by Eq. (23). They will be referred to as "SC I" and "SC II" in the tables. The minimization problems in the scalar checking methods are solved using the MATLAB function fmincon, which is based on sequential quadratic programming. Estimation methods compared in Ref. ? such as TWOSTEP, the EKF, and the Unscented Filter can also be used to solve the minimization problems. The termination tolerance values for fmincon, TolX and TolFun, are set to 10^{-9}. The first EM algorithm that uses both the magnetometer and gyro data is referred to in the tables as "EM I" and the second EM algorithm that uses the magnetometer data only "EM II." Both the EM algorithms and the MATLAB function fmincon are iterative algorithms. They are initialized with random samples between 0 and twice the true values. An augmented EKF, referred to as "Aug. EKF," is included to compare with the first EM algorithm. The augmented EKF estimates the attitude, the gyro biases, and the nine calibration parameters simultaneously from the original magnetometer data and the gyro data. It is a straightforward extension of the EKF for attitude and gyro bias estimation in a previous section. True noise parameters are used in the gyro model in the augmented EKF and the first EM algorithm.

Fifty Monte Carlo runs were executed. The root-mean-square (RMS) errors of the methods after they arrive at their convergence are given in Table 1. The first aspect we note from Table 1 is

Table 1 RMS Estimation Errors in the Calibration Parameters (with σ = 0.5 mG)

Parameter	SC I	SC II	EM II	EM I	Aug. EKF
$\Delta b_1/(\text{mG})$	0.1179	0.1179	0.1179	0.0720	0.0720
$\Delta b_2/(\text{mG})$	0.2086	0.2101	0.2098	0.1133	0.1127
$\Delta b_3/(\text{mG})$	0.1699	0.1699	0.1697	0.0855	0.0855
ΔD_{11}	0.1022e-3	0.1022e-3	0.1021e-3	0.0425e-3	0.0425e-3
ΔD_{22}	0.7908e-3	0.7942e-3	0.7932e-3	0.4607e-3	0.4576e-3
ΔD_{33}	0.0925e-3	0.0925e-3	0.0925e-3	0.0448e-3	0.0448e-3
ΔD_{12}	0.2765e-3	0.2765e-3	0.2765e-3	0.1699e-3	0.1698e-3
ΔD_{13}	0.0546e-3	0.0546e-3	0.0546e-3	0.0313e-3	0.0313e-3
ΔD_{23}	0.3879e-3	0.3878e-3	0.3875e-3	0.1911e-3	0.1911e-3

that the estimation errors using the second EM algorithm are very similar to those using the scalar checking methods and that the estimation errors using the first EM algorithm are very similar to those using the augmented EKF. The discrepancy among algorithms of close performance occurs mainly in Δb_2 and ΔD_{22}. The cause of this discrepancy may be related to the predominant rotation of the spacecraft about the y axis but is not yet well understood. The scalar checking methods are comparable in accuracy to the EM algorithm that assumes a very coarse attitude dynamics model and makes no use of the gyro measurements. With the aid of the gyro measurements, the magnetometer parameter estimation errors are reduced by approximately 40 to 50 percent. This shows the importance of integrating extra information in sensor calibration. In fact, even without the gyro data, the magnetometer measurements coupled with a constant angular velocity model or other tracking models are likely to increase the accuracy of magnetometer parameter estimation significantly because the low dynamics of the simulated spacecraft can be well modeled by the tracking models. The achievable accuracy with perfect knowledge about the angular velocity or about the attitude and the same amount of magnetometer data is given in Table 2. Note that Δb_2 and ΔD_{22} are much greater than the others.

Table 2 RMS Estimation Errors in the Calibration Parameters with Exact Angular Velocity or Attitude Knowledge (with σ = 0.5 mG)

Parameter	Exact Angular Velocity	Exact Attitude
$\Delta b_1/(\text{mG})$	0.0151	0.0133
$\Delta b_2/(\text{mG})$	0.1102	0.0720
$\Delta b_3/(\text{mG})$	0.0140	0.0133
ΔD_{11}	0.0426e-4	0.0428e-4
ΔD_{22}	0.4706e-4	0.3179e-4
ΔD_{33}	0.0470e-4	0.0427e-4
ΔD_{12}	0.0506e-4	0.0424e-4
ΔD_{13}	0.0332e-4	0.0302e-4
ΔD_{23}	0.0505e-4	0.0423e-4

The two scalar checking algorithms and the second EM algorithm, though similar in accuracy, are not identical. This can be seen more clearly from the test of 200 trials with $\sigma = 5$ mG and all other parameters unchanged. The results of the test are given in Table 3. It is also worth mentioning that not only is the error in b_2 is the largest, but b_2 is always underestimated.

The performance of the first EM algorithm and the augmented EKF depends on the quality of the gyro measurement model as well as the gyro measurements. In previous tests, we have assumed that the true noise parameters are available to the two estimators. Now it is assumed that the value of σ_v used by the two estimators is 400 times the true value. The actual gyro noise $\boldsymbol{\eta}_v$ is still simulated using the true value of σ_v. Because of the large error in σ_v, the performance degradation is significant, as shown in Table 4, and the benefits of integrating gyro data are lost. It is thus necessary to estimate σ_v along with the mangnetometer parameters $\boldsymbol{\theta}$. The first EM algorithm is extended to estimate all the parameters. The E-Step of it remains the same; the update of σ_v, which is independent of the estimation of $\boldsymbol{\theta}$, is added to the M-Step. A typical result of 10,000 iterations is shown in Figure 2. It can be seen that the magnetometer parameters converge within 1000 iterations while σ_v takes much longer to converge. In fact, after 10,000 iterations, the estimate of σ_v is still twice the true value. The slow convergence of some parameters is typical of the EM algorithm.

CONCLUSIONS

Two EM algorithms are presented for magnetometer calibration. The first EM algorithm assumed that accurate knowledge about the angular velocity is available and the second assumed a very coarse

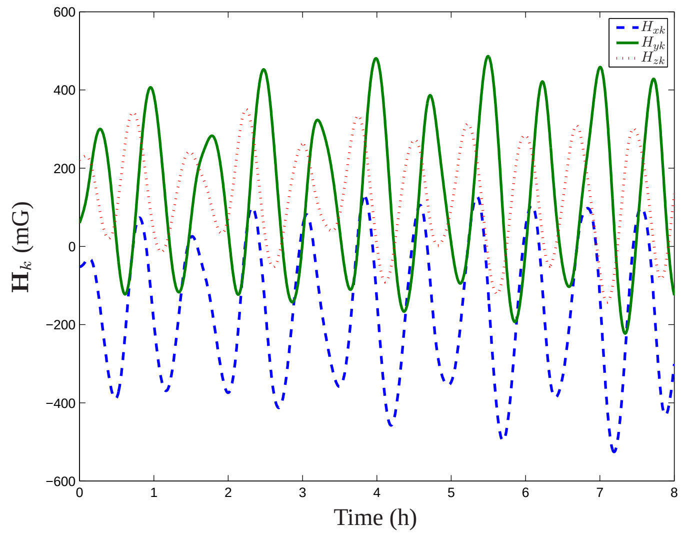
Table 3 RMS Estimation Errors in the Calibration Parameters (with σ = 5 mG)

Parameter	SC I	SC II	EM II
$\Delta b_1/(\text{mG})$	2.1207	2.1176	2.1132
$\Delta b_2/(\text{mG})$	11.5442	11.7678	11.7599
$\Delta b_3/(\text{mG})$	2.5427	2.5334	2.5194
ΔD_{11}	0.00325	0.00325	0.00324
ΔD_{22}	0.04424	0.04474	0.04470
ΔD_{33}	0.00367	0.00367	0.00366
ΔD_{12}	0.00553	0.00553	0.00552
ΔD_{13}	0.00066	0.00066	0.00066
ΔD_{23}	0.00591	0.00590	0.00587

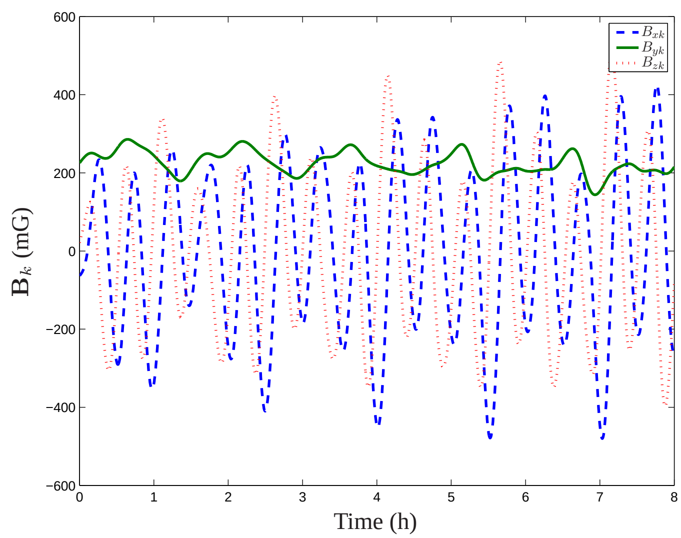
Table 4 RMS Estimation Errors in the Calibration Parameters (with σ = 0.5 mG and Incorrect σ_v)

$\frac{\Delta b_1}{(\text{mG})}$	$\frac{\Delta b_2}{(\text{mG})}$	$\frac{\Delta b_3}{(\text{mG})}$	ΔD_{11}	ΔD_{22}	ΔD_{33}	ΔD_{12}	ΔD_{13}	ΔD_{23}
0.1052	0.3203	0.1650	0.113e-3	1.203e-3	0.115e-3	0.254e-3	0.051e-3	0.379e-3

dynamics model for the attitude motion which essentially ignores the angular velocity information. The first EM algorithm is 40 to 50 percent more accurate than the second EM algorithm because of the additional information contained in the gyro data. However, when large errors exist in the noise parameters of the gyro measurement model, the advantage of integrating the gyro data in magnetometer calibration becomes minimal. The first EM algorithm can be easily extended to estimate other parameters such as the noise parameters of the gyro. Numerical study indicated that the scalar checking methods are close in accuracy to the second EM algorithm. Within the EM framework, the extended Kalman smoother or filter can be replaced by more robust nonlinear estimators in the presence of high nonlinearity or large uncertainty. The EM algorithm is a batch method, which requires more data storage than a recursive algorithm. The recursive version of the EM algorithm will be further investigated.

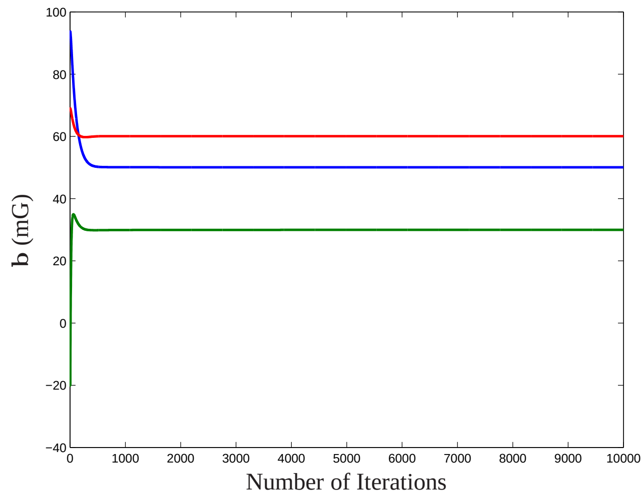


(a) representations w.r.t. reference frame

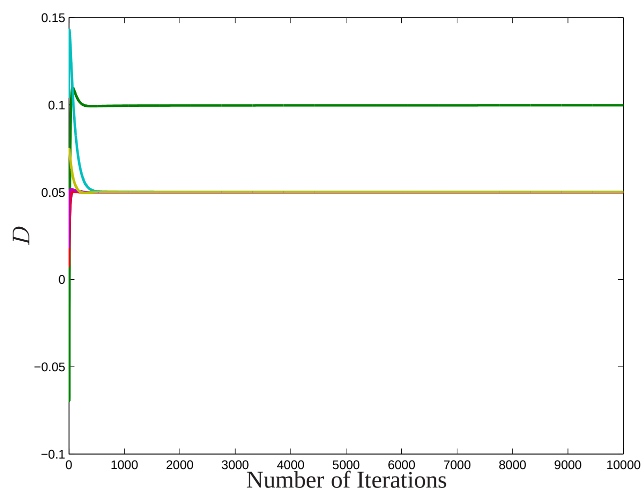


(b) representations w.r.t. body frame

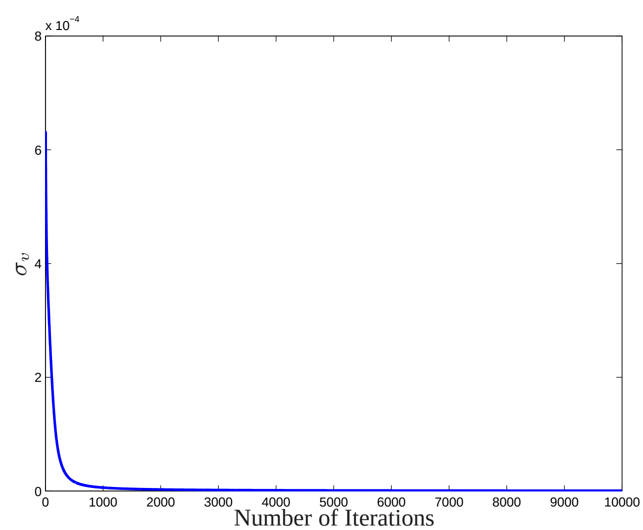
Figure 1   Time History of Earth's Magnetic Field



(a)



(b)



(c)

Figure 2 Convergence of Calibration Parameters