Flight Mechanics/Estimation Theory Symposium, NASA Goddard Space Flight Center, Greenbelt, Maryland, May 17–19, 1994, pp. 513–527

A NEW ALGORITHM FOR ATTITUDE-INDEPENDENT MAGNETOMETER CALIBRATION

Roberto Alonso * **and Malcolm D. Shuster** †

A new algorithm is developed for inflight magnetometer bias determination without knowledge of the attitude. This algorithm combines the fast convergence of an heuristic algorithm currently in use with the correct treatment of the statistics and without discarding data. The algorithm performance is examined using simulated data and compared with previous algorithms.

Introduction

At orbit injection, the only attitude sensor which may be operating is often the vector magnetometer. Frequently, the spacecraft is spinning rapidly, and, if the spacecraft is not in an equatorial orbit or at too high an altitude, it is possible on the basis of this sensor alone to determine the spin rate and the spin-axis attitude of the spacecraft. At the same time, the accuracy of the magnetometer data may be compromised by large systematic magnetic disturbances on the spacecraft, often the result of space charging during launch or from electrical currents within the spacecraft. Thus, some means is usually needed to quickly determine this bias. Since the three-axis attitude of the spacecraft usually cannot be determined at this stage, the desired algorithm must not require a knowledge of the attitude as input.

A number of algorithms have been proposed for estimating the magnetometer bias. The simplest is to solve for the bias vector by minimizing the weighted sum of the squares of residuals which are the differences in the squares of the magnitudes of the measured and modeled magnetic fields [1]. This approach has the disadvantage that the cost function is quartic in the magnetometer bias, and therefore admits multiple minima. If these solutions are close to one another, then convergence of the algorithm may be poor. Typically, one initiates the least-squares procedure by assuming that the initial magnetometer bias vector

* Head, Attitude Control Group, Comisión Nacional de Actividades Espaciales, 1425 Buenos Aires, Argentina

† Senior Professional Staff, Space Department, The Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723.

vanishes, which may lead to slow convergence if the magnetometer bias is large compared to the ambient magnetic field.

Gambhir [1,2] advocated centering the data to remove the quartic dependence. This leads to a cost function which is quadratic in the bias and, therefore, has a unique solution. The algorithm embodying this centering is called RESIDG (supposedly, "G" for Gambhir) and has been employed with good results for nearly two decades. The centering, however, necessarily discards part of the data, and the effect of this loss of data on the accuracy has never been studied. In addition, RESIDG does not make any attempt to treat the statistics correctly, so that a it is not possible to assess the accuracy of the estimation adequately.

A second approach has been put forth by Thompson [3,4], who preferred to construct a fixed-point algorithm, which he chose to call, with obvious reference, RESIDT. Fixed-point algorithms have the advantage of converging quickly when one is far from the solution, but can become intolerably slow as one approaches the solution. Thompson's algorithm was successfully employed in support of the AMPTE spacecraft.

Davenport [5] has proposed another approach to solving the quartic cost function by finding an approximate solution for the magnetometer bias and using this as an initial value for the iterative solution of the least-squares problem. The approximate solution produced by this algorithm, unfortunately, makes approximations which destroy its consistency. Hence, the approximate solution cannot approach the true solution as the number of data becomes infinite. However, the inconsistency seems to be no worse than about ten per cent for biases as large as one third of the ambient field. Higher accuracy can then be obtained by an iterative procedure, using the approximate estimate as a starting value. This algorithm has been applied to the magnetometers of the Hubble Space Telescope.

The present work proposes a superior solution which: is almost as fast as the centered algorithm of Refs. 1 and 2, without discarding data or ignoring the correlations introduced by centering; does not suffer from the convergence problems of a fixed-point algorithm such as in Ref. 3 and 4; is much more direct than the algorithm of Ref. 5; and is consistent as well at every stage. It does this in several important ways by (1) treating the statistics more completely and correctly, (2) correcting for the centering operation, and (3) estimating scale factors as well as biases. The authors do not call this algorithm either RESIDA or RESIDS.

The Model

All treatments begin with the model

$$\mathbf{B}_k = A_k \mathbf{H}_k + \mathbf{b} + \boldsymbol{\epsilon}_k , \quad k = 1, \ldots , N , \tag{1}$$

where $\mathbf{B}_k$ is the measurement of the magnetic field (more exactly, magnetic induction) by the magnetometer at time t_k; $\mathbf{H}_k$ is the corresponding value of the geomagnetic field with respect to an Earth-fixed coordinate system; A_k is the attitude of the magnetometer with respect to the Earth-fixed coordinates; $\mathbf{b}$ is the magnetometer bias; and $\boldsymbol{\epsilon}_k$ is the measurement noise. The measurement noise, which includes both sensor errors and geomagnetic field model uncertainties, is generally assumed to be white and Gaussian. This is probably a poor approximation, since the errors in the geomagnetic field model are certainly highly correlated, and, in fact, generally dominate the instrument errors. However, for the sake of argument we shall assume here that the errors are white and Gaussian.

To eliminate the dependence on the attitude, we transpose terms in equation (1) and compute the square, so that at each time

$$|\mathbf{H}_k|^2 = |A_k\mathbf{H}_k|^2 = |\mathbf{B}_k - \mathbf{b} - \boldsymbol{\epsilon}_k|^2 \tag{2ab}$$

$$= |\mathbf{B}_k|^2 - 2\,\mathbf{B}_k \cdot \mathbf{b} + |\mathbf{b}|^2 - 2(\mathbf{B}_k - \mathbf{b}) \cdot \boldsymbol{\epsilon}_k + |\boldsymbol{\epsilon}_k|^2 . \tag{2c}$$

If we now define effective measurements and measurement noise according to

$$z_k \equiv |\mathbf{B}_k|^2 - |\mathbf{H}_k|^2 , \tag{3a}$$

$$v_k \equiv 2(\mathbf{B}_k - \mathbf{b}) \cdot \boldsymbol{\epsilon}_k - |\boldsymbol{\epsilon}_k|^2 , \tag{3b}$$

then we can write

$$z_k = 2\,\mathbf{B}_k \cdot \mathbf{b} - |\mathbf{b}|^2 + v_k , \quad k = 1, \ldots, N . \tag{4}$$

This is the starting point for the derivation of all of the algorithms. (Note that in equations (3b) and (4), $\mathbf{B}_k$ is the value about which the measurement is linearized and therefore must be interpreted as the sample value of the measured magnetic field and not a random variable.)

Even with the assumption that the original measurement noise is white and Gaussian, the effective measurement noise is not white and Gaussian. Assuming that $\boldsymbol{\epsilon}_k$ is white and Gaussian, so that

$$\boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \Sigma_k) , \tag{5}$$

and

$$E\{\, \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_\varrho^{\mathrm{T}} \,\} = 0 \quad \text{for} \quad k \neq \varrho , \tag{6}$$

then

$$\mu_k \equiv E\{\, v_k \,\} = -\mathrm{tr}\,(\Sigma_k) , \tag{7a}$$

$$\sigma_k^2 \equiv E\{\, v_k^2 \,\} - \mu_k^2 = 4\,(\mathbf{B}_k - \mathbf{b})^{\mathrm{T}} \Sigma_k (\mathbf{B}_k - \mathbf{b}) + 2 \sum_{i=1}^{3} (\Sigma_k)_{ii}^2 , \tag{7b}$$

so that v_k must contain both Gaussian and χ^2 components, as is evident from equation (3b). Here $\mathrm{tr}\,(\cdot)$ denotes the trace operation. In addition,

$$E\{\, v_k\, v_\varrho \,\} = \mu_k \mu_\varrho , \tag{8}$$

so that the v_k are uncorrelated but not white. If we assume, however, that the noise $\boldsymbol{\epsilon}_k$ is small compared to the geomagnetic field, then to a large degree v_k is Gaussian and we can write approximately

$$v_k \sim \mathcal{N}(\mu_k, \sigma_k^2) , \tag{9}$$

keeping only the first term in equation (7b).

Maximum Likelihood Estimate of the Bias and Scoring

Given the statistical model above, the negative-log-likelihood function [6] for the magnetometer bias is given by

$$J(\mathbf{b}) = \frac{1}{2} \sum_{k=1}^{N} \left[\frac{1}{\sigma_k^2} (z_k - 2\,\mathbf{B}_k \cdot \mathbf{b} + |\mathbf{b}|^2 - \mu_k)^2 + \log \sigma_k^2 + \log 2\pi \right] , \tag{10}$$

which is quartic in $\mathbf{b}$. The maximum-likelihood estimate maximizes the likelihood of the estimate, which is the probability density of the measurements (evaluated at their sampled values) given as a function of the magnetometer bias. Hence, it minimizes the negative logarithm of the likelihood (equation (10)), which thus provides a cost function.

Since the domain of $\mathbf{b}$ has no boundaries, the maximum-likelihood estimate for $\mathbf{b}$, which we denote by $\mathbf{b}^*$, which minimizes the negative-log-likelihood function, must satisfy

$$\left.\frac{\partial J}{\partial \mathbf{b}}\right|_{\mathbf{b}^*} = \mathbf{0}. \tag{11}$$

Note that only the first of the three terms under the summation depends on the magnetometer bias. Unless one wishes to estimate parameters of the measurement noise, there is no reason to retain the remaining two terms. This quartic dependence can be avoided if complete three-axis attitude information is available, since the bias term then enters linearly into the measurement model (q.v. equation (1)) as in the work of Lerner and Shuster [7].

The most direct solution is obtained by scoring, which in this case is the Newton–Raphson approximation. We consider the sequence[1]

$$\mathbf{b}_0^{\mathrm{NR}} = \mathbf{0}, \qquad \mathbf{b}_{i+1}^{\mathrm{NR}} = \mathbf{b}_i^{\mathrm{NR}} - \left[\frac{\partial^2 J}{\partial \mathbf{b}\partial \mathbf{b}^{\mathrm{T}}}\left(\mathbf{b}_i^{\mathrm{NR}}\right)\right]^{-1} \frac{\partial J}{\partial \mathbf{b}}\left(\mathbf{b}_i^{\mathrm{NR}}\right). \tag{12}$$

This series is obtained by expanding $J(\mathbf{b})$ to quadratic order in $(\mathbf{b}-\mathbf{b}_i^{\mathrm{NR}})$, setting the gradient of the truncated series to zero, and solving for $\mathbf{b}_{i+1}$. If for some value of i we are sufficiently close to the maximum-likelihood estimate, then it will be true that

$$\lim_{i\to\infty} \mathbf{b}_i^{\mathrm{NR}} \to \mathbf{b}^*. \tag{13}$$

We have made the convention here that the partial derivative of a scalar function with respect to a column vector is again a column vector. The gradient vector $\partial J/\partial \mathbf{b}$ is the 3×1 matrix

$$\frac{\partial J}{\partial \mathbf{b}} = -\sum_{k=1}^{N} \frac{1}{\sigma_k^2}\left(z_k - 2\,\mathbf{B}_k\cdot\mathbf{b} + |\mathbf{b}|^2 - \mu_k\right) 2\left(\mathbf{B}_k - \mathbf{b}\right), \tag{14}$$

and the Hessian matrix $\partial^2 J/\partial \mathbf{b}\partial \mathbf{b}^{\mathrm{T}}$ is given by the 3×3 matrix

$$\frac{\partial^2 J}{\partial \mathbf{b}\partial \mathbf{b}^{\mathrm{T}}} = \sum_{k=1}^{N} \frac{1}{\sigma_k^2}\left[4\left(\mathbf{B}_k - \mathbf{b}\right)\left(\mathbf{B}_k - \mathbf{b}\right)^{\mathrm{T}} + 2\left(z_k - 2\,\mathbf{B}_k\cdot\mathbf{b} + |\mathbf{b}|^2 - \mu_k\right) I_{3\times 3}\right]. \tag{15}$$

Generally, the second term in the brackets will be much smaller than the first and can be discarded.

A second approach to scoring is the Gauss–Newton approximation [8]. In this case, we replace the Hessian matrix by its expectation, the Fisher information matrix F. Since

$$E\{\left(z_k - 2\,\mathbf{B}_k\cdot\mathbf{b} + |\mathbf{b}|^2 - \mu_k\right)\} = 0, \tag{16}$$

[1]Throughout this work we shall use k as the time index and i as the iteration index.

this amounts to discarding the second term. According to the law of large numbers, as the number of independent identically distributed (i.i.d.) samples of a random variable becomes infinite (the asymptotic limit), the average of these samples approaches the expectation value of the random variable. Our measurements are not identically distributed because of the dependence on $\mathbf{B}_k$. However, if the distribution of the values of $A_k\mathbf{H}_k$ is regularly repeated, then we may regard the measurements as being i.i.d. for each value of $A_k\mathbf{H}_k$. Except for the replacement of the Hessian matrix in equations (12) by the Fisher information matrix, the iteration proceeds as before.

For both the Newton–Raphson and the Gauss–Newton method, the estimate error covariance matrix is given in the limit of infinitely large data samples by

$$P_{bb} \to F_{bb}^{-1} = \left[\sum_{k=1}^{N} \frac{1}{\sigma_k^2} 4\,(\mathbf{B}_k - \mathbf{b})(\mathbf{B}_k - \mathbf{b})^{\mathrm{T}} \right]^{-1} . \tag{17}$$

If the measurement noise is Gaussian, then the asymptotic limit is true, in fact, for finite data samples. In most cases, the Fisher information matrix is simpler to evaluate than the Hessian matrix of the negative-log-likelihood function, and often can be evaluated independently of the data.

The earliest estimates of the magnetometer bias were accomplished by the method culminating in equations (12) though usually the weights were not chosen according to a statistical criterion.

The Centered Estimate

In order to avoid the minimization of a quartic cost function, let us define the following weighted averages

$$\overline{z} \equiv \overline{\sigma}^2 \sum_{k=1}^{N} \frac{1}{\sigma_k^2} z_k , \quad \overline{\mathbf{B}} \equiv \overline{\sigma}^2 \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \mathbf{B}_k , \quad \overline{v} \equiv \overline{\sigma}^2 \sum_{k=1}^{N} \frac{1}{\sigma_k^2} v_k , \quad \overline{\mu} \equiv \overline{\sigma}^2 \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \mu_k , \tag{18}$$

where

$$\frac{1}{\overline{\sigma}^2} \equiv \sum_{k=1}^{N} \frac{1}{\sigma_k^2} . \tag{19}$$

This is similar to the centering approximation of Gambhir [1, 2], who, however, did not determine the weights from any statistical quantities. It follows that

$$\overline{z} = 2\,\overline{\mathbf{B}} \cdot \mathbf{b} - |\mathbf{b}|^2 + \overline{v} . \tag{20}$$

If we define now

$$\tilde{z}_k \equiv z_k - \overline{z} , \qquad \widetilde{\mathbf{B}}_k \equiv \mathbf{B}_k - \overline{\mathbf{B}} , \qquad \tilde{v}_k \equiv v_k - \overline{v} , \qquad \tilde{\mu}_k \equiv \mu_k - \overline{\mu} , \tag{21}$$

then subtracting equation (20) from equation (4) leads to

$$\tilde{z}_k = 2\,\widetilde{\mathbf{B}}_k \cdot \mathbf{b} + \tilde{v}_k , \quad k = 1, \ldots , N . \tag{22}$$

This operation is called *centering*.

The centered measurement is no longer quadratic in the magnetometer bias vector. However, the centered measurement noise is no longer uncorrelated. Thus, one can no longer write the negative-log-likelihood function in the form of equation (10). Nonetheless, in practice one has ignored this and determined the bias from a cost function of the form [2]

$$J^{\text{approx}}(\mathbf{b}) = \frac{1}{2}\sum_{k=1}^{N-1}\frac{1}{\sigma_k^2}\left(\tilde{z}_k - 2\,\widetilde{\mathbf{B}}_k\cdot\mathbf{b} - \tilde{\mu}_k\right)^2, \tag{23}$$

and achieved reasonable results in spite of the lack of consistency, arguing that one was only discarding a single measurement out of many. We shall see below that one can discard much more than $1/N$ of the accuracy by this operation, but we shall see also that equation (23) is closer to being correct than one might have imagined. Note that the sum is from 1 to $N-1$, since the centered measurements are not independent.

$$\sum_{k=1}^{N}\tilde{z}_k = 0\,. \tag{24}$$

Minimizing $J^{\text{approx}}(\mathbf{b})$ over $\mathbf{b}$ leads to

$$\mathbf{b}^{*\,\text{approx}} = P_{bb}^{\text{approx}}\sum_{k=1}^{N-1}\frac{1}{\sigma_k^2}\left(\tilde{z}_k - \tilde{\mu}_k\right)2\widetilde{\mathbf{B}}_k\,, \tag{25}$$

with the estimate error covariance matrix given approximately by

$$P_{bb}^{\text{approx}} \approx \left(F_{bb}^{\text{approx}}\right)^{-1} = \left[\sum_{k=1}^{N-1}\frac{1}{\sigma_k^2}\,4\,\widetilde{\mathbf{B}}_k\widetilde{\mathbf{B}}_k^{\mathrm{T}}\right]^{-1} \tag{26}$$

Note that $\tilde{\mu}_k$ will vanish if the original measurement noise ϵ_k, $k = 1, \ldots, N$, is identically distributed. The centered estimator converges in a single iteration because the cost function is exactly quadratic.

Fixed-Point Method

To avoid the loss of data from centering, Thompson, Neal and Shuster [3, 4] proposed a fixed-point algorithm. Define the quantities

$$G \equiv \sum_{k=1}^{N}\frac{1}{\sigma_k^2}\left[4\,\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}} + 2\left(z_k - \mu_k\right)I_{3\times 3}\right], \tag{27a}$$

$$\mathbf{a} \equiv \sum_{k=1}^{N}\frac{1}{\sigma_k^2}\left(z_k - \mu_k\right)2\mathbf{B}_k\,, \tag{27b}$$

$$\mathbf{f}(\mathbf{b}) \equiv \sum_{k=1}^{N}\frac{1}{\sigma_k^2}\left[4\left(\mathbf{B}_k\cdot\mathbf{b}\right)\mathbf{b} + 2\,|\mathbf{b}|^2(\mathbf{B}_k - \mathbf{b})\right]. \tag{27c}$$

[2] In actual fact, these calculations have almost always assumed a constant weighting and neglected the contribution of μ_k.

Then the gradient of the negative-log-likelihood function becomes

$$\frac{\partial J(\mathbf{b})}{\partial \mathbf{b}} = G\,\mathbf{b} - \mathbf{a} - \mathbf{f}(\mathbf{b}) = 0\,. \tag{28}$$

which can be solved implicitly to yield

$$\mathbf{b}^* = G^{-1}\left[\mathbf{a} + \mathbf{f}(\mathbf{b}^*)\right]\,. \tag{29}$$

This equation must be solved iteratively,

$$\mathbf{b}_0^{\mathrm{FP}} = \mathbf{0}\,, \qquad \mathbf{b}_{i+1}^{\mathrm{FP}} = G^{-1}\left[\mathbf{a} + \mathbf{f}(\mathbf{b}_i^{\mathrm{FP}})\right]\,, \tag{30}$$

and we expect that once $\mathbf{b}_i^{\mathrm{FP}}$ is sufficiently close to the solution that

$$\lim_{i\to\infty} \mathbf{b}_i^{\mathrm{FP}} = \mathbf{b}^*\,. \tag{31}$$

Davenport's Approximation

Davenport and his collaborators [5] have offered an approximate form for the bias vector estimator. He begins by writing an approximate cost function as

$$J_D(\mathbf{b}) = \frac{1}{2}\sum_{k=1}^{N} \frac{1}{\sigma_k^2}\,(z_k - 2\,\mathbf{B}_k \cdot \mathbf{b} + \lambda^2 - \mu_k)^2\,, \tag{32}$$

where λ is a constant. This cost function would agree with that of equation (10) when $\lambda = |\mathbf{b}|$. Davenport, however, allows λ to be a free parameter.

The cost function of equation (32) is only quadratic in $\mathbf{b}$. Differentiating this cost function with respect to $\mathbf{b}$ and setting the gradient equal to zero leads to a solution of the form

$$\mathbf{b}_D^* = \mathbf{f}_D(\lambda)\,, \tag{33}$$

that is, the estimate of the bias is a function of the parameter λ. The "consistent" value of this parameter is obtained by solving

$$|\mathbf{f}_D(\lambda)|^2 = \lambda^2\,. \tag{34}$$

Because Davenport's algorithm effectively changes the dependence of the non-random part of the measurement on the bias even in the absence of noise, it cannot be consistent. Thus, as more data is accumulated the accuracy will not improve. However, it can be used as the starting point for a Newton-Raphson or Gauss-Newton iteration of the quartic cost function.

A Statistically Correct Centered Algorithm

The original data, z_k, $k = 1, \ldots, N$, may be replaced by the centered data, $\tilde{z}_k$, $k = 1, \ldots, N-1$, and the center value $\bar{z}$. The measurement equations are given by

equations (20) and (22). The centered data have the advantage of depending only linearly on the magnetometer bias. However, they have the disadvantage that the centered measurement noise is correlated. Therefore, the cost function for the centered data alone cannot be written as the sum of $N-1$ squares. To write a statistically correct cost function for the centered data (making the approximation that the measurement noise v_k is Gaussian) we define

$$\widetilde{\mathcal{Z}} \equiv [\,\tilde{z}_1,\, \tilde{z}_2,\, \ldots,\, \tilde{z}_{N-1}\,]^{\mathrm{T}}\,, \qquad \widetilde{\mathcal{B}} \equiv \left[\,\widetilde{\mathbf{B}}_1,\, \widetilde{\mathbf{B}}_2,\, \ldots,\, \widetilde{\mathbf{B}}_{N-1}\,\right]^{\mathrm{T}}\,, \tag{35ab}$$

$$\widetilde{\mathcal{M}} \equiv [\,\tilde{\mu}_1,\, \tilde{\mu}_2,\, \ldots,\, \tilde{\mu}_{N-1}\,]^{\mathrm{T}}\,, \qquad \widetilde{\mathcal{V}} \equiv [\,\tilde{v}_1,\, \tilde{v}_2,\, \ldots,\, \tilde{v}_{N-1}\,]^{\mathrm{T}}\,, \tag{35cd}$$

and write formally

$$\widetilde{\mathcal{Z}} = 2\,\widetilde{\mathcal{B}}\,\mathbf{b} + \widetilde{\mathcal{V}}\,, \tag{36}$$

with

$$\widetilde{\mathcal{V}} \sim \mathcal{N}\left(\widetilde{\mathcal{M}},\, \widetilde{\mathcal{R}}\right)\,. \tag{37}$$

Here $\widetilde{\mathcal{R}}$ is the covariance matrix of $\widetilde{\mathcal{V}}$. (Note that $\widetilde{\mathcal{B}}$ is an $(N-1)\times 3$ matrix.)

The negative-log-likelihood function for this stacked centered measurement is simply

$$\widetilde{J}(\mathbf{b}) = \frac{1}{2}\left[\left(\widetilde{\mathcal{Z}} - 2\,\widetilde{\mathcal{B}}\,\mathbf{b} - \widetilde{\mathcal{M}}\right)^{\mathrm{T}}\widetilde{\mathcal{R}}^{-1}\left(\widetilde{\mathcal{Z}} - 2\,\widetilde{\mathcal{B}}\,\mathbf{b} - \widetilde{\mathcal{M}}\right) + \log\det\widetilde{\mathcal{R}} + (N-1)\log 2\pi\right]\,. \tag{38}$$

Equation (23) made the assumption that $\widetilde{\mathcal{R}}$ was diagonal. We do not make this approximation here. Minimizing this negative-log-likelihood function leads directly to

$$\widetilde{\mathbf{b}}^{*} = \left(4\,\widetilde{\mathcal{B}}^{\mathrm{T}}\widetilde{\mathcal{R}}^{-1}\widetilde{\mathcal{B}}\right)^{-1}\,2\widetilde{\mathcal{B}}^{\mathrm{T}}\widetilde{\mathcal{R}}^{-1}\left(\widetilde{\mathcal{Z}} - \widetilde{\mathcal{M}}\right)\,, \tag{39}$$

with estimate error covariance matrix

$$\widetilde{P}_{bb} = \left(4\,\widetilde{\mathcal{B}}^{\mathrm{T}}\widetilde{\mathcal{R}}^{-1}\widetilde{\mathcal{B}}\right)^{-1}\,. \tag{40}$$

For large quantities of data, the naive evaluation of equations (39) and (36) can be a formidable task. Therefore, we seek the means of inverting the matrix in equation (38) explicitly. Clearly,

$$\widetilde{\mathcal{R}}_{k\ell} = E\{\,(\tilde{v}_k - \tilde{\mu}_k)(\tilde{v}_\ell - \tilde{\mu}_\ell)\,\} = \sigma_k^2\,\delta_{k\ell} - \overline{\sigma}^2\,, \tag{41}$$

which shows the correlation explicitly. However, this matrix has the simple inverse

$$\left(\widetilde{\mathcal{R}}^{-1}\right)_{k\ell} = \frac{1}{\sigma_k^2}\,\delta_{k\ell} + \frac{\sigma_N^2}{\sigma_k^2\sigma_\ell^2}\,, \tag{42}$$

where σ_N^2 is the variance of v_N. Substituting this expression into equation (38) leads to

$$\widetilde{J}(\mathbf{b}) = \frac{1}{2}\sum_{k=1}^{N}\frac{1}{\sigma_k^2}\,(\tilde{z}_k - 2\widetilde{\mathbf{B}}_k\cdot\mathbf{b} - \tilde{\mu}_k)^2 + \text{terms independent of } \mathbf{b}\,. \tag{43}$$

The statistically correct cost function for the centered data looks exactly like the naive expression of equation (23) except that the summation is now from 1 to N. The minimization is simple and leads directly to

$$\tilde{\mathbf{b}}^* = \tilde{P}_{bb} \sum_{k=1}^{N} \frac{1}{\sigma_k^2} (\tilde{z}_k - \tilde{\mu}_k) 2\tilde{\mathbf{B}}_k , \tag{44}$$

and the estimate error covariance of the centered estimate is given by

$$\tilde{P}_{bb} = \tilde{F}_{bb}^{-1} = \left[\sum_{k=1}^{N} \frac{1}{\sigma_k^2} 4 \tilde{\mathbf{B}}_k \tilde{\mathbf{B}}_k^{\mathrm{T}} \right]^{-1} . \tag{45}$$

The centered estimate is seen now to be much more attractive than before. It is simple, and by a very trivial alteration (replacing the sum from 1 to $N-1$ by a sum from 1 to N) it can be made to treat the statistics of the measurement noise correctly. It is very different in character from that the centered estimate of Gambhir [1,2]. It is thus to be preferred to Thompson's algorithm [3,4], whose convergence can be problematic, and to Davenport's approximation [5], which is not consistent. The greatest drawback to the centered algorithm lies in the exclusion of certain data, the effect of which we now investigate.

The Complete Solution

For N large, the naive centered algorithm presented earlier is hardly worse than the rigorously centered algorithm derived above. From the standpoint of computation burden, the more rigorous treatment of the statistics has merely added one more term (out of N) to the summation. However, equation (45), because it has been derived rigorously, will afford us the possibility of computing the correction from the discarded measurement $\tilde{z}$.

Instead of the measurement set $\tilde{z}_k$, $k = 1, \ldots, N-1$, $\bar{z}$, we may now consider the measurements to be effectively $\tilde{\mathbf{b}}^*$ and $\bar{z}$. Therefore, to determine the exact maximum likelihood estimate $\mathbf{b}^*$, we must develop the statistics of these two effective measurements more completely.

Let us substitute equation (22) into equation (44). This leads to

$$\tilde{\mathbf{b}}^* = \tilde{P}_{bb} \sum_{k=1}^{N} \frac{1}{\sigma_k^2} (2\tilde{\mathbf{B}}_k \cdot \mathbf{b} + \tilde{v}_k - \tilde{\mu}_k) 2\tilde{\mathbf{B}}_k , \tag{46}$$

which we may rewrite as

$$\tilde{\mathbf{b}}^* = \mathbf{b} + \tilde{P}_{bb} \sum_{k=1}^{N} \frac{1}{\sigma_k^2} 2\tilde{\mathbf{B}}_k (\tilde{v}_k - \tilde{\mu}_k) \tag{47a}$$

$$= \mathbf{b} + \tilde{\mathbf{v}}_b . \tag{47b}$$

The last term is just the (zero-mean) estimate error. Obviously,

$$\widetilde{\mathbf{v}}_b \sim \mathcal{N}\left(\mathbf{0},\, \widetilde{P}_{bb}\right). \tag{48}$$

It follows that we can write

$$\widetilde{J}(\mathbf{b}) = \frac{1}{2}\left(\mathbf{b} - \widetilde{\mathbf{b}}^*\right)^{\mathrm{T}} \widetilde{P}_{bb}^{-1}\left(\mathbf{b} - \widetilde{\mathbf{b}}^*\right) + \text{terms independent of } \mathbf{b}, \tag{49}$$

which can be verified by expanding equation (43) and completing the square in **b**. The estimate $\widetilde{\mathbf{b}}^*$ is thus a sufficient statistic for **b** [6]. Equation (49) is very useful, because it allows us to investigate the effect of corrections to the centered formula using only our knowledge of $\widetilde{\mathbf{b}}^*$ and $\widetilde{P}$. We do not have to refer again to the N centered measurements $\tilde{z}_k$, $k = 1, \ldots, N$.

We must now combine $\widetilde{\mathbf{b}}^*$ and $\overline{z}$ to obtain a complete representation of our data for the computation of **b**. Recall equation (20),

$$\overline{z} = 2\,\overline{\mathbf{B}} \cdot \mathbf{b} - |\mathbf{b}|^2 + \overline{v}, \tag{20}$$

with

$$\overline{v} \sim \mathcal{N}(\overline{\mu},\, \overline{\sigma}^2). \tag{50}$$

Note that $\overline{z}$, which, unfortunately, is a nonlinear function of **b**, is nonetheless an extremely accurate measurement, more accurate than the other measurements by typically a factor of $1/\sqrt{N}$, because $\overline{\sigma}$ is smaller typically than the other variances by this factor. Thus, simply centering the data can entail a significant loss of accuracy if $\overline{\mathbf{B}} - \mathbf{b}$ is not significantly smaller than typical values of $\mathbf{B}_k$.

What is the correlation between $\widetilde{\mathbf{v}}_b$ and $\overline{v}$? Calculating this explicitly, gives

$$E\{\,\widetilde{\mathbf{v}}_b(\overline{v} - \overline{\mu})\,\} = \widetilde{P}\sum_{k=1}^{N} \frac{1}{\sigma_k^2}\widetilde{\mathbf{B}}_k E\{\,(\tilde{v}_k - \tilde{\mu}_k)(\overline{v} - \overline{\mu})\,\} \tag{51a}$$

$$= \widetilde{P}\sum_{k=1}^{N} \frac{1}{\sigma_k^2}\widetilde{\mathbf{B}}_k \overline{\sigma}^2 = \mathbf{0}, \tag{51bc}$$

and we have used equation (21). Thus, $\widetilde{\mathbf{v}}_b$ and $\overline{v}$ are uncorrelated. It follows, that the negative-log-likelihood functions add and

$$J(\mathbf{b}) = \widetilde{J}(\mathbf{b}) + \overline{J}(\mathbf{b}), \tag{52}$$

with $\widetilde{J}(\mathbf{b})$ given by equation (49) and

$$\overline{J}(\mathbf{b}) = \frac{1}{2}\left[\frac{1}{\overline{\sigma}^2}(\overline{z} - 2\,\overline{\mathbf{B}} \cdot \mathbf{b} + |\mathbf{b}|^2 - \overline{\mu})^2 + \log\overline{\sigma}^2 + \log 2\pi\right] \tag{53}$$

The weight associated with the center term $\overline{J}(\mathbf{b})$ is equal to the sum of all the weights of $\widetilde{J}(\mathbf{b})$. Thus, when $\overline{\mathbf{B}}$ is not small, the loss of accuracy from discarding the center time can

be substantial. We can determine the relative importance of these terms to the estimate accuracy by computing the Fisher information matrix F_{bb} to obtain

$$F_{bb} = \tilde{P}_{bb}^{-1} + \frac{4}{\bar{\sigma}^2}(\bar{\mathbf{B}} - \mathbf{b})(\bar{\mathbf{B}} - \mathbf{b})^{\mathrm{T}} = P_{bb}^{-1} . \tag{54}$$

The estimate error covariance matrix will be the inverse of this quantity. If the distribution of the magnetometer measurements is "isotropic," that is, if $\bar{\mathbf{B}} - \mathbf{b}$ vanishes, then $\bar{J}(\mathbf{b})$ will be insensitive to $\mathbf{b}$. It is in this case that the centering approximation obviously leads to the best results. If, however, one attempts to determine the magnetometer bias from a short data span, say, from an inertially stabilized or Earth-pointing spacecraft, then $\bar{\mathbf{B}} - \mathbf{b}$ will be equal to the similar expression for a typical value of the magnetic field, and the formerly discarded center term which will provide half of the accuracy, especially for the component along $\bar{\mathbf{B}} - \mathbf{b}$.

Thus, our new algorithm is as follows:

- We compute the centered estimate $\tilde{\mathbf{b}}^*$ of the magnetometer bias and the covariance matrix $\tilde{P}_{bb}$ using the centered data and equations (44) and (45).

- Using the centered estimate $\tilde{\mathbf{b}}^*$ as an initial estimate, the correction due to the center term is computed using the Gauss–Newton method

$$\mathbf{b}_{i+1} = \mathbf{b}_i - F_{bb}^{-1}(\mathbf{b}_i)\,\mathbf{g}(\mathbf{b}_i) , \tag{55}$$

  where the Fisher information matrix $F_{bb}(\mathbf{b})$ is given by equation (54), and the gradient vector is given by the sum of the gradients of equations (49) and (53)

$$\mathbf{g}(\mathbf{b}) = \tilde{P}_{bb}^{-1}(\mathbf{b} - \tilde{\mathbf{b}}^*) - \frac{1}{\bar{\sigma}^2}(\bar{z} - 2\bar{\mathbf{B}} \cdot \mathbf{b} + |\mathbf{b}|^2 - \bar{\mu})\,2(\bar{\mathbf{B}} - \mathbf{b}) . \tag{56}$$

- The iteration is continued until

$$\eta_i \equiv (\mathbf{b}_i - \mathbf{b}_{i-1})^{\mathrm{T}} F_{bb}(\mathbf{b}_{i-1})(\mathbf{b}_i - \mathbf{b}_{i-1}) \tag{57}$$

  is less than some predetermined small quantity.

Numerical Examples

The algorithms treated in this work have been examined for an inertially stabilized spacecraft. The spacecraft orbit has been chosen to be circular with an altitude of 560 km and an inclination of 38 deg. This is, in fact, the orbit of the SAC-B spacecraft (Satelite de Aplicaciones Científicas), the first spacecraft to be developed by Argentina, which will be inertially stabilized to observe the Sun. The geomagnetic field in our studies has been simulated using the International Geomagnetic Reference Field (IGRF (1985)) [9], which has been extrapolated to 1994. More recent field models are available, but IGRF (1985) is adequate for our simulation needs.

For purposes of simulation we have assumed an effective white Gaussian magnetometer measurement error with a standard deviation per axis of 2.0 mG, corresponding to an angular error of approximately 0.5 deg at the equator. We have assumed that no axis of the magnetometer is predominantly parallel to the spacecraft spin axis or the geomagnetic field. The data were sampled once every ten seconds.

We examine first Davenport's approximation. To highlight the inconsistency of this method, we examine its behavior and that of the centered estimate for noise-free data. The results for half an orbit of data for the spinning spacecraft are shown in Table 1. The equivalent results for noisy data are presented in Table 2.

For small values of the magnetometer bias, Davenport's approximation yields acceptable results. For values of the magnetometer bias comparable to or greater than the magnitude of the ambient magnetic field, the errors in Davenport's approximation become unacceptably large. These statements hold both for the noise-free and the noisy data.

We can gain a greater appreciation of the behavior of these two algorithms if we examine the normalized errors, $\tilde{\eta}$ and η_D defined by

Table 1. Comparison of Davenport's Approximation and Centered Estimate for Noise-Free Data

Model Bias (mG)	Centered Estimate	Davenport's Approximation
[10., 20., 30.]	[10., 20., 30.]	[10., 20., 30.]
[30., 60., 90.]	[30., 60., 90.]	[30., 50., 90.]
[60., 129., 180.]	[60., 129., 180.]	[60., 129., 180.]
[100., 200., 300.]	[100., 200., 300.]	[101., 208., 261.]
[200., 400., 600.]	[200., 400., 600.]	[180., 539., 161.]

$$\eta \equiv \frac{1}{\sqrt{6}} \left[(\mathbf{b}^{\text{true}} - \mathbf{b}^*)^{\text{T}} \tilde{P}_{bb}^{-1} (\mathbf{b}^{\text{true}} - \mathbf{b}) - 3 \right] , \tag{58}$$

which should have mean zero and standard deviation unity. A comparison of these quantities is given in Table 3. The inconsistency of Davenport's algorithm is evident.

To see the advantages of this algorithm over naive quartic scoring consider the estimation of a magnetometer bias whose true value is (10., 20., 30.) mG. The results of successive iterations for naive quartic scoring and the new algorithm are shown in Table 4. For the new algorithm, the first algorithm is the statistically correct centered algorithm and successive iteration are obtained by applying the Gauss–Newton method to the complete cost function as given by equation (53). The 1σ error brackets, computed from the Fisher information matrix, are (±.13, ±.19, ±.12) mG. The results of the two methods are nearly identical in this case and the convergence is equally rapid. Small differences in the results are due to the slightly different treatment of the roundoff errors.

Table 2. Comparison of Davenport's Approximation and Centered Estimate for Noisy Data

Model Bias (mG)	Centered Estimate	Davenport
[10., 20., 30.]	[9.88 ± 0.35, 20.47 ± 0.82, 27.77 ± 2, 44]	[10.27, 19.88, 29.70]
[30., 60., 90.]	[29.69 ± 0.30, 60.52 ± 0.69, 89.95 ± 2.03]	[29.66, 60.60, 89.66]
[60., 129., 180.]	[59.47 ± 0.25, 130.98 ± 0.65, 174.54 ± 1.58]	[54.88, 129.28, 179.75]
[100., 200., 300.]	[100.33 ± 0.22, 201.13 ± 0.56, 296.78 ± 1.60]	[101.94, 214.30, 230.13]
[200., 400., 600.]	[199.82 ± 0.37, 400.76 ± 6.34, 598.28 ± 3.44]	[178, 92, 538.68, 160.33]

Table 3. Comparison of Normalized Errors for Davenport's Approximation and the Centered Estimate for Noisy Data

Model Bias (mG)	$\tilde{\eta}$	η_D
[10., 20., 30.]	0,82	-1.18
[30., 60., 90.]	0.68	3.87
[60., 129., 180.]	0.814	0.81
[100., 200., 300.]	2.66	2.86×10^3
[200., 400., 600.]	-0.83	0.65×10^3

Consider now the case where the magnetometer bias vector is large compared with the ambient field, say (100., 200., 300.) mG. In this case we obtain the value presented in Table 5. The 1σ error brackets here are found to be (±.12, ±.10, ±.12) mG.

In this case naive quartic scoring does not even converge to the correct answer, nor does the method of Thompson, Neal and Shuster [3,4], which does not converge at all. Naive quartic scoring converges, in fact, to a local minimum. The new algorithm, on the other hand, works very well. Note that a single iteration of the center correction is sufficient. The errors for the new algorithm are clearly consistent with the computed confidence intervals.

Discussion

The new algorithm for attitude-independent magnetometer bias determination produces excellent results in all situations. Since it begins with a very good initial estimate for the bias, it is more likely to converge to the correct minimum than does naive scoring [1] or the fixed-point method of Thompson *et al.* [3,4], which begin at $\mathbf{b} = \mathbf{0}$. Unlike the centered algorithm of RESIDG fame [2], it does not discard data and does the centering in a statistically correct way, apart from the approximation that the measurement errors on the attitude-independent derived measurement are Gaussian and uncorrelated, which is almost certainly not the case. It is amusing to speculate that the statistically correctly centered cost function of equation (43) would probably be rejected as statistically incorrect by heuristic algorithm developers unschooled in Statistics, because

it appears to use redundant data. Its initial centered estimate for the magnetometer bias is clearly a better approximation than ignoring the quadratic behavior of $|\mathbf{b}|^2$ as in the work of Davenport *et al.* [5]. The new algorithm is certainly more sophisticated statistically than its predecessors, and more efficient computationally. Perhaps, most importantly, the new algorithm makes manifest the physical quantities which determine the behavior of the bias estimator.

Table 4. Comparison of Naive Quartic Scoring and the New Algorithm. The true value of the magnetometer bias vector is (10., 20., 30.) **mG.**

Iteration	Naive Quartic Scoring	New Algorithm
1	[10.08, 19.27, 33.04]	[9.82, 20.08, 29.05]
2	[9.84, 20.18, 29.91]	[9.90, 19.83, 29.94]
3	[9.84, 20.19, 29.89]	[9.90, 19.83, 29.93]
4	[9.84, 20.19, 29.89]	[9.90, 19.83, 29.93]

Table 5. Comparison of Naive Quartic Scoring and the New Algorithm. The true value of the magnetometer bias vector is (100., 200., 300.) **mG.**

Iteration	Naive Quartic Scoring	New Algorithm
1	[107.62, 259.77, 2.85]	[99.82, 200.63, 298.02]
2	[51.51, 398.62, −368.88]	[99.97, 200.11, 299.81]
3	[70.35, 358.17, −196.33]	[99.97, 200.11, 299.81]
4	[72.13, 340.88, −145.65]	...
5	[71.78, 338.71, −140.60]	...
6	[71.70, 338.64, −140.62]	...
7	[71.70, 338.64, −140.62]	...

Acknowledgments

The authors are grateful to F. Landis Markley for interesting discussions and helpful criticisms. R. Alonso is grateful to the Guidance and Control Branch of NASA Goddard Space Flight Center (GSFC), NASA, for their hospitality during which this work was carried out. He is thankful also to the Comisión Nacional de Actividades Espaciales (CONAE) and the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) for their financial support during this period.

References

[1] GAMBHIR, B. "Determination of Magnetometer Biases Using Module RESIDG," Computer Sciences Corporation, Report No. 3000-32700-01TN, March 1975.

[2] LERNER, G. M. "Scalar Checking," in Wertz, J. R. (editor) *Spacecraft Attitude Determination and Control*, Kluwer Academic, Dordrecht, 1978, pp. 328–334.

[3] THOMPSON, R. H., NEAL, G. N., and SHUSTER, M. D. "Spin-Axis Attitude Estimation and Magnetometer Bias Determination for the AMPTE Mission," *Proceedings, Flight Mechanics/Estimation Theory Symposium*, NASA Goddard Space Flight Center, Greenbelt, Maryland, October 1981.

[4] THOMPSON, R. H., NEAL, G. N., and SHUSTER, M. D. "Magnetometer Bias Determination and Spin-Axis Attitude Estimation for the AMPTE Mission,"*Journal of Guidance Control and Dynamics*, Vol. 7, No. 4, 1984, pp. 505–507.

[5] DAVENPORT, P. B., RUMPL, W. M., and WELTER, G. L. "In-Flight Determination of Spacecraft Magnetic Bias Independent of Attitude," *Proceedings, Flight Mechanics/Estimation Theory Symposium*, NASA Goddard Space Flight Center, Greenbelt, Maryland, May 10-11, 1988, pp. 326–343.

[6] SORENSON, H. W. *Parameter Estimation*, Marcel Dekker, New York, 1980.

[7] LERNER, G. M., and SHUSTER, M. D. "Inflight Magnetometer Calibration and Attitude Determination for Near-Earth Spacecraft," *Journal of Guidance and Control*, Vol. 4, No. 5, 1981, pp. 518- -522.

[8] GUPTA, N. K., and MEHRA, R. K. "Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations," *IEEE Transactions on Automatic Control*, Vol. AC-19, No. 6, 1974, pp. 774–783.

[9] LANGEL, R. A. "The Main Field," in J. A. JACOBS (editor), *Geomagnetism*, Academic Press, New York, 1987, pp. 348–349.

Please take note that the present document does not correspond page for page to the article that was printed in the conference proceedings. This has occured, because the second author has now adopted a scalable version of the Times Roman font whose character widths differ from those of the bit-mapped fonts (Bitstream Dutch) used to set the conference article. The text of the article has not been altered except for nonsubstantive changes to accommodate the placement of the tables.