The Journal of the Astronautical Sciences, Vol. 54, No. 1, January–March 2006, pp. 113–123

The TRIAD Algorithm as Maximum Likelihood Estimation

Malcolm D. Shuster[1]

But what likelihood is in that?

William Shakespeare (1564–1616)
Measure for Measure, Act IV, scene ii

Abstract

The TRIAD algorithm is shown to be derivable as a maximum-likelihood estimator. In particular, using the QUEST measurement model, the TRIAD attitude error covariance matrix can be derived as the inverse of the Fisher information matrix. The treatment here gives a microscopic analysis of the algorithm and its connection to the QUEST algorithm. It also sheds valuable light on the origin of discrete degeneracies in deterministic attitude estimation.

Introduction

The TRIAD algorithm [1, 2] in the consciousness of most workers in attitude estimation is the deterministic attitude estimation algorithm *par excellence.* This differentiates it from optimal algorithms, for example, QUEST [2–4], Markley's SVD algorithm [4, 5], Markley's FOAM algorithm [4, 6], Mortari's many ESOQ algorithms [4, 7, 8], and the new MRAD algorithm of Bruccoleri, Lee, and Mortari [9] using the modified Rodrigues parameters [10], which are all solutions of the Wahba problems [11] and all maximum-likelihood estimators for the QUEST measurement model [2]. For these the attitude-error covariance matrix can be calculated easily as the inverse of the Fisher information matrix, rather than by the brute-force method we have believed to be required by deterministic methods like TRIAD.[2] We shall show that every deterministic attitude estimator can be recast as a maximum-likelihood estimator. Our efforts to do this for the TRIAD algorithm

[1]Director of Research, Acme Spacecraft Company, 13017 Wisteria Drive, Box 328, Germantown, Maryland 20874. Email: mdshuster@comcast.net.
[2]Before it had been demonstrated that the Wahba problem could be cast as a maximum-likelihood estimator [3], the covariance matrix for the QUEST algorithm had also been calculated by brute force [2].

will lead to a better understanding of this algorithm, the Wahba problem, and deterministic algorithms in general.

The idea that there is a connection between the deterministic TRIAD algorithm and the maximum-likelihood QUEST algorithm is not new. In the earliest journal publication of QUEST [2] it was shown that for the case when there are only two direction measurements and in the limit that the weights a_1 and a_2 satisfy $a_2/a_1 \to 0$ (equivalently $\sigma_1/\sigma_2 \to 0$) the QUEST algorithm becomes the TRIAD algorithm. Naturally, the QUEST attitude error covariance matrix also becomes the TRIAD attitude error covariance matrix in that limit. The connection between the TRIAD and QUEST algorithms that we will examine here will be of a different kind and not confined to a limiting case.

The arguments used to reach those earlier conclusions about the connection of TRIAD and QUEST have had one practical consequence. Although not stated in that work, they were in part the motivation for the SCAD algorithm [12], a suboptimal algorithm which, for a star tracker, performs almost as well as QUEST.

The methods developed here for recasting a deterministic attitude estimator into a maximum-likelihood estimator do not lead necessarily to the development of more practical estimators for deterministic algorithms. For the TRIAD algorithm in particular, it is hard to imagine constructing a more efficient algorithm than equations (4) through (6) below. For more general deterministic algorithms, however, it can certainly simplify the construction of the attitude solution by replacing complicated (and possibly nonexistent) algebraic operations by well-known iterative approaches. Most importantly, these methods provide a more practical means for calculating the attitude covariance matrix for deterministic algorithms like TRIAD and offer important insights into the origin of the finite degeneracies of deterministic attitude solutions encountered earlier [13].

The TRIAD Algorithm: Basics

Suppose we are given the values of two direction measurements $\hat{\mathsf{W}}_1'$ and $\hat{\mathsf{W}}_2'$ which are the respective realizations of two direction random variables $\hat{\mathsf{W}}_1^{\text{r.v.}}$ and $\hat{\mathsf{W}}_2^{\text{r.v.}}$, with respective true values $\hat{\mathsf{W}}_1^{\text{true}}$ and $\hat{\mathsf{W}}_2^{\text{true}}$. These measurements are assumed to be the representation of two directions with respect to spacecraft body axes. The corresponding representations with respect to space axes are $\hat{\mathsf{V}}_1$ and $\hat{\mathsf{V}}_2$, assumed to be nonrandom.[3] Thus, we write for the three types of variables, namely[4]

$$\hat{\mathsf{W}}_1^{\text{r.v.}} = A^{\text{true}}\hat{\mathsf{V}}_1 + \Delta\hat{\mathsf{W}}_1^{\text{r.v.}}, \qquad \hat{\mathsf{W}}_2^{\text{r.v.}} = A^{\text{true}}\hat{\mathsf{V}}_2 + \Delta\hat{\mathsf{W}}_2^{\text{r.v.}} \tag{1ab}$$

$$\hat{\mathsf{W}}_1' = A^{\text{true}}\hat{\mathsf{V}}_1 + \Delta\hat{\mathsf{W}}_1', \qquad \hat{\mathsf{W}}_2' = A^{\text{true}}\hat{\mathsf{V}}_2 + \Delta\hat{\mathsf{W}}_2' \tag{2ab}$$

$$\hat{\mathsf{W}}_1^{\text{true}} = A^{\text{true}}\hat{\mathsf{V}}_1, \qquad \hat{\mathsf{W}}_2^{\text{true}} = A^{\text{true}}\hat{\mathsf{V}}_2 \tag{3ab}$$

A variable without superscript would denote simply a free (nonrandom) variable. The true attitude (direction-cosine) matrix [10] is obviously not a random variable, because it is not sampled from a distribution. The same is not true for the attitude estimator A^*, which is a random matrix, or for the attitude estimate $A^{*\prime}$, which is its realization. Estimators (denoted by an asterisk) are always random variables and do not require an additional superscript "r.v." Note that it is the true attitude matrix

[3]The treatment of random reference directions $\hat{\mathsf{V}}_1$ and $\hat{\mathsf{V}}_2$ is not difficult [2] but an unnecessary complication in the present work.
[4]The vectors in this work are all column vectors rather than (abstract) physical vectors. Hence, following the conventions of reference [10], we denote these by bold unslanted sans-serif letters.

which appears in each of equations (1) through (3). The attitude matrix as a free variable will be denoted by A without superscript.

The TRIAD attitude estimate $A^{*\prime}$ is constructed in the familiar way. One defines

$$\hat{\mathbf{r}}_1 = \hat{\mathbf{V}}_1, \qquad \hat{\mathbf{r}}_2 = \frac{\hat{\mathbf{V}}_1 \times \hat{\mathbf{V}}_2}{|\hat{\mathbf{V}}_1 \times \hat{\mathbf{V}}_2|}, \qquad \hat{\mathbf{r}}_3 = \hat{\mathbf{r}}_1 \times \hat{\mathbf{r}}_2 \tag{4abc}$$

$$\hat{\mathbf{s}}_1' = \hat{\mathbf{W}}_1', \qquad \hat{\mathbf{s}}_2' = \frac{\hat{\mathbf{W}}_1' \times \hat{\mathbf{W}}_2'}{|\hat{\mathbf{W}}_1' \times \hat{\mathbf{W}}_2'|}, \qquad \hat{\mathbf{s}}_3' = \hat{\mathbf{s}}_1' \times \hat{\mathbf{s}}_2' \tag{5abc}$$

(and similarly for $\hat{\mathbf{s}}_1^{\text{r.v.}}$, $\hat{\mathbf{s}}_2^{\text{r.v.}}$, and $\hat{\mathbf{s}}_3^{\text{r.v.}}$ in terms of $\hat{\mathbf{W}}_1^{\text{r.v.}}$ and $\hat{\mathbf{W}}_2^{\text{r.v.}}$). The TRIAD attitude estimate is then

$$A^{*\prime} = [\,\hat{\mathbf{s}}_1' \quad \hat{\mathbf{s}}_2' \quad \hat{\mathbf{s}}_3'\,][\,\hat{\mathbf{r}}_1 \quad \hat{\mathbf{r}}_2 \quad \hat{\mathbf{r}}_3\,]^{\mathrm{T}} \tag{6}$$

and the TRIAD attitude estimator

$$A^{*} = [\,\hat{\mathbf{s}}_1^{\text{r.v.}} \quad \hat{\mathbf{s}}_2^{\text{r.v.}} \quad \hat{\mathbf{s}}_3^{\text{r.v.}}\,][\,\hat{\mathbf{r}}_1 \quad \hat{\mathbf{r}}_2 \quad \hat{\mathbf{r}}_3\,]^{\mathrm{T}} \tag{7}$$

The TRIAD attitude estimate satisfies $\hat{\mathbf{W}}_1' = A^{*\prime}\hat{\mathbf{V}}_1$ exactly and would satisfy $\hat{\mathbf{W}}_2' = A^{*\prime}\hat{\mathbf{V}}_2$ for $\Delta\hat{\mathbf{W}}_2' = \mathbf{0}$. Hence, we usually choose $\hat{\mathbf{W}}_1'$ to be the measured direction of greater accuracy.

The TRIAD algorithm is a deterministic algorithm (no figure of merit has been optimized). Hence, it must compute the attitude by first discarding data beyond what it requires, accomplished in this case by the construction of the ancillary triads. We will examine these and the discarded data in more detail below.

Deterministic Estimates as Optimal Estimates

Let $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]$ be a parameter vector of dimension n and let z_k, $k = 1, \ldots, N > n$, be a set of N scalar random measurements satisfying[5]

$$z_k = f_k(\boldsymbol{\theta}^{\text{true}}) + v_k, \qquad k = 1, \ldots, N \tag{8}$$

where the v_k, $k = 1, \ldots, N$, are zero-mean random noise. Since $N > n$, there is no solution, in general, to equation (8), because $\boldsymbol{\theta}$ is overdetermined. We can make the solution determinate[6] by choosing n of the N measurements (say z_k', $k = 1, \ldots, n$, if the Jacobian determinant $|\partial(f_1, \ldots, f_n)/\partial(\theta_1, \ldots, \theta_n)|$ does not vanish) and solving the corresponding equations

$$z_k' = f_k(\boldsymbol{\theta}^{*\prime}), \qquad k = 1, \ldots, n \tag{9}$$

for the estimate $\boldsymbol{\theta}^{*\prime}$.[7] This is a *deterministic* algorithm for $\boldsymbol{\theta}^{*\prime}$.

If the measurement noise is Gaussian, and R, the covariance matrix of $[\,v_1, v_2, \ldots, v_n\,]$, is invertible, then we can write the cost function as

[5]Rather than overburden our notation we do not place primes on z_k or on v_k, which are always understood to denote random variables.

[6]By "determinate" we mean here that the solution has no degeneracies. The antonym of "determinate" is "indeterminate." A solution may be indeterminate due either to continuous or to discrete degeneracies. Among the determinate solutions, we may distinguish "deterministic" solutions, which use a minimum number of data to arrive at a unique attitude solution, or "optimal" solutions, which use more data than are necessary and minimize some criterion. Most often this criterion is of a statistical nature (minimum variance, maximum likelihood, maximum entropy, etc.). We will frequently, following common practice, use "deterministic" to mean determinate from a minimal data set but for a discrete degeneracy, as in reference [13].

[7]Likewise, we could solve the same equations with z_k' replaced by z_k to obtain the estimator $\boldsymbol{\theta}^{*}$.

$$J_\theta(\boldsymbol{\theta}) = \frac{1}{2}[\mathbf{z}' - \mathbf{f}(\boldsymbol{\theta})]^{\mathrm{T}} R^{-1} [\mathbf{z}' - \mathbf{f}(\boldsymbol{\theta})] \tag{10}$$

with $\mathbf{z}' = [z_1', \ldots, z_n']^{\mathrm{T}}$ the concatenated $n \times 1$ measurement vector, and $\mathbf{f}(\boldsymbol{\theta}) = [f_1(\boldsymbol{\theta}), \ldots, f_n(\boldsymbol{\theta})]^{\mathrm{T}}$ the corresponding concatenated $n \times 1$ measurement function. Then $\boldsymbol{\theta}^{*\prime}$, the optimizing value of $\boldsymbol{\theta}$ for equation (10), is the *optimal* estimate of $\boldsymbol{\theta}$ given the measurements $z_1', \ldots, z_n'$ and R. It is also the deterministic estimate, since the cost function will vanish exactly at the deterministic solution. If the measurement noise is Gaussian, then the $\boldsymbol{\theta}^{*\prime}$ which minimizes $J_\theta(\boldsymbol{\theta})$ is the maximum-likelihood estimate and also satisfies equation (9). Voilà![8]

The "optimal" character of the deterministic estimate is thus a very banal property of the latter. But it is not a property without consequences, because for zero-mean Gaussian noise we know that the inverse estimate-error covariance matrix $P_{\theta\theta}^{-1}$ associated with the maximum-likelihood estimate is just the Fisher information matrix, which we write approximately as[9]

$$P_{\theta\theta}^{-1} = \frac{\partial \mathbf{f}^{\mathrm{T}}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}(\boldsymbol{\theta})\, R^{-1} \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}) \tag{11}$$

or equivalently

$$[P_{\theta\theta}^{-1}]_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} \frac{\partial f_k}{\partial \theta_i}(\boldsymbol{\theta})\, [R^{-1}]_{kl} \frac{\partial f_l}{\partial \theta_j}(\boldsymbol{\theta}) \tag{12}$$

which, theoretically, we should evaluate at the true value of $\boldsymbol{\theta}$ but which, in practice, we must evaluate at $\boldsymbol{\theta}^{*\prime}$. This result for the Fisher information matrix was used without comment or justification in reference [13].

Clearly, then, we can write the TRIAD algorithm equivalently as

$$A_{\mathrm{TRIAD}}^{*\prime} = \arg\min_{A} \left\{ \frac{1}{2} \sum_{k=1}^{3} a_k |\hat{\mathbf{s}}_k' - A\, \hat{\mathbf{r}}_k|^2 \right\} \tag{13}$$

for a_k, $k = 1, 2, 3$, arbitrary but positive, or in an infinitude of other forms yielding the same result, but that does not help us construct a useful expression for the TRIAD covariance matrix. For one thing the three measurement triad vectors are correlated with one another (see equation (5c)), so that although the appropriate maximum-likelihood cost function may be quadratic, it will certainly not have the form inside the braces of equation (13). The task of much of the remainder of this work will be to construct the appropriate cost function.

The TRIAD Algorithm: Statistical Analysis

The sampled measurements $\hat{\mathbf{W}}_1'$ and $\hat{\mathbf{W}}_2'$ are unit vectors and satisfy, therefore

$$\hat{\mathbf{W}}_i' \cdot \Delta\hat{\mathbf{W}}_i' = 0 + O(|\Delta\hat{\mathbf{W}}_i'|^2), \qquad i = 1, 2 \tag{14}$$

and, with very large probability

$$|\Delta\hat{\mathbf{W}}_i'| \ll 1 \tag{15}$$

[8]The value of this approach may be overstated, since it is seldom more difficult to treat all of the data in a least-square estimation problem than a subset which makes the algorithm deterministic. For our analysis of the TRIAD algorithm, practicality is not the overriding issue but understanding.

[9]We use the convention that the derivative of a scalar with respect to a column vector is a row vector.

For a sensor of accuracy 10 arc seconds

$$E\{|\Delta\hat{\mathbf{W}}_i^{\text{r.v.}}|^2\} \lesssim 10^{-3}\ \text{arcsec} \tag{16}$$

in angle-equivalent error. Thus, the zero-mean component of $\hat{\mathbf{W}}_i^{\text{r.v.}}$, $i = 1, 2$, lies essentially in the plane perpendicular to $\hat{\mathbf{W}}_i^{\text{true}}$, $i = 1, 2$. If the coordinate axes of this nonrandom plane are $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{b}}_i$, $i = 1, 2$, then the four random components in the two measurements are to $O(\sigma^2)$

$$\hat{\mathbf{a}}_i \cdot \hat{\mathbf{W}}_i^{\text{r.v.}} \quad \text{and} \quad \hat{\mathbf{b}}_i \cdot \hat{\mathbf{W}}_i^{\text{r.v.}} \qquad i = 1, 2 \tag{17a}$$

For completeness we define as well the "constraint" measurements

$$\hat{\mathbf{c}}_i \cdot \hat{\mathbf{W}}_i^{\text{r.v.}} \equiv \hat{\mathbf{W}}_i^{\text{true}} \cdot \hat{\mathbf{W}}_i^{\text{r.v.}} \approx 1 \qquad i = 1, 2 \tag{17b}$$

where $\{\hat{\mathbf{a}}_i, \hat{\mathbf{b}}_i, \hat{\mathbf{c}}_i\}$, $i = 1, 2$, is a nonrandom righthand orthonormal triad. The scalars presented in equation (17b) are not truly measurements, since they are in the direction of the unit-norm constraint (heuristically, we could imagine them as being measurements with zero measurement error), but they will be very helpful in understanding the TRIAD algorithm. Thus, we may decompose the two unit-vector measurements $\hat{\mathbf{W}}_1^{\text{r.v.}}$ and $\hat{\mathbf{W}}_2^{\text{r.v.}}$ into six scalar measurements[10]

$$z_1 = \hat{\mathbf{a}}_1 \cdot \hat{\mathbf{W}}_1^{\text{r.v.}}, \quad z_2 = \hat{\mathbf{b}}_1 \cdot \hat{\mathbf{W}}_1^{\text{r.v.}}, \quad z_3 = \hat{\mathbf{c}}_1 \cdot \hat{\mathbf{W}}_1^{\text{r.v.}} \tag{18a}$$

$$z_4 = \hat{\mathbf{a}}_2 \cdot \hat{\mathbf{W}}_2^{\text{r.v.}}, \quad z_5 = \hat{\mathbf{b}}_2 \cdot \hat{\mathbf{W}}_2^{\text{r.v.}}, \quad z_6 = \hat{\mathbf{c}}_2 \cdot \hat{\mathbf{W}}_2^{\text{r.v.}} \tag{18b}$$

Note that the true values of z_1, z_2, z_4, and z_5 are zero and that the true values of z_3 and z_6 are unity. Thus, besides the (nonrandom) constraint measurements, there are four (random) equivalent scalar measurements, the "true" scalar measurements (not to be confused with the *true values* of the scalar measurements), available for the estimation of the attitude, of which a deterministic algorithm can use only three. We have still to specify $\hat{\mathbf{a}}_1$ and $\hat{\mathbf{a}}_2$, from which the specification of $\hat{\mathbf{b}}_1$ and $\hat{\mathbf{b}}_2$ will follow.

It is possible to construct the TRIAD attitude matrix given only two of $\hat{\mathbf{s}}_1'$, $\hat{\mathbf{s}}_2'$, and $\hat{\mathbf{s}}_3'$, since the remaining direction can be constructed from the other two using the vector product. The first measured direction $\hat{\mathbf{W}}_1'$ enters the TRIAD algorithm whole as $\hat{\mathbf{s}}_1'$. The triad vector $\hat{\mathbf{s}}_3'$, however, satisfies

$$\hat{\mathbf{s}}_3' = -\text{unit}(\hat{\mathbf{W}}_2' - (\hat{\mathbf{W}}_1' \cdot \hat{\mathbf{W}}_2')\hat{\mathbf{W}}_1') \tag{19}$$

where unit$(\cdot)$ is the function which unitizes a row or column array. Thus, apart from the adjusted norm and the overall sign, $\hat{\mathbf{s}}_3'$ consists of $\hat{\mathbf{W}}_2'$ from which the component along $\hat{\mathbf{W}}_1'$ has been removed. This is the truncation necessary to create a deterministic algorithm from a surfeit of data. We recognize in equation (19) the application of the Gram-Schmidt orthogonalization, as has been remarked earlier in an examination of the TRIAD algorithm in three *or more* dimensions [14]. The transposition of the TRIAD problem to two dimensions is also of interest [15].

As a result of these operations, the random part of the unit random vector $\hat{\mathbf{s}}_3^{\text{r.v.}}$ cannot have any component along either $\hat{\mathbf{W}}_1^{\text{r.v.}} = \hat{\mathbf{s}}_1^{\text{r.v.}}$, hence not, to order σ_2^2, along

[10]For convenience, we do not write the superscript "true" on $\hat{\mathbf{a}}_i$, $\hat{\mathbf{b}}_i$, $\hat{\mathbf{c}}_i$, $i = 1, 2$, which are always nonrandom. Naturally, in a calculation with real data, we would need to replace these column vectors with $\{\hat{\mathbf{a}}_i', \hat{\mathbf{b}}_i', \hat{\mathbf{c}}_i'\}$, $i = 1, 2$, which are calculated analogously from $\hat{\mathbf{W}}_i'$, $i = 1, 2$.

$\hat{\mathbf{s}}_1^{\text{true}}$, or along $\hat{\mathbf{s}}_3^{\text{true}}$. To order σ_2^2 it can have a single random component only along $\hat{\mathbf{s}}_2^{\text{true}}$, which we define to be $\hat{\mathbf{a}}_2$. Following this example, we define $\hat{\mathbf{a}}_1$ to be $\hat{\mathbf{s}}_2^{\text{true}}$ as well. This is always possible, since the only requirement on $\hat{\mathbf{a}}_1$ and $\hat{\mathbf{a}}_2$ is that they be perpendicular to $\hat{\mathbf{W}}_1^{\text{true}}$ and $\hat{\mathbf{W}}_2^{\text{true}}$, respectively. Hence, $\hat{\mathbf{b}}_1 = \hat{\mathbf{W}}_1^{\text{true}} \times \hat{\mathbf{a}}_1 = \hat{\mathbf{W}}_1^{\text{true}} \times \hat{\mathbf{s}}_2^{\text{true}} = \hat{\mathbf{s}}_3^{\text{true}}$ and $\hat{\mathbf{b}}_2 = \hat{\mathbf{W}}_2^{\text{true}} \times \hat{\mathbf{a}}_2 = \hat{\mathbf{W}}_2^{\text{true}} \times \hat{\mathbf{s}}_2^{\text{true}} \equiv \hat{\mathbf{s}}_4^{\text{true}}$. Thus, in summary

$$\hat{\mathbf{a}}_1 = \hat{\mathbf{s}}_2^{\text{true}}, \quad \hat{\mathbf{b}}_1 = \hat{\mathbf{W}}_1^{\text{true}} \times \hat{\mathbf{s}}_2^{\text{true}} = \hat{\mathbf{s}}_3^{\text{true}}, \quad \hat{\mathbf{c}}_1 = \hat{\mathbf{W}}_1^{\text{true}} \tag{20a}$$

$$\hat{\mathbf{a}}_2 = \hat{\mathbf{s}}_2^{\text{true}}, \quad \hat{\mathbf{b}}_2 = \hat{\mathbf{W}}_2^{\text{true}} \times \hat{\mathbf{s}}_2^{\text{true}} = \hat{\mathbf{s}}_4^{\text{true}}, \quad \hat{\mathbf{c}}_2 = \hat{\mathbf{W}}_2^{\text{true}} \tag{20b}$$

On the basis of the above analysis, we take the three equivalent scalar measurements for the TRIAD algorithm to be

$$z_1 = \hat{\mathbf{a}}_1 \cdot \hat{\mathbf{W}}_1^{\text{r.v.}} = \hat{\mathbf{s}}_2^{\text{true}} \cdot \hat{\mathbf{W}}_1^{\text{r.v.}}, \quad z_2 = \hat{\mathbf{b}}_1 \cdot \hat{\mathbf{W}}_1^{\text{r.v.}} = \hat{\mathbf{s}}_3^{\text{true}} \cdot \hat{\mathbf{W}}_1^{\text{r.v.}}, \quad \text{and}$$
$$z_4 = \hat{\mathbf{s}}_2^{\text{true}} \cdot \hat{\mathbf{W}}_2^{\text{r.v.}} \tag{21abc}$$

Thus, we write

$$z_1 = \hat{\mathbf{s}}_2^{\text{trueT}} A^{\text{true}} \hat{\mathbf{V}}_1 + v_1, \quad z_2 = \hat{\mathbf{s}}_3^{\text{trueT}} A^{\text{true}} \hat{\mathbf{V}}_1 + v_2, \quad \text{and}$$
$$z_4 = \hat{\mathbf{s}}_2^{\text{trueT}} A^{\text{true}} \hat{\mathbf{V}}_2 + v_4 \tag{22abc}$$

and

$$v_1 = \hat{\mathbf{s}}_2^{\text{true}} \cdot \Delta\hat{\mathbf{W}}_1^{\text{r.v.}}, \quad v_2 = \hat{\mathbf{s}}_3^{\text{true}} \cdot \Delta\hat{\mathbf{W}}_1^{\text{r.v.}}, \quad \text{and} \quad v_4 = \hat{\mathbf{s}}_2^{\text{true}} \cdot \Delta\hat{\mathbf{W}}_2^{\text{r.v.}} \tag{23abc}$$

which have mean zero and variances given by

$$E\{v_1^2\} = \sigma_1^2, \quad E\{v_2^2\} = \sigma_1^2, \quad \text{and} \quad E\{v_4^2\} = \sigma_2^2 \tag{24abc}$$

One recognizes directly that the three measurement-noise terms for our effective scalar measurements are mutually uncorrelated. In fact, all six effective scalar measurements are uncorrelated, because $\hat{\mathbf{W}}_1^{\text{r.v.}}$ and $\hat{\mathbf{W}}_2^{\text{r.v.}}$ are uncorrelated.

One could also have constructed the third measurement from $\hat{\mathbf{s}}_2'$ in which case one would have obtained

$$z_4^{(2)} = \hat{\mathbf{W}}_2^{\text{true}} \cdot \hat{\mathbf{s}}_2^{\text{r.v.}} = \hat{\mathbf{W}}_2^{\text{trueT}} A\, \hat{\mathbf{r}}_2 + v_4^{(2)} \tag{25}$$

and, like v_4, $v_4^{(2)}$ is uncorrelated with v_1 and v_2, has mean zero and variance σ_2^2. The advantage of this alternate measurement set is that all effective scalar measurements are projections of the triad vectors, emphasizing the statement that the TRIAD attitude matrix is the maximum-likelihood estimate of the attitude given the triad of ancillary vector measurements. We prefer, for greater clarity, to project our scalar measurements from the original measurement vectors, $\hat{\mathbf{W}}_1^{\text{r.v.}}$ and $\hat{\mathbf{W}}_2^{\text{r.v.}}$. Note the relationship of v_4 to $v_4^{(2)}$ in which the character of random vector and non-random vector are interchanged for the two vectors in the scalar product. Likewise, one could have projected the third effective scalar measurement from $\hat{\mathbf{s}}_3$.

The above development suggests that we take as the cost function, i.e., the data-dependent part of the sampled negative-log likelihood function [16], the function

$$J'(A) = \frac{1}{2}\left\{\frac{1}{\sigma_1^2}\left|z_1' - \hat{\mathbf{s}}_2'^{\text{T}} A\, \hat{\mathbf{r}}_1\right|^2 + \frac{1}{\sigma_1^2}\left|z_2' - \hat{\mathbf{s}}_3'^{\text{T}} A\, \hat{\mathbf{r}}_1\right|^2 + \frac{1}{\sigma_2^2}\left|z_4' - \hat{\mathbf{s}}_2'^{\text{T}} A \hat{\mathbf{V}}_2\right|^2\right\} \tag{26}$$

The effective sampled measurement values z_1', z_2', and z_4' will all vanish. This cost function, while instructive, is not quite adequate, as we shall see.

The TRIAD Fisher Information Matrix

Examine a typical term of the cost function of equation (26), which we can write in obvious notation as

$$J(A) = j_1(A) + j_2(A) + j_4(A) \tag{27}$$

Each term has the general (sampled) form

$$j'(A) = \frac{1}{\sigma^2} |z' - \hat{\mathbf{p}}'^{\mathrm{T}} A \, \hat{\mathbf{q}}|^2 \tag{28}$$

Writing[11]

$$A = (I_{3\times3} + [[\Delta\boldsymbol{\theta}]] + [[\Delta\boldsymbol{\theta}]]^2/2) \, A^{\mathrm{true}} + O(|\Delta\boldsymbol{\theta}|^3) \tag{29}$$

equation (28) becomes

$$j(\Delta\boldsymbol{\theta}) = \frac{1}{\sigma^2} |z' - \hat{\mathbf{p}}'^{\mathrm{T}} A^{\mathrm{true}} \hat{\mathbf{q}} + \hat{\mathbf{p}}'^{\mathrm{T}} [[A^{\mathrm{true}} \hat{\mathbf{q}}]] \Delta\boldsymbol{\theta} - (1/2) \Delta\boldsymbol{\theta}^{\mathrm{T}} [[\hat{\mathbf{p}}']] \, [[A^{\mathrm{true}} \hat{\mathbf{q}}]] \Delta\boldsymbol{\theta}|^2 \tag{30}$$

The partial derivatives with respect to $\Delta\boldsymbol{\theta}$ are now carried out easily. The contribution of this term to the Fisher information matrix is then simply (recalling that the Fisher information matrix is evaluated at the true values of the measurements)

$$f_{\theta\theta} = \frac{1}{\sigma^2} (\hat{\mathbf{p}}^{\mathrm{true}} \times (A^{\mathrm{true}} \hat{\mathbf{q}})) \, (\hat{\mathbf{p}}^{\mathrm{true}} \times (A^{\mathrm{true}} \hat{\mathbf{q}}))^{\mathrm{T}} \tag{31}$$

Making the appropriate substitutions in equation (31) now yields

$$\begin{aligned} F_{\theta\theta} = (P_{\theta\theta}^{\mathrm{TRIAD}})^{-1} &= \frac{1}{\sigma_1^2} \hat{\mathbf{s}}_3^{\mathrm{true}} \hat{\mathbf{s}}_3^{\mathrm{trueT}} + \frac{1}{\sigma_1^2} \hat{\mathbf{s}}_2^{\mathrm{true}} \hat{\mathbf{s}}_2^{\mathrm{trueT}} + \frac{1}{\sigma_2^2} \hat{\mathbf{s}}_4^{\mathrm{true}} \, \hat{\mathbf{s}}_4^{\mathrm{trueT}} \\ &= \frac{1}{\sigma_1^2} (I_{3\times3} - \hat{\mathbf{s}}_1^{\mathrm{true}} \hat{\mathbf{s}}_1^{\mathrm{trueT}}) + \frac{1}{\sigma_2^2} \hat{\mathbf{s}}_4^{\mathrm{true}} \hat{\mathbf{s}}_4^{\mathrm{trueT}} \end{aligned} \tag{32}$$

which is the familiar expression [2].[12]

Degeneracy of the Solutions

As pointed out in reference [13], a least-square cost function in terms of three independent arc lengths (or, equivalently, the cosines of three independent arc lengths) will lead to an eight-fold degeneracy in the solutions. While there is no continuous degeneracy of the attitude estimate, there is a discrete degeneracy. In the present case, where the reference directions for two of the arc-length measurements

[11]Since we are going to compute second derivatives with respect to $\Delta\boldsymbol{\theta}$ at $\Delta\boldsymbol{\theta} = \mathbf{0}$, we must retain terms through second order in $\Delta\boldsymbol{\theta}$.
[12]For additional equivalent expressions see reference [17].

are the same, the degeneracy will be only four-fold. Thus, the cost function of equation (26) is insufficient for the determination of a unique attitude estimate. It yields the correct Fisher information matrix, as we have seen, but that is because the Fisher information matrix is the same for each of the four minima (here zeros) of the cost function.

The inference to be drawn here is that we have removed too much from the original measurement model. Obviously, we must remove at least one of the true measurements, or our estimator would yield the QUEST and not the TRIAD attitude solution. But there are six components of $\hat{\mathbf{W}}_1$ and $\hat{\mathbf{W}}_2$, and the remaining two of these, the constraint measurements, must be able to remove the degeneracy.

Consider one of these, namely z_3, for which the contribution to the (sampled) cost function is

$$j_3'(A) = \frac{1}{\sigma_3^2}\,|z_3' - \hat{\mathbf{W}}_1' \cdot A\,\hat{\mathbf{V}}_1|^2 \tag{33}$$

The weight $1/\sigma_3^2$ in $j_3(A)$ is left unspecified for the moment. Heuristically, it would seem to be zero since the constraint provides no statistical information. We shall see that we can choose its value at our convenience.

The quantity $j_3(A)$ will be zero when $\hat{\mathbf{W}}_1' = A^{*\prime\mathrm{TRIAD}}\hat{\mathbf{V}}_1$. Its Fisher information matrix with respect to $\Delta\boldsymbol{\theta}$, calculated analogously to equation (31), will be the expectation of the Hessian matrix whose sampled values are

$$\left.\frac{\partial j_3'(\Delta\boldsymbol{\theta})}{\partial\Delta\boldsymbol{\theta}^{\mathrm{T}}\partial\Delta\boldsymbol{\theta}}\right|_{\Delta\boldsymbol{\theta}=0} = \frac{1}{\sigma_3^2}\,(\hat{\mathbf{W}}_1' \times (A\hat{\mathbf{V}}_1))\,(\hat{\mathbf{W}}_1' \times (A\hat{\mathbf{V}}_1))^{\mathrm{T}} \tag{34}$$

This expression vanishes identically when A is the TRIAD attitude estimate $A^{*\prime\mathrm{TRIAD}}$. Thus, this term does not affect the value of the Fisher information matrix, as should be expected since it contains no statistical information. However, $j_3'(A)$ does make a positive contribution to the cost function when A is different from the TRIAD attitude, so it does affect the degeneracy. Since the Fisher information is not affected, we choose, for convenience, $\sigma_3 = \sigma_1$. This means that we can combine $j_1'(A)$, $j_2'(A)$, $j_3'(A)$, and $j_4'(A)$ in the (sampled) cost function to obtain

$$J'(A) = \frac{1}{\sigma_1^2}\,|\hat{\mathbf{W}}_1' - A\hat{\mathbf{V}}_1|^2 + \frac{1}{\sigma_2^2}\,|z_4' - \hat{\mathbf{s}}_2'^{\mathrm{T}} A\hat{\mathbf{V}}_2|^2 \tag{35}$$

The effective measurements for the estimation consist now of one direction and one arc length, also examined in reference [13]. Hence, the degeneracy of the attitude solution is now only two-fold, as reference [13] has shown. Equation (34) holds analogously for the constraint measurement z_6. Thus, setting $\sigma_6 = \sigma_2$, the sampled cost function combining $j_1'(A)$, $j_2'(A)$, $j_3'(A)$, $j_4'(A)$, and $j_6'(A)$ becomes

$$J^{\mathrm{TRIAD}\prime}(A) = \frac{1}{\sigma_1^2}\,|\hat{\mathbf{W}}_1' - A\hat{\mathbf{V}}_1|^2 + \frac{1}{\sigma_2^2}\,\{|z_4' - \hat{\mathbf{s}}_2'^{\mathrm{T}} A\hat{\mathbf{V}}_2|^2 + |z_6' - \hat{\mathbf{W}}_2'^{\mathrm{T}} A\hat{\mathbf{V}}_2|^2\} \tag{36}$$

and yields the TRIAD attitude unambiguously and with the correct Fisher information matrix.[13]

[13] The construction of the attitude solution would not lead directly to equation (6) but to an equivalent expression much like the SCAD algorithm [12]. Small additional terms will arise in the Fisher information for this solution because of the appearance of $\hat{\mathbf{a}}_2'$ and $\hat{\mathbf{c}}_2'$, but these will be of higher order in σ_2 than the usual terms and always discarded.

The inclusion of these extra terms in the TRIAD cost function is similar to the transformation of negative-log-likelihood function for $\hat{\mathbf{W}}_1'$ and $\hat{\mathbf{W}}_2'$ in reference [3] to obtain the Wahba cost function [18] and the unit-vector filter [19].

The Missing Piece

Of the six components of the two direction measurements only a single component remains, namely

$$z_5 = \hat{\mathbf{s}}_4^{\text{true}} \cdot \hat{\mathbf{W}}_2 = \hat{\mathbf{s}}_4^{\text{trueT}} A \hat{\mathbf{V}}_2 + v_5 \tag{37}$$

This is the only component which cannot be constructed from the three ancillary triad vectors. It is simply not there. It is a simple matter to show that v_5 has mean zero and variance σ_2^2. Following equations (28) through (31), the Fisher information associated with this term is just

$$f_{\theta\theta} = \frac{1}{\sigma_2^2} \hat{\mathbf{s}}_2^{\text{true}} \hat{\mathbf{s}}_2^{\text{trueT}} \tag{38}$$

Adding this term to the Fisher information matrix of equation (32) yields

$$\begin{aligned} F_{\theta\theta} &= (P_{\theta\theta}^{\text{TRIAD}})^{-1} + \frac{1}{\sigma_2^2} \hat{\mathbf{s}}_2^{\text{true}} \hat{\mathbf{s}}_2^{\text{trueT}} \\ &= \frac{1}{\sigma_1^2}(I_{3\times 3} - \hat{\mathbf{W}}_1^{\text{true}} \hat{\mathbf{W}}_1^{\text{trueT}}) + \frac{1}{\sigma_2^2}(I_{3\times 3} - \hat{\mathbf{W}}_2^{\text{true}} \hat{\mathbf{W}}_2^{\text{trueT}}) = (P_{\theta\theta}^{\text{QUEST}})^{-1} \end{aligned} \tag{39}$$

which was to be expected, since $J(A)$ is now the Wahba cost function.

It is instructive to make a table of the results of the estimation process for several of the possible measurement sets treated above. The first column of Table 1 gives the set of measurements to be considered, the second column gives the attitude estimator which results from the application of maximum-likelihood estimation to this measurement set, and the third column gives the multiplicity (degeneracy) of the attitude solution. The table is meant to be illustrative rather than exhaustive. For example, the measurement set $\{z_1, z_2, z_3, z_4, z_6\}$ also yields the QUEST attitude with multiplicity two.

We can summarize the maximum-likelihood cost functions which lead to nondegenerate solutions as

$$J^{\text{TRIAD}\prime}(A) = J'(A|\hat{\mathbf{W}}_1', \hat{\mathbf{s}}_2') = J'(A|\hat{\mathbf{W}}_1', \hat{\mathbf{s}}_3') = J'(A|z_1', z_2', z_3', z_4', z_6') \tag{40a}$$

$$J^{\text{QUEST}\prime}(A) = J'(A|\hat{\mathbf{W}}_1', \hat{\mathbf{W}}_2') = J'(A|z_1', z_2', z_3', z_4', z_5', z_6') \tag{40b}$$

TABLE 1. Estimation Results for Different Scalar Measurement Sets

Measurement Set	Attitude Estimator	Multiplicity
z_1, z_2, z_4	TRIAD	4
z_1, z_2, z_3, z_4	TRIAD	2
z_1, z_2, z_3, z_4, z_6	TRIAD	1
z_1, z_2, z_4, z_5	QUEST	4
z_1, z_2, z_3, z_4, z_5	QUEST	2
$z_1, z_2, z_3, z_4, z_5, z_6$	QUEST	1

Conclusions

We have seen that it is possible to treat deterministic attitude estimators as maximum-likelihood estimators, and this is especially easy to accomplish if the measurement noise is linear and Gaussian. This equivalence with maximum-likelihood estimators is, in fact, a common property of deterministic estimators. Its value is that it allows us an easier way to compute the attitude covariance matrix for the algorithm if the minimal set of measurements can be identified easily. To see that this method is easier to implement for the computation of the attitude covariance matrix of the TRIAD algorithm than the brute-force method, one need only examine the procedures of references [2] and [17]. This method, in fact, was employed without formal justification in reference [13].

We have seen also the connection of the ambiguity of deterministic solutions to the constraint portions of the measurements, not appreciated before reference [13]. While these constraint measurements have no continuous information, they permit the removal of the degeneracy. To a large degree, the present work expands reference [13] by relating the discrete degeneracies to the amputation of the constraint measurements (metronotomy).[14]

In truth, the methods developed here do not have practical value for constructing the attitude estimator. In order to remove the inherent degeneracy of a deterministic attitude solution, we need to construct the constraint measurements. But a constraint measurement like (in its sampled form) $z' = \hat{\mathbf{W}}'^{\mathrm{T}} A \hat{\mathbf{V}} = 1$ is equivalent to $\hat{\mathbf{W}}' = A\hat{\mathbf{V}}$, a vector measurement. For the TRIAD algorithm we know $\hat{\mathbf{W}}'$, but if we are given only scalar measurements from the start, we do not. The only way to eliminate the degeneracy then is to include more "true" measurements, which makes the attitude estimation problem no longer deterministic.

Unless we are fortunate, as in the TRIAD algorithm, to have an algebraic solution method which yields a unique solution, we may have no choice but to calculate all eight possible degenerate "deterministic" solutions and use additional constraints to eliminate the unwanted seven. However, to construct these constraints, in general, we must have more than just the minimal three data. For the construction of the TRIAD solution, as we saw, we really used *five* "data" to remove the degeneracies and, in general, we may require six or more. Some day we may be able to show that in the strictest sense (i.e., free of degeneracy) there really aren't any deterministic solutions at all from only three scalar measurements.

The practical value of the material here thus lies chiefly in the simpler calculation of the attitude covariance matrix rather than the attitude itself. For this operation, certainly, it has real value.

Very likely, the TRIAD algorithm has still further mysteries to reveal.

Acknowledgment

The author is grateful to F. Landis Markley for many helpful suggestions.

References

[1] BLACK, H. D. "A Passive System for Determining the Attitude of a Satellite," *AIAA Journal,* Vol. 2, July 1964, pp. 1350–1351.

[14]Metronotomy has been constructed from the Ancient Greek *μέτρον, τό*, "measurement," and Ancient Greek *τόμη, ἡ*, "cutting (n.)." Thus, we may speak of TRIAD and QUEST as metronotomical and ametronotomical solution algorithms, respectively, of the Wahba problem.

[2] SHUSTER, M. D. and OH, S. D. "Three-Axis Attitude Determination from Vector Observations," *Journal of Guidance and Control,* Vol. 4, No. 1, January-February 1981, pp. 70–77.
[3] SHUSTER, M. D. "Maximum Likelihood Estimation of Spacecraft Attitude," *The Journal of the Astronautical Sciences,* Vol. 37, No. 1, January-March 1989, pp. 79–88.
[4] MARKLEY, F. L. and MORTARI M. "Quaternion Attitude Estimation Using Vector Measurements, *The Journal of the Astronautical Sciences,* Vol. 48, Nos. 2 and 3, April-September 2000, pp. 359–380.
[5] MARKLEY, F. L. "Attitude Determination Using Vector Observations and the Singular Value Decomposition," *The Journal of the Astronautical Sciences,* Vol. 36, No. 3, July-September 1988, pp. 245–258.
[6] MARKLEY, F. L. "Attitude Determination Using Vector Observations: A Fast Optimal Matrix Algorithm," *The Journal of the Astronautical Sciences,* Vol. 41, No. 2, April-June 1993, pp. 261–280.
[7] MORTARI, D. "ESOQ: A Closed-Form Solution to the Wahba Problem," *The Journal of the Astronautical Sciences,* Vol. 45, No. 2, April-June 1997, pp. 195–204.
[8] MORTARI, D. "Second Estimator for the Optimal Quaternion," *Journal of Guidance, Control and Dynamics,* Vol. 23, No. 4, September-October 2000, pp. 885–888.
[9] BRUCCOLERI, C., LEE, D.-J., and MORTARI, D. "Single-Point Optimal Attitude Determination Using Modified Rodrigues Parameters," presented as paper AAS-05-459 at the AAS Astronautics Symposium, Grand Island, New York, 2005; reprinted in *Advances in the Astronautical Sciences,* Vol. 122, 2006, pp. 127–148.
[10] SHUSTER, M. D. "A Survey of Attitude Representations," *The Journal of the Astronautical Sciences,* Vol. 41, No. 4, October-December 1993, pp. 439–517.
[11] WAHBA, G. "Problem 65-1: A Least Squares Estimate of Spacecraft Attitude," *SIAM Review,* Vol. 7, No. 3, July 1965, p. 409.
[12] SHUSTER, M. D. "SCAD—A Fast Algorithm for Star Camera Attitude Determination," *The Journal of the Astronautical Sciences,* Vol. 52, No. 3, July-September 2004, pp. 391–404.
[13] SHUSTER, M. D. "Deterministic Three-Axis Attitude Determination," *The Journal of the Astronautical Sciences,* Vol. 52, No. 3, July-September 2004, pp. 405–419.
[14] SHUSTER, M. D. "Attitude Determination in Higher Dimensions" (Engineering Note), *Journal of Guidance, Control and Dynamics,* Vol. 16, No. 2, March-April 1993, pp. 393–395.
[15] SHUSTER, M. D. "Attitude Analysis in Flatland: The Plane Truth," *The Journal of the Astronautical Sciences,* Vol. 52, Nos. 1 and 2, January-June 2004, pp. 195–209.
[16] SORENSEN, H. W. *Parameter Estimation,* Marcel Dekker, New York 1980.
[17] CHENG, Y. and SHUSTER, M. D. "QUEST and the Anti-QUEST: Good and Evil Attitude Estimation," *The Journal of the Astronautical Sciences,* Vol. 53, No. 3, July-September 2005, pp. 337–351.
[18] SHUSTER, M. D. "Kalman Filtering of Spacecraft Attitude and the QUEST Model," *The Journal of the Astronautical Sciences,* Vol. 38, No. 3, July-September 1990, pp. 377–393; Erratum: Vol. 51, No. 3, July-September 2003, p. 359.
[19] SEDLAK, J. and CHU, D. "Kalman Filter Estimation of Attitude and Gyro Bias with the QUEST Observation Model," presented as paper AAS-93-297 at the AAS/GSFC International Symposium on Space Flight Dynamics, NASA Goddard Space Flight Center, April 26-30, 1993.