The Journal of the Astronautical Sciences, Vol. 41, No. 2, April-June 1993, pp. 261–280

Attitude Determination Using Vector Observations: A Fast Optimal Matrix Algorithm

F. Landis Markley[1]

Abstract

The attitude matrix minimizing Wahba's loss function is computed directly by a method that is either faster or more robust than any previously known algorithm for finding this optimal estimate. Analysis of the special case of two vector observations identifies those cases for which the TRIAD or algebraic method minimizes Wahba's loss function. The new method also provides an estimate of the attitude error covariance matrix, including an especially convenient representation of the two-observation covariance matrix.

Introduction

In 1965, Wahba posed the problem of finding the proper orthogonal matrix A that minimizes the non-negative loss function [1]

$$L(A) \equiv \frac{1}{2} \sum_{i=1}^{n} a_i |\mathbf{b}_i - A\mathbf{r}_i|^2, \tag{1}$$

where the unit vectors $\mathbf{r}_i$ are representations in a reference frame of the directions to some observed objects, the $\mathbf{b}_i$ are the unit vector representations of the corresponding observations in the spacecraft body frame, the a_i are positive weights, and n is the number of observations. The motivation for this loss function is that if the vectors are error-free and the true attitude matrix A_{true} is assumed to be the same for all the measurements, then $\mathbf{b}_i$ is equal to $A_{\text{true}}\mathbf{r}_i$ for all i and the loss function is equal to zero for A equal to A_{true}.

Attitude determination algorithms based on minimizing this loss function have been used for many years [2–6]. The original solutions to Wahba's problem solved for the spacecraft attitude matrix directly [2], but most practical applications have been based on Davenport's q-method [3–5], which solves for the quaternion representing the attitude matrix. In this paper, we present a new method that solves for

[1]Assistant Head, Guidance and Control Branch, Code 712, Goddard Space Flight Center, Greenbelt, MD 20771.

the attitude matrix directly, as well as the covariance matrix, and which is competitive with the well known QUEST algorithm [6] in speed. Examination of test cases uncovers some where the new algorithm is more robust than QUEST. Analysis of the special case of two observations serves to relate this method to the TRIAD or algebraic method [5–7], and also leads to a convenient representation of the two-observation covariance matrix.

Statement of the Problem

Simple matrix manipulations transform the loss function into

$$L(A) = \lambda_0 - \mathrm{tr}(AB^{\mathrm{T}}), \tag{2}$$

where

$$\lambda_0 \equiv \sum_{i=1}^{n} a_i, \tag{3}$$

$$B \equiv \sum_{i=1}^{n} a_i \mathbf{b}_i \mathbf{r}_i^{\mathrm{T}}, \tag{4}$$

tr denotes the trace, and the superscript T denotes the matrix transpose. Thus Wahba's problem is equivalent to the problem of finding the proper orthogonal matrix A that maximizes the trace of the matrix product AB^{T}. The weights are often chosen so that $\lambda_0 = 1$, but this is not always the most convenient choice, as will be discussed below.

This optimization problem has an interesting relation to a matrix norm. The Euclidean norm (also known as the Schur, Frobenius, or Hilbert-Schmidt norm) is defined for a general real matrix M by [8, 9]

$$\|M\|^2 \equiv \sum M_{ij}^2 = \mathrm{tr}(MM^{\mathrm{T}}), \tag{5}$$

where the sum is over all the matrix elements. The assumed orthogonality of A and properties of the trace give

$$\|A - B\|^2 = \mathrm{tr}[(A - B)(A - B)^{\mathrm{T}}] = \mathrm{tr}\, I - 2\mathrm{tr}(AB^{\mathrm{T}}) + \|B\|^2, \tag{6}$$

where I is the 3×3 identity matrix. The orthogonal matrix A that maximizes $\mathrm{tr}(AB^{\mathrm{T}})$ minimizes this norm, so Wahba's problem is also equivalent to the problem of finding the proper orthogonal matrix A that is closest to B in the Euclidean norm [10]. It is also related to the problem of finding a "procrustean transformation" of B [8, 9].

The matrix B can be shown to have the decomposition [11]

$$B = U_+ \,\mathrm{diag}[S_1, S_2, S_3] V_+^{\mathrm{T}} \tag{7}$$

where U_+ and V_+ are proper orthogonal matrices; diag[. . .] denotes a matrix with the indicated elements on the main diagonal and zeros elsewhere; and S_1, S_2, and $|S_3|$, the singular values of B, obey the inequalities

$$S_1 \geq S_2 \geq |S_3|. \tag{8}$$

The optimal attitude estimate is given in terms of these matrices by [11]

$$A_{\text{opt}} = U_+ V_+^{\mathrm{T}} . \tag{9}$$

Equation (7) differs from the singular value decomposition (SVD) [8,9] in that U_+ and V_+ are required to have positive determinants. In reference [11], S_3 was denoted by ds_3, where $d = \pm 1$ and $s_3 \geq 0$.

The SVD provides a robust method for computing the matrices U_+ and V_+, and thus the optimal attitude estimate, but it is not very efficient [11]. The purpose of this paper is to present a more efficient method to estimate the attitude.

Computation of the Attitude Matrix

The decomposition of equation (7) allows the Euclidean norm, determinant, and adjoint of B to be written as

$$\|B\|^2 = S_1^2 + S_2^2 + S_3^2 , \tag{10}$$

$$\det B = S_1 S_2 S_3 , \tag{11}$$

and

$$\operatorname{adj} B^{\mathrm{T}} = U_+ \operatorname{diag}[S_2 S_3, S_3 S_1, S_1 S_2] V_+^{\mathrm{T}} . \tag{12}$$

These quantities, as well as the product

$$BB^{\mathrm{T}}B = U_+ \operatorname{diag}[S_1^3, S_2^3, S_3^3] V_+^{\mathrm{T}} \tag{13}$$

can also be evaluated without performing the singular value decomposition, but the representations in terms of the singular values are useful in deriving the matrix identities in this paper. In particular, they can be used to show that the optimal attitude estimate is given by

$$A_{\text{opt}} = [(\kappa + \|B\|^2)B + \lambda \operatorname{adj} B^{\mathrm{T}} - BB^{\mathrm{T}}B]/\zeta , \tag{14}$$

where

$$\kappa \equiv S_2 S_3 + S_3 S_1 + S_1 S_2 , \tag{15}$$

$$\lambda \equiv S_1 + S_2 + S_3 , \tag{16}$$

and

$$\zeta \equiv (S_2 + S_3)(S_3 + S_1)(S_1 + S_2) . \tag{17}$$

The matrices in equation (14) can be computed without performing the singular value decomposition, but this equation is an improvement over equation (9) only because the scalar coefficients κ, λ, and ζ can also be computed without the SVD, as we will show below.

Iterative Computation of the Scalar Coefficients

We first find expressions for the other scalar coefficients in terms of λ. A little algebra shows that

$$\kappa = \frac{1}{2}(\lambda^2 - \|B\|^2) \tag{18}$$

and

$$\zeta = \kappa\lambda - \det B . \tag{19}$$

Let $A(\lambda)$ denote the expression for the attitude matrix given by equations (14), (18), and (19) as a function of λ and B. This is equal to $A_{\rm opt}$ if λ is given by equation (16). Equations (7), (9), and (16) give

$$\lambda = \operatorname{tr}(A_{\rm opt}B^{\rm T}) , \tag{20}$$

so λ can be computed as a solution of the equation

$$0 = \lambda - \operatorname{tr}[A(\lambda)B^{\rm T}] = \lambda - \operatorname{tr}[(\kappa + \|B\|^2)BB^{\rm T} + \lambda(\det B)I - (BB^{\rm T})^2]/\zeta . \tag{21}$$

Multiplication by 2ζ, substitution of equations (18) and (19), and use of the identity

$$\begin{aligned}\|B\|^4 - \operatorname{tr}[(BB^{\rm T})^2] &= (S_1^2 + S_2^2 + S_3^2)^2 - (S_1^4 + S_2^4 + S_3^4)\\ &= 2(S_2^2S_3^2 + S_3^2S_1^2 + S_1^2S_2^2) = 2\|\operatorname{adj} B\|^2\end{aligned} \tag{22}$$

let us write this as

$$0 = p(\lambda) \equiv (\lambda^2 - \|B\|^2)^2 - 8\lambda \det B - 4\|\operatorname{adj} B\|^2. \tag{23}$$

It can be shown that the quartic polynomial $p(\lambda)$ is the same polynomial that is used in QUEST, although the explicit form of the coefficients is different. Substitution of equations (10), (11), and (12) into equation (23) gives the four roots of the quartic in terms of S_1, S_2, and S_3:

$$\begin{aligned}p(\lambda) = {}&(\lambda - S_1 - S_2 - S_3)(\lambda - S_1 + S_2 + S_3)\\ &\times (\lambda + S_1 - S_2 + S_3)(\lambda + S_1 + S_2 - S_3) .\end{aligned} \tag{24}$$

This form is useful for analysis, but not for computation, since the singular values are not known in the iterative method before the appropriate root of the quartic has been found. The roots are all real, and they are the four eigenvalues of the K matrix in the q-method, as is well known [4, 6]. Equation (16) and the inequalities of the singular values expressed by equation (8) show that we require the maximum root. This root is distinct unless $S_2 + S_3 = 0$, in which case the attitude solution is not unique, as is discussed in Markley [11]. In the method introduced in this paper, $S_2 + S_3 = 0$ gives $\zeta = 0$, and all the elements of $A_{\rm opt}$ have the indefinite form 0/0.

We now note from equations (2) and (20) that

$$L(A_{\rm opt}) = \lambda_0 - \lambda \geq 0 , \tag{25}$$

which also shows that choosing the largest root of $p(\lambda)$ minimizes the loss function. For small measurement errors, the loss function should be close to zero, so the maximum root of equation (23) should be close to λ_0 [6]. Thus we can find λ by Newton's method, starting with this value. This defines a sequence of estimates of λ by

$$\lambda_i = \lambda_{i-1} - p(\lambda_{i-1})/p'(\lambda_{i-1}), \qquad i = 1, 2, \ldots , \tag{26}$$

where

$$p'(\lambda) = 8\zeta \tag{27}$$

is the derivative of $p(\lambda)$ with respect to λ. Substitution of equation (24) shows that this sequence would be monotonically decreasing with infinite-precision arithmetic, but a computation with finite-precision arithmetic eventually finds a $\lambda_i \geq \lambda_{i-1}$. At this point, the iterations are terminated and λ_{i-1} is taken to be the desired root to full computer precision. This iteration converges extremely rapidly in practice, except in the case that the maximum root of $p(\lambda)$ is not unique. In that case the derivative in the denominator of equation (26) goes to zero as the root is approached, so the computation is terminated and a warning is issued that the attitude is indeterminate.

Newton's method has quadratic convergence. Higher-order methods, such as Halley's method [12], would give convergence in fewer iterations, but would require more computations per iteration. Since convergence with Newton's method is quite rapid, higher-order methods were not investigated further.

It is important to carry out the computation of λ to full machine precision, since otherwise the computed attitude matrix will not be orthogonal. Straightforward but tedious matrix computation using the Cayley-Hamilton theorem [9] for the matrix BB^{T} gives

$$A(\lambda)A^{\mathrm{T}}(\lambda) = I - (2\zeta)^{-2}p(\lambda)\,(\lambda^2 I - BB^{\mathrm{T}})\,. \tag{28}$$

This shows the orthogonality of the computed attitude matrix if λ is a root of $p(\lambda)$, and estimates the departure from orthogonality otherwise.

Analytic Computation of the Scalar Coefficients

The scalar coefficients can also be computed as functions of the largest singular value S_1 of B by

$$\kappa = S_1(S_2 + S_3) + S_2 S_3 = S_1(S_2 + S_3) + S_1^{-1} \det B\,, \tag{29}$$

$$\lambda = S_1 + (S_2 + S_3)\,, \tag{30}$$

and

$$\zeta = (\kappa + S_1^2)\,(S_2 + S_3)\,, \tag{31}$$

where

$$S_2 + S_3 = \{S_1^{-2}[\|\mathrm{adj}\ B\|^2 - (S_1^{-1} \det B)^2] + 2S_1^{-1} \det B\}^{1/2}. \tag{32}$$

This form is chosen to avoid near-cancellations in near-singular cases. The largest singular value is found as the positive square root of the largest root of the cubic characteristic equation of the matrix BB^{T} [4]:

$$\begin{aligned} 0 &= (S_1^2)^3 - \mathrm{tr}(BB^{\mathrm{T}})\,(S_1^2)^2 + \mathrm{tr}[\mathrm{adj}(BB^{\mathrm{T}})]S_1^2 - \det(BB^{\mathrm{T}}) \\ &= (S_1^2)^3 - \|B\|^2(S_1^2)^2 + \|\mathrm{adj}\ B\|^2 S_1^2 - (\det B)^2. \end{aligned} \tag{33}$$

The largest root of this equation is given by [4, 13]

$$S_1^2 = \frac{1}{3}\left\{\|B\|^2 + 2\alpha \cos\left[\frac{1}{3}\cos^{-1}(\alpha^{-3}\beta)\right]\right\}, \tag{34}$$

where

$$\alpha \equiv (\|B\|^4 - 3\|\text{adj } B\|^2)^{1/2}, \tag{35}$$

and

$$\beta \equiv \|B\|^6 - (9/2)\|B\|^2\|\text{adj } B\|^2 + (27/2)(\det B)^2. \tag{36}$$

Equations (10), (11), and (12) can be used to show that the argument of the square root in equation (35) is greater than or equal to zero, with equality if and only if $S_1 = S_2 = |S_3|$, in which case $\beta = 0$ also. Thus we have a complete analytic solution of Wahba's problem, which would appear to be preferable to the iterative method. In practice, however, the five transcendental function evaluations required by equations (29) to (36) result in a slower algorithm without providing any additional accuracy, as shown by tests below.

Computation of the Covariance Matrix

The quality of the attitude estimate is best expressed in terms of the covariance of the three-component column vector $\boldsymbol{\phi}$ of attitude error angles in the spacecraft body frame. This parameterization gives the following relation between the estimated and true attitude matrices A and A_{true}:

$$A = \{\exp[(-\boldsymbol{\phi})\times]\}A_{\text{true}} = \{I - [\boldsymbol{\phi}\times] + \frac{1}{2}[\boldsymbol{\phi}\times]^2 + \ldots\}A_{\text{true}}, \tag{37}$$

where the matrix $[\mathbf{u}\times]$ is defined for a general three-component column vector $\mathbf{u}$ as

$$[\mathbf{u}\times] \equiv \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}. \tag{38}$$

This notation reflects the equality of the matrix product $[\mathbf{u}\times]\mathbf{v}$ and the cross product $\mathbf{u} \times \mathbf{v}$.

Shuster [14] has recast the Wahba problem as a maximum likelihood estimation problem, which leads to a very convenient method for computing the covariance matrix. Asymptotically, as the amount of data becomes infinite, the covariance matrix tends to the inverse of the Fisher information matrix F, which is the expected value of the Hessian of the negative-log-likelihood function J [15],

$$F_{jk} \equiv \mathrm{E}[\partial^2 J/\partial\phi_j\partial\phi_k]. \tag{39}$$

The distribution of the components of the ith measurement error vector perpendicular to the true vector are assumed to be Gaussian and uniformly distributed in phase about the true vector with variance σ_i^2 per axis [11, 14]. These variances of unit vectors can be interpreted as angular variances in radians. Then the negative-log-likelihood function for this problem is

$$J = \frac{1}{2}\sum_{i=1}^{n} \sigma_i^{-2}|\mathbf{b}_i - A\mathbf{r}_i|^2 + \ldots, \tag{40}$$

where the omitted terms are independent of attitude. This error distribution is only an approximation for real sensors, and the amount of data is always finite, but these approximations only affect the covariance estimate and not the attitude estimate. The resulting covariance estimate has been found to be completely adequate in applications [14].

For any positive λ_0 and with

$$\sigma_{\text{tot}}^2 \equiv \left(\sum_{i=1}^{n} \sigma_i^{-2}\right)^{-1}, \tag{41}$$

the weights

$$a_i = \lambda_0 \sigma_{\text{tot}}^2 / \sigma_i^2 \tag{42}$$

are positive and satisfy equation (3). With this choice

$$J = \lambda_0^{-1} \sigma_{\text{tot}}^{-2} L(A) + \ldots, \tag{43}$$

where the ellipsis has the same significance as in equation (40). Equation (43) shows that the solution to Wahba's problem with these weights is a maximum-likelihood estimate, since it minimizes the negative-log-likelihood function. Substituting equation (37) into equation (2) and using the identity

$$[\mathbf{u}\times][\mathbf{v}\times] = -(\mathbf{v}^{\mathrm{T}}\mathbf{u})I + \mathbf{v}\mathbf{u}^{\mathrm{T}} \tag{44}$$

gives, to second order,

$$\begin{aligned} L(A) &= \lambda_0 - \mathrm{tr}(A_{\text{true}}B^{\mathrm{T}}) + \mathrm{tr}\{[\boldsymbol{\phi}\times]A_{\text{true}}B^{\mathrm{T}}\} + \frac{1}{2}\,\mathrm{tr}\{[(\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\phi})I - \boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}]A_{\text{true}}B^{\mathrm{T}}\} \\ &= \lambda_0 - \mathrm{tr}(A_{\text{true}}B^{\mathrm{T}}) + \mathrm{tr}\{[\boldsymbol{\phi}\times]A_{\text{true}}B^{\mathrm{T}}\} + \frac{1}{2}\,\boldsymbol{\phi}^{\mathrm{T}}[\mathrm{tr}(A_{\text{true}}B^{\mathrm{T}})I - A_{\text{true}}B^{\mathrm{T}}]\boldsymbol{\phi}. \end{aligned} \tag{45}$$

Inserting this into equation (43) gives the attitude-dependent part of the negative-log-likelihood function. Only the part quadratic in $\boldsymbol{\phi}$ contributes to the Fisher information matrix

$$F = \lambda_0^{-1}\sigma_{\text{tot}}^{-2}[\mathrm{tr}(A_{\text{true}}B^{\mathrm{T}})I - \frac{1}{2}(A_{\text{true}}B^{\mathrm{T}} + BA_{\text{true}}^{\mathrm{T}})]. \tag{46}$$

Matrix inversion then gives the covariance matrix

$$P = \lambda_0\sigma_{\text{tot}}^{2}[\mathrm{tr}(A_{\text{true}}B^{\mathrm{T}})I - \frac{1}{2}(A_{\text{true}}B^{\mathrm{T}} + BA_{\text{true}}^{\mathrm{T}})]^{-1}. \tag{47}$$

The true attitude matrix is not known in a real attitude estimation problem, of course, so A_{opt} must be used in place of A_{true} in computing the covariance. Making this replacement in equation (47) gives, with equation (20) and the symmetry of the matrix product $A_{\text{opt}}B^{\mathrm{T}}$, which follows from equations (7) and (9),

$$P = \lambda_0\sigma_{\text{tot}}^2(\lambda I - A_{\text{opt}}B^{\mathrm{T}})^{-1} = \lambda_0\sigma_{\text{tot}}^2\,\mathrm{adj}(\lambda I - A_{\text{opt}}B^{\mathrm{T}})/\det(\lambda I - A_{\text{opt}}B^{\mathrm{T}}). \tag{48}$$

Equation (48) is one of the forms for the covariance matrix given in Appendix B of Markley [11], which is also the result obtained in Markley [16], simplified to the case that only the attitude is estimated. The computation of the matrix inverse can be avoided as follows [17]. Equations (7), (9), and (16) show that

$$\lambda I - A_{\mathrm{opt}} B^{\mathrm{T}} = U_+ \operatorname{diag}[S_2 + S_3, S_3 + S_1, S_1 + S_2] U_+^{\mathrm{T}}. \tag{49}$$

The determinant of this matrix is given by equation (17) as

$$\det(\lambda I - A_{\mathrm{opt}} B^{\mathrm{T}}) = \zeta, \tag{50}$$

and its adjoint is given by equations (7), (15), and (49) as

$$\begin{aligned} \operatorname{adj}(\lambda I - A_{\mathrm{opt}} B^{\mathrm{T}}) &= U_+ \operatorname{diag}[(S_3 + S_1)(S_1 + S_2), (S_1 + S_2)(S_2 + S_3), \\ &\qquad (S_2 + S_3)(S_3 + S_1)] U_+^{\mathrm{T}} \\ &= \kappa I + B B^{\mathrm{T}}. \end{aligned} \tag{51}$$

These yield the desired manifestly symmetric result

$$P = \lambda_0 \sigma_{\mathrm{tot}}^2 (\kappa I + B B^{\mathrm{T}})/\zeta. \tag{52}$$

We see that the covariance matrix is infinite when $\zeta = 0$, which agrees with the conditions for indeterminacy of the attitude solution discussed above.

The angular units of σ_i were assumed above to be radians, in which case the dimensionality of the covariance matrix is radians squared. It is often more convenient to allow the σ_i to be specified in arbitrary angular units. In this case, equations (40) and (43) are only true up to a constant multiplicative factor that does not affect the conclusions of this section. The units of the covariance matrix are then the square of the angular units used for σ_i.

Normalization of the Weights

The results above are valid for any positive value of the parameter λ_0, but only two choices are useful:

$$\lambda_0 = 1 \quad \text{(normalized weights)} \tag{53}$$

or

$$\lambda_0 = \sigma_{\mathrm{tot}}^{-2} \quad \text{(unnormalized weights)}. \tag{54}$$

Past treatments of this problem have generally used normalized weights, which give a B matrix with elements of order unity. This is convenient in computations using fixed-point arithmetic, but floating-point arithmetic is an option on virtually all present-day computers. The normalized form may also be useful if the measurement weights are arbitrarily assigned.

The unnormalized form is more natural if the weights are computed in terms of measurement variances, as in equation (42), since the unnormalized weights are just equal to the inverse variances. The unnormalized form also simplifies the computation of the covariance, as shown by equation (52), but this form can potentially lead to numerical problems. The elements of B are of order $\sigma_{\mathrm{tot}}^{-2}$ if the weights are not normalized, which means that $\|\operatorname{adj} B\|^2$ is of order $\sigma_{\mathrm{tot}}^{-8}$. Since σ_{tot}

can be of order 10^{-6} radians for highly accurate sensors, $\|\text{adj } B\|^2$ can be of order 10^{48}, leading to exponent overflow in floating-point representations that do not provide an adequate exponent range. This is not a problem with double-precision arithmetic in conformity with ANSI/IEEE Standard 754-1985 for binary floating-point arithmetic [18], since this standard mandates eleven bits for the exponent, allowing representation of numbers as large as 10^{308}. The Standard Apple Numerical Environment [19] and VAX G_FLOATING [20] double-precision arithmetic employ eleven-bit exponents, but VAX D_FLOATING double-precision arithmetic allots only eight bits for the exponent. This is the same as in IEEE-standard single-precision arithmetic, and allows representation of numbers only as large as 10^{38}. Single-precision arithmetic would lead to exponent overflow problems for measurement variances σ_{tot}^2 less than about 10^{-9}, but double-precision arithmetic is certainly preferred in such cases.

These overflow problems can often be avoided by a suitable choice of the units for σ_i. However, the optimized loss function computed with unnormalized weights in radian units has the nice property of being roughly equal to the number of measurements n. This property is lost when other angular units are used, but it can be recovered by the appropriate rescaling.

Algorithm Implementation

Two forms of the new algorithm, the form with the iterative solution for λ (FOAM—Fast Optimal Attitude Matrix), and the form with the analytic solution for S_1 (SOMA—Slower Optimal Matrix Algorithm), were implemented in double-precision FORTRAN and executed on a DEC VAX 8830 computer. Both were implemented in G_FLOATING arithmetic with unnormalized weights, and FOAM was also implemented with normalized weights in both G_FLOATING and D_FLOATING arithmetic.

The computational flow of FOAM is as follows. The input observation and reference vectors are normalized and λ_0 and the B matrix are computed according to equations (3) and (4), using equations (41), (42), and either (53) or (54) for the weights. Some efficiencies in the normalization process were found and applied to all the algorithms in the tests. The scalars det B, $\|B\|^2$, and $\|\text{adj } B\|^2$ are calculated next, and λ is computed by equations (23) and (26). Then the optimal attitude estimate is found from equation (14), using equations (18) and (19) for κ and ζ, respectively. Finally, the covariance estimate is given by equation (52). This last step is optional; but it is not very expensive, since it largely uses previously computed quantities.

In the case of near-indeterminacy of the attitude estimate, the singular values are approximately $S_1 \approx \lambda, S_2 \approx S_3 \approx 0$ [11], which gives the covariance

$$P \approx \lambda_0 \sigma_{\text{tot}}^2 U_+ \operatorname{diag}[\lambda^2/\zeta, \lambda^{-1}, \lambda^{-1}] U_+^{\text{T}}, \tag{55}$$

where ζ is very small. The iterative solution for λ by equation (26) is terminated if

$$p'(\lambda) = 8\zeta < 8\lambda_0^3 \sigma_{\text{tot}}^2 \phi_{\text{tol}}^{-2}, \tag{56}$$

where ϕ_{tol} is some suitably chosen tolerance, since equation (55) predicts attitude estimation error standard deviations larger than $(\lambda/\lambda_0)\phi_{\text{tol}}$ when this inequality is satisfied. This error can be much less than ϕ_{tol} only if $\lambda \ll \lambda_0$, in which case the attitude estimate is poor because the loss function is large. The angular units used for ϕ_{tol} must be the same as those of σ_i, of course. In practice, ϕ_{tol} can be quite large; a value of 2 radians was used for the test runs below.

The computational flow of SOMA is identical to that of FOAM, except for the computation of λ, κ, and ζ, which are found from equations (34) to (36) and (29) to (32).

Algorithm Test–Accuracy

The new algorithms, FOAM and SOMA, were compared with the SVD method [11] and Shuster's QUEST (QUaternion ESTimation) algorithm [6] for minimizing Wahba's loss function. The SVD method was implemented with unnormalized weights in D_FLOATING arithmetic, and QUEST was implemented with normalized weights in G_FLOATING arithmetic. In addition to the reference and observation vectors and the measurement standard deviations, QUEST requires the input of five control parameters, which were taken as `QUIBBL` = 0.1, `IMETH` = 1 (two parameters used to avoid a singularity for 180 degree rotations), `FIBBL` = 10^{-10} (similar to ϕ_{tol}), `QUACC` = 10^{-8} (a criterion for convergence of the iteration for λ), and `NEWT` = 10 (the maximum number of iterations allowed). The SVD method is expected to be the most accurate, since it uses very well-tested and robust algorithms to make the best possible use of the information contained in the B matrix.

Twelve test cases were analyzed. Each test case was specified by a set of measurement vectors $\mathbf{r}_i$ and measurement standard deviations σ_i. The observation vectors were computed as

$$\mathbf{b}_i = A_{true}\mathbf{r}_i + \mathbf{n}_i , \tag{57}$$

where $\mathbf{n}_i$ is a vector of measurement errors, and

$$A_{true} = \begin{bmatrix} 0.352 & 0.864 & 0.360 \\ -0.864 & 0.152 & 0.480 \\ 0.360 & -0.480 & 0.800 \end{bmatrix}, \tag{58}$$

which has all non-zero matrix elements with exact decimal representations and is otherwise arbitrary. The tests were run both with $\mathbf{n}_i = \mathbf{0}$ and with measurement errors simulated by zero-mean Gaussian white noise on the components of $\mathbf{n}_i$. The specified measurement standard deviations in each case were used to compute the measurement weights and also the level of simulated measurement errors.

The twelve test cases were specified as follows:

Case 1 used the three reference vectors

$$\mathbf{r}_1 = [1,0,0]^T, \qquad \mathbf{r}_2 = [0,1,0]^T, \qquad \mathbf{r}_3 = [0,0,1]^T, \tag{59}$$

with measurement standard deviations $\sigma_1 = \sigma_2 = \sigma_3 = 10^{-6}$ rad. This reference vector set models three fine sensors with orthogonal boresights along the body axes.

Case 2 used the two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ from the above set with $\sigma_1 = \sigma_2 = 10^{-6}$ rad.

Case 3 was the same as case 1 but with $\sigma_1 = \sigma_2 = \sigma_3 = 0.01$ rad, modeling three orthogonal coarse sensors.

Case 4 was the same as case 2 but with $\sigma_1 = \sigma_2 = 0.01$ rad.

Case 5 used the two reference vectors

$$\mathbf{r}_1 = [0.6, 0.8, 0]^{\mathrm{T}}, \qquad \mathbf{r}_2 = [0.8, -0.6, 0]^{\mathrm{T}}, \tag{60}$$

with $\sigma_1 = 10^{-6}$ rad and $\sigma_2 = 0.01$ rad. This models one fine and one coarse sensor with orthogonal boresights not along the spacecraft body axes.

Case 6 used the three reference vectors

$$\mathbf{r}_1 = [1, 0, 0]^{\mathrm{T}}, \qquad \mathbf{r}_2 = [1, 0.01, 0]^{\mathrm{T}}, \qquad \mathbf{r}_3 = [1, 0, 0.01]^{\mathrm{T}}, \tag{61}$$

with $\sigma_1 = \sigma_2 = \sigma_3 = 10^{-6}$ rad. This reference vector set models three star measurements in a single star sensor with a small field-of-view.

Case 7 used the two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ from equation (61) with $\sigma_1 = \sigma_2 = 10^{-6}$ rad.

Case 8 was the same as case 6 but with $\sigma_1 = \sigma_2 = \sigma_3 = 0.01$ rad, modeling a star sensor with large errors, to stress the algorithms.

Case 9 was the same as case 7 but with $\sigma_1 = \sigma_2 = 0.01$ rad.

Case 10 used the three references vectors

$$\mathbf{r}_1 = [1, 0, 0]^{\mathrm{T}}, \qquad \mathbf{r}_2 = [0.96, 0.28, 0]^{\mathrm{T}}, \qquad \mathbf{r}_3 = [0.96, 0, 0.28]^{\mathrm{T}}. \tag{62}$$

with $\sigma_1 = 10^{-6}$ rad and $\sigma_2 = \sigma_3 = 0.01$ rad. This models one fine sensor with its boresight along the body x-axis and two less accurate reference vectors 16.26 degrees off this axis.

Case 11 used the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ from equation (62) with $\sigma_1 = 10^{-6}$ rad and $\sigma_2 = 0.01$ rad.

Case 12 was the same as case 11 but with $\sigma_1 = 0.01$ rad and $\sigma_2 = 10^{-6}$ rad. This models the case that the boresight of the fine sensor is 16.26 degrees off the body x-axis.

The estimation error, which is defined as the rotation angle between the true and estimated attitudes, is the accuracy measure of most interest in applications. The estimation error is mathematically defined by

$$\begin{aligned} A_{\mathrm{opt}} &= \{\exp[(-\phi_{\mathrm{err}}\mathbf{e})\times]\}A_{\mathrm{true}} \\ &= \{\cos\phi_{\mathrm{err}} I + (1 - \cos\phi_{\mathrm{err}})\mathbf{e}\mathbf{e}^{\mathrm{T}} - \sin\phi_{\mathrm{err}}[\mathbf{e}\times]\}A_{\mathrm{true}}\,, \end{aligned} \tag{63}$$

where $\mathbf{e}$ is the unit vector defining the axis of the rotation that takes the true attitude to the optimal estimate. This gives, after some algebra,

$$\phi_{\mathrm{err}} = 2\sin^{-1}(\|A_{\mathrm{opt}}A_{\mathrm{true}}^{\mathrm{T}} - I\|/\sqrt{8}) = 2\sin^{-1}(\|A_{\mathrm{opt}} - A_{\mathrm{true}}\|/\sqrt{8})\,. \tag{64}$$

The estimation errors for the twelve test cases, computed with simulated measurement errors, are presented in Table 1. These errors are the same for FOAM, SOMA, and the SVD method, to the accuracy of the computation errors. The QUEST estimation errors are also identical, except for cases 10, 11, and 12 where they are much larger. The QUEST algorithm returns a flag warning of failure in

TABLE 1. Estimation Errors

Case	FOAM, SOMA, and SVD method ϕ_{err} (rad)	$L(A_{opt})$	ϕ_{cov} (rad)	QUEST ϕ_{err} (rad)	$L(A_{opt})$
1	1.23×10^{-6}	1.81	1.22×10^{-6}	1.23×10^{-6}	1.81
2	1.79×10^{-6}	1.15	1.58×10^{-6}	1.79×10^{-6}	1.15
3	1.25×10^{-2}	1.86	1.22×10^{-2}	1.25×10^{-2}	1.86
4	1.81×10^{-2}	1.18	1.58×10^{-2}	1.81×10^{-2}	1.18
5	1.21×10^{-2}	0.07	1.00×10^{-2}	1.21×10^{-2}	0
6	3.10×10^{-5}	2.19	8.66×10^{-5}	3.10×10^{-5}	1.87
7	3.94×10^{-5}	1.70	1.41×10^{-4}	3.94×10^{-5}	2.22
8	0.235	2.26	0.866	0.235	2.26
9	0.105	1.78	1.414	0.105	1.78
10	2.17×10^{-2}	2.13	2.53×10^{-2}	0.599	−9045
11	4.22×10^{-2}	0.13	3.57×10^{-2}	0.683	−9769
12	2.74×10^{-2}	2.34	3.57×10^{-2}	2.348	−9210

these cases, but it issues the same warning in cases 5 through 9 where it computes the attitude successfully. This poor behavior could possibly be corrected by tuning the control parameters in QUEST, but it is not clear how to accomplish this.

The optimized loss function $L(A_{opt})$, computed with unnormalized weights, is also given in the table. The QUEST column was computed by dividing the normalized loss function by σ_{tot}^2, which is the proportionality factor between the two loss functions given by equations (1), (42), (53), and (54). The unnormalized loss function for FOAM, SOMA, and the SVD method is on the order of the number of measurements, as expected. A larger value indicates either failure of the algorithm or the use of incorrect weights. The large negative values of $L(A_{opt})$ in cases 10, 11, and 12 provide a better warning of the failure of QUEST in these cases than the warning flag provided by the algorithm.

The FOAM, SOMA, and SVD covariance estimate of the error, ϕ_{cov}, which is identical for these three methods, is also given in Table 1. This is computed as the square root of the trace of the covariance matrix. The covariance estimate for QUEST differs, but not significantly, because it is computed differently. It can be seen that the covariance gives a good order-of-magnitude estimate of the expected estimation errors in most cases, and that it errs on the conservative side in cases where it differs significantly.

Since the estimation error is the same for FOAM, SOMA, and the SVD method, other error measures were investigated in order to differentiate among these methods. Table 2 presents the computation error for all methods, computed for the same cases but with $\mathbf{n}_i = \mathbf{0}$,

$$COMP = \|A_{opt} - A_{true}\| = \|A_{opt}A_{true}^T - I\|, \tag{65}$$

and the maximum orthogonality error, with or without simulated measurement errors,

$$ORTH = \|A_{opt}A_{opt}^T - I\|. \tag{66}$$

TABLE 2. Computation and Orthogonality Errors

Case	SVD method *COMP*	SVD method *ORTH*	FOAM and SOMA *COMP*	FOAM and SOMA *ORTH*	QUEST *COMP*	QUEST *ORTH*
1	3.27×10^{-17}	1.57×10^{-16}	4.61×10^{-16}	1.12×10^{-15}	3.10×10^{-16}	2.72×10^{-16}
2	3.27×10^{-17}	0.90×10^{-16}	3.05×10^{-16}	6.11×10^{-16}	1.69×10^{-16}	5.79×10^{-16}
3	4.72×10^{-17}	2.16×10^{-16}	5.27×10^{-16}	1.01×10^{-15}	1.69×10^{-16}	2.88×10^{-16}
4	2.43×10^{-17}	0.40×10^{-16}	3.05×10^{-16}	1.12×10^{-15}	1.69×10^{-16}	2.65×10^{-16}
5	1.63×10^{-10}	0.93×10^{-16}	7.83×10^{-9}	2.73×10^{-8}	1.09×10^{-9}	6.88×10^{-16}
6	6.62×10^{-17}	1.55×10^{-16}	4.66×10^{-12}	8.94×10^{-12}	4.16×10^{-9}	5.82×10^{-16}
7	3.74×10^{-15}	0.73×10^{-16}	7.84×10^{-12}	1.54×10^{-11}	1.33×10^{-12}	5.02×10^{-16}
8	2.48×10^{-15}	1.13×10^{-16}	4.04×10^{-12}	7.50×10^{-12}	4.16×10^{-9}	5.17×10^{-16}
9	9.91×10^{-16}	0.71×10^{-16}	5.70×10^{-12}	1.12×10^{-11}	1.13×10^{-12}	5.02×10^{-16}
10	3.67×10^{-17}	0.84×10^{-16}	1.49×10^{-7}	2.97×10^{-7}	1.91×10^{-8}	9.16×10^{-16}
11	6.28×10^{-17}	0.57×10^{-16}	1.45×10^{-7}	2.87×10^{-7}	2.55	8.31×10^{-16}
12	2.10×10^{-9}	1.40×10^{-16}	3.01×10^{-7}	6.00×10^{-7}	2.32×10^{-8}	3.09×10^{-16}

The FOAM and SOMA columns in the table give the maximum errors for all variants of these methods; no significant differences were seen between FOAM and SOMA or between normalized and unnormalized weights. The SVD method gives the smallest orthogonality and computation errors. QUEST gives computation errors comparable to the FOAM and SOMA errors, except for its failure in case 11, but has orthogonality errors similar to those of the SVD method, because it always produces a normalized quaternion. D_FLOATING arithmetic is about one decimal digit more precise than G_FLOATING arithmetic, as expected [20]. This is not significant in general, since the computation and orthogonality errors are much less than the estimation errors in all cases with realistic noise, except for QUEST in case 11. It is clear that cases with widely differing measurement accuracies furnish the greatest computational challenges to all the methods.

Algorithm Test–Speed

The above methods were also compared for computational speed. The measured CPU times were computed for sets of two to twelve observations, similar to case 6 in the previous section, and were effectively the same for normalized and unnormalized weights. They consist of a part that is independent of the number of observations processed and a part proportional to the number of observations:

$$t_{\text{QUEST}} = 0.24 + 0.09n \text{ msec}, \tag{67}$$

$$t_{\text{FOAM}} = 0.26 + 0.07n \text{ msec}, \tag{68}$$

$$t_{\text{SOMA}} = 0.36 + 0.07n \text{ msec}, \tag{69}$$

$$t_{\text{SVD}} = (3 \pm 1) + 0.07n \text{ msec}. \tag{70}$$

The n-dependent time in FOAM, SOMA, and SVD is the time required to normalize the input vectors and form the B matrix. The greater n-dependent time in QUEST is due to the computation of the information matrix, which QUEST uses to compute the covariance matrix. The n-independent time is the time required to perform all other computations, including the covariance matrix. The computation of λ generally requires one or two iterations in QUEST and two to six iterations in FOAM, due to the need to iterate to convergence in the latter method, which accounts for the greater n-independent time in FOAM. The transcendental function calls in SOMA account for its longer running time compared to FOAM, which is definitely preferable to SOMA since it is faster and no less accurate. The range of times for the SVD method is related to the rank and conditioning of the B matrix. This method is significantly slower than all the other methods tested, as has been noted previously; but the SVD method may still find applications in nearly singular estimation problems.

The absolute execution times are not of great significance, but the relative times are interesting. The exact CPU times will vary from case to case, and the time required for either FOAM or QUEST appears to be quite modest in comparison with other computations performed in spacecraft attitude determination.

It should be pointed out that FOAM computes the attitude matrix, while QUEST computes an attitude quaternion. If an attitude matrix is required from QUEST, as in the previous section, an additional step is required to compute it from the quaternion. This requires only multiplications and additions, though, and no transcendental function evaluations. If it is desired to compute a quaternion from FOAM, the standard method for extracting it from the attitude matrix can be used [21]. This requires the evaluation of one square root, but FOAM is faster than QUEST for more than three observations even with this addition.

Two-Observation Case

In the special case of two observations, the rank of B is at most two, so $\det B = 0$, which gives with equations (18), (19), and (23)

$$\kappa = \|\text{adj } B\|, \tag{71}$$

$$\lambda = (2\kappa + \|B\|^2)^{1/2}, \tag{72}$$

and

$$\zeta = \kappa\lambda\,. \tag{73}$$

Both κ and λ must be positive in order for λ to be the largest root of $p(\lambda)$. The explicit form for B as a function of the reference and observation vectors then yields

$$\text{adj } B^{\mathrm{T}} = a_1 a_2 (\mathbf{b}_1 \times \mathbf{b}_2)(\mathbf{r}_1 \times \mathbf{r}_2)^{\mathrm{T}}, \tag{74}$$

$$\kappa = a_1 a_2 |\mathbf{b}_1 \times \mathbf{b}_2|\,|\mathbf{r}_1 \times \mathbf{r}_2|, \tag{75}$$

and

$$\lambda = \{a_1^2 + 2a_1 a_2[|\mathbf{b}_1 \times \mathbf{b}_2|\,|\mathbf{r}_1 \times \mathbf{r}_2| + (\mathbf{b}_1^{\mathrm{T}}\mathbf{b}_2)(\mathbf{r}_1^{\mathrm{T}}\mathbf{r}_2)] + a_2^2\}^{1/2}. \tag{76}$$

The attitude is indeterminate if either the two reference vectors or the two observation vectors are parallel or antiparallel. Thus we will assume that both θ_r, the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$, and θ_b, the angle between $\mathbf{b}_1$ and $\mathbf{b}_2$, are strictly greater than zero and strictly less than pi. Now set $\lambda_0 = a_1 + a_2 = 1$ for the remainder of the discussion in this section, define

$$\varepsilon \equiv (\theta_b - \theta_r)/2\,, \tag{77}$$

and note that $|\varepsilon| < \pi/2$. The parameter ε is zero for perfect measurements since the angle between the reference vectors is equal to the angle between the observation vectors. The expression for λ can be written more compactly as

$$\lambda = (1 - 4a_1a_2\sin^2\varepsilon)^{1/2}, \tag{78}$$

which is equivalent to equation (72) in Shuster and Oh [6].

It is convenient to write the optimal attitude estimate in terms of the orthonormal triads:

$$\mathbf{r}_+ \equiv (\mathbf{r}_2 + \mathbf{r}_1)/[2\cos(\theta_r/2)], \tag{79a}$$

$$\mathbf{r}_- \equiv (\mathbf{r}_2 - \mathbf{r}_1)/[2\sin(\theta_r/2)], \tag{79b}$$

$$\mathbf{r}_+ \times \mathbf{r}_- = (\mathbf{r}_1 \times \mathbf{r}_2)/|\mathbf{r}_1 \times \mathbf{r}_2|\,, \tag{79c}$$

and

$$\mathbf{b}_+ \equiv (\mathbf{b}_2 + \mathbf{b}_1)/[2\cos(\theta_b/2)], \tag{80a}$$

$$\mathbf{b}_- \equiv (\mathbf{b}_2 - \mathbf{b}_1)/[2\sin(\theta_b/2)], \tag{80b}$$

$$\mathbf{b}_+ \times \mathbf{b}_- = (\mathbf{b}_1 \times \mathbf{b}_2)/|\mathbf{b}_1 \times \mathbf{b}_2|\,. \tag{80c}$$

Other orthogonal triads can be defined, but these preserve the maximum symmetry between the two measurements. The optimal attitude matrix expressed in terms of these triads is

$$\begin{aligned} A_{\text{opt}} = {} & [\cos\varepsilon(\mathbf{b}_+\mathbf{r}_+^{\mathrm{T}} + \mathbf{b}_-\mathbf{r}_-^{\mathrm{T}}) + (a_1 - a_2)\sin\varepsilon(\mathbf{b}_+\mathbf{r}_-^{\mathrm{T}} - \mathbf{b}_-\mathbf{r}_+^{\mathrm{T}})]/\lambda \\ & + (\mathbf{b}_+ \times \mathbf{b}_-)(\mathbf{r}_+ \times \mathbf{r}_-)^{\mathrm{T}}. \end{aligned} \tag{81}$$

It is interesting to note that a factor of a_1a_2 in the denominator of equation (14) has cancelled an identical factor in the numerator. Thus the attitude estimate has a well-defined limit as either a_1 or a_2 tends to zero, even though Wahba's loss function does not have a unique minimum in either limit. Another interesting property of the two-observation case is that the optimal estimate is independent of the weights when $\varepsilon = 0$. Equations (25) and (78) with $\lambda_0 = 1$ show that the optimized loss function is zero if any of a_1, a_2, or ε is zero.

We now investigate the conditions under which this optimal attitude estimate can be obtained by a generalization of the simpler TRIAD or algebraic method [5–7]. This is a well-known algorithm for computing an attitude matrix from two vector observations by forming orthonormal triads from the reference and observation vectors. One of the vectors in the reference triad is the normalized cross product of the two reference vectors, and the other two are orthonormal linear combinations of the two reference vectors. The most general form for the reference triad that we will consider is:

$$\mathbf{r}_{\mathrm{I}} \equiv \cos\psi_r \mathbf{r}_+ - \sin\psi_r \mathbf{r}_- = [\sin(\psi_r + \theta_r/2)\mathbf{r}_1 - \sin(\psi_r - \theta_r/2)\mathbf{r}_2]/\sin\theta_r, \tag{82a}$$

$$\mathbf{r}_{\mathrm{II}} \equiv \cos\psi_r \mathbf{r}_- + \sin\psi_r \mathbf{r}_+ = [\cos(\psi_r - \theta_r/2)\mathbf{r}_2 - \cos(\psi_r + \theta_r/2)\mathbf{r}_1]/\sin\theta_r, \tag{82b}$$

$$\mathbf{r}_{\mathrm{I}} \times \mathbf{r}_{\mathrm{II}} = \mathbf{r}_+ \times \mathbf{r}_-, \tag{82c}$$

where ψ_r is some rotation angle in the plane spanned by $\mathbf{r}_1$ and $\mathbf{r}_2$. The observation triad is

$$\mathbf{b}_{\mathrm{I}} \equiv \cos\psi_b \mathbf{b}_+ - \sin\psi_b \mathbf{b}_- = [\sin(\psi_b + \theta_b/2)\mathbf{b}_1 - \sin(\psi_b - \theta_b/2)\mathbf{b}_2]/\sin\theta_b, \tag{83a}$$

$$\mathbf{b}_{\mathrm{II}} \equiv \cos\psi_b \mathbf{b}_- + \sin\psi_b \mathbf{b}_+ = [\cos(\psi_b - \theta_b/2)\mathbf{b}_2 - \cos(\psi_b + \theta_b/2)\mathbf{b}_1]/\sin\theta_b, \tag{83b}$$

$$\mathbf{b}_{\mathrm{I}} \times \mathbf{b}_{\mathrm{II}} = \mathbf{b}_+ \times \mathbf{b}_-, \tag{83c}$$

similarly.

Different choices of the angles ψ_r and ψ_b give different variants of the TRIAD method. The choice $\psi_r = \psi_b = 0$, for example, treats the two measurements symmetrically, using all three components of each. The choice $\psi_r = \theta_r/2$ and $\psi_b = \theta_b/2$ gives

$$\mathbf{r}_{\mathrm{I}} = \mathbf{r}_1, \tag{84a}$$

$$\mathbf{r}_{\mathrm{II}} = (\mathbf{r}_2 - \cos\theta_r \mathbf{r}_1)/\sin\theta_r, \tag{84b}$$

and similar relations for $\mathbf{b}_{\mathrm{I}}$ and $\mathbf{b}_{\mathrm{II}}$. This makes no use of the component of $\mathbf{r}_2$ along $\mathbf{r}_1$ or the component of $\mathbf{b}_2$ along $\mathbf{b}_1$. The choice $\psi_r = -\theta_r/2$ and $\psi_b = -\theta_b/2$, on the other hand, gives

$$\mathbf{r}_{\mathrm{I}} = \mathbf{r}_2, \tag{85a}$$

$$\mathbf{r}_{\mathrm{II}} = -(\mathbf{r}_1 - \cos\theta_r \mathbf{r}_2)/\sin\theta_r, \tag{85b}$$

and similarly for $\mathbf{b}_{\mathrm{I}}$ and $\mathbf{b}_{\mathrm{II}}$. This choice ignores the components of $\mathbf{r}_1$ and $\mathbf{b}_1$ along $\mathbf{r}_2$ and $\mathbf{b}_2$, respectively. The key point is that ψ_r is some function of θ_r and the measurement weights, and ψ_b is the *same* function of θ_b and the weights. Note that this does not imply that $\psi_r = \psi_b$ except in the case that $\varepsilon = 0$. Often, the TRIAD method is understood to mean only the special cases of equations (84) or (85), rather than the generalized method specified by equations (82) and (83). These special cases are the most common cases, since TRIAD is often employed where one vector is much better known than the other, so one component of the less-well-known vector is not used.

The TRIAD attitude estimate is given by

$$\begin{aligned} A_{\mathrm{TRIAD}} &= [\mathbf{b}_{\mathrm{I}} \vdots \mathbf{b}_{\mathrm{II}} \vdots \mathbf{b}_{\mathrm{I}} \times \mathbf{b}_{\mathrm{II}}][\mathbf{r}_{\mathrm{I}} \vdots \mathbf{r}_{\mathrm{II}} \vdots \mathbf{r}_{\mathrm{I}} \times \mathbf{r}_{\mathrm{II}}]^{\mathrm{T}} \\ &= \mathbf{b}_{\mathrm{I}}\mathbf{r}_{\mathrm{I}}^{\mathrm{T}} + \mathbf{b}_{\mathrm{II}}\mathbf{r}_{\mathrm{II}}^{\mathrm{T}} + (\mathbf{b}_{\mathrm{I}} \times \mathbf{b}_{\mathrm{II}})(\mathbf{r}_{\mathrm{I}} \times \mathbf{r}_{\mathrm{II}})^{\mathrm{T}} \\ &= \cos(\psi_b - \psi_r)(\mathbf{b}_+\mathbf{r}_+^{\mathrm{T}} + \mathbf{b}_-\mathbf{r}_-^{\mathrm{T}}) + \sin(\psi_b - \psi_r)(\mathbf{b}_+\mathbf{r}_-^{\mathrm{T}} - \mathbf{b}_-\mathbf{r}_+^{\mathrm{T}}) \\ &\quad + (\mathbf{b}_+ \times \mathbf{b}_-)(\mathbf{r}_+ \times \mathbf{r}_-)^{\mathrm{T}}. \end{aligned} \tag{86}$$

We now attempt to find angles ψ_r and ψ_b such that the TRIAD solution gives the optimal attitude estimate of equation (81). We immediately find such angles in four special cases:

1) If $\varepsilon = 0$, then $\psi_r = \psi_b$ automatically, all TRIAD solutions are the same, and they all agree with the optimal estimate, which is independent of the weights in the loss function.
2) If $a_1 = a_2 = 1/2$, the TRIAD solution with $\psi_r = \psi_b = 0$ and with vector triads given by equations (79) and (80) gives the optimal estimate.
3) If $a_1 = 1$, $a_2 = 0$, the TRIAD solution with $\psi_r = \theta_r/2$, $\psi_b = \theta_b/2$ and with triads as in equations (84) gives the optimal estimate.
4) If $a_1 = 0, a_2 = 1$, the TRIAD solution with $\psi_r = -\theta_r/2$, $\psi_b = -\theta_b/2$ and with triads as in equations (85) gives the optimal estimate.

We will now show that the TRIAD solution does not minimize Wahba's loss function except in these four special cases. Comparing equations (81) and (86) gives the following necessary condition for agreement of the TRIAD and optimal attitude estimates:

$$\tan(\psi_b - \psi_r) = (a_1 - a_2)\tan\varepsilon\,. \tag{87}$$

Set $\theta_r = \theta_0$, some arbitrarily chosen angle, and denote the corresponding value of ψ_r by ψ_0, which is also a function of the observation weights. Then

$$\tan(\psi_b - \psi_0) = (a_1 - a_2)\tan[(\theta_b - \theta_0)/2] \equiv (a_1 - a_2)\tau_b\,. \tag{88}$$

This equation must hold for any θ_r, with ψ_0 and θ_0 regarded as fixed parameters, since ψ_b is required to be a function of θ_b and the weights only, and not of θ_r. Now setting $\theta_b = \theta_0$ in equation (87) gives $\psi_b = \psi_0$ and

$$\tan(\psi_r - \psi_0) = (a_1 - a_2)\tan[(\theta_r - \theta_0)/2] \equiv (a_1 - a_2)\tau_r\,, \tag{89}$$

which must hold for any θ_b. In fact, equation (89) could have been written directly in analogy with equation (88), since ψ_r is required to be the same function of θ_r and the measurement weights as ψ_b is of θ_b and the weights. Now combining equations (88) and (89) with some elementary trigonometry gives

$$\begin{aligned}\tan(\psi_b - \psi_r) &= \tan[(\psi_b - \psi_0) - (\psi_r - \psi_0)]\\ &= (a_1 - a_2)(\tau_b - \tau_r)/[1 + (a_1 - a_2)^2\tau_b\tau_r]\\ &= (a_1 - a_2)\tan\varepsilon(1 + \tau_b\tau_r)/[1 + (1 - 4a_1a_2)\tau_b\tau_r]\,.\end{aligned} \tag{90}$$

Equating the right sides of equations (87) and (90) gives, after some cancellations, the necessary condition

$$4a_1a_2\tau_b\tau_r(a_1 - a_2)\tan\varepsilon = 0\,, \tag{91}$$

which is satisfied in the four special cases discussed above. It is also satisfied if either τ_b or τ_r is zero, but these conditions cannot be satisfied in general since θ_0 is an arbitrarily chosen angle. Thus the TRIAD method cannot find the optimal attitude minimizing Wahba's loss function in the general case, but only in the special cases $\varepsilon = 0$, $a_1 = 0$, $a_2 = 0$, and $a_1 = a_2$.

Two-Observation Case–Covariance

The covariance matrix in the two-observation case can be written, using equations (4), (42), (52), (73), and (75), as

$$P = \lambda_0\sigma_{tot}^2[|\mathbf{b}_1 \times \mathbf{b}_2||\mathbf{r}_1 \times \mathbf{r}_2|I + \sigma_2^2\sigma_1^{-2}\mathbf{b}_1\mathbf{b}_1^T + (\mathbf{r}_1^T\mathbf{r}_2)(\mathbf{b}_1\mathbf{b}_2^T + \mathbf{b}_2\mathbf{b}_1^T) + \sigma_1^2\sigma_2^{-2}\mathbf{b}_2\mathbf{b}_2^T]/(\lambda|\mathbf{b}_1 \times \mathbf{b}_2||\mathbf{r}_1 \times \mathbf{r}_2|) . \qquad (92)$$

Now we make the approximation that $\theta_r = \theta_b$, which should be valid for computing the covariance matrix if the measurement errors are much less than the angular separation of the two vector observations. This gives $\lambda = \lambda_0$ and allows the covariance matrix to be expressed as a function of the observation vectors only:

$$P = \sigma_{tot}^2|\mathbf{b}_1 \times \mathbf{b}_2|^{-2}[|\mathbf{b}_1 \times \mathbf{b}_2|^2 I + \sigma_2^2\sigma_1^{-2}\mathbf{b}_1\mathbf{b}_1^T + (\mathbf{b}_1^T\mathbf{b}_2)(\mathbf{b}_1\mathbf{b}_2^T + \mathbf{b}_2\mathbf{b}_1^T) + \sigma_1^2\sigma_2^{-2}\mathbf{b}_2\mathbf{b}_2^T] . \qquad (93)$$

This expression, which is mathematically equivalent to equation (100) in Shuster and Oh [6], can be further simplified by writing the identity matrix as a dyadic in the orthonormal triad $\mathbf{b}_I$, $\mathbf{b}_{II}$, and $\mathbf{b}_I \times \mathbf{b}_{II}$:

$$I = \mathbf{b}_I\mathbf{b}_I^T + \mathbf{b}_{II}\mathbf{b}_{II}^T + (\mathbf{b}_I \times \mathbf{b}_{II})(\mathbf{b}_I \times \mathbf{b}_{II})^T$$
$$= |\mathbf{b}_1 \times \mathbf{b}_2|^{-2}[\mathbf{b}_1\mathbf{b}_1^T + \mathbf{b}_2\mathbf{b}_2^T + (\mathbf{b}_1 \times \mathbf{b}_2)(\mathbf{b}_1 \times \mathbf{b}_2)^T - (\mathbf{b}_1^T\mathbf{b}_2)(\mathbf{b}_1\mathbf{b}_2^T + \mathbf{b}_2\mathbf{b}_1^T)] . \qquad (94)$$

Substituting this into equation (93) and using equation (41) gives

$$P = |\mathbf{b}_1 \times \mathbf{b}_2|^{-2}[\sigma_2^2\mathbf{b}_1\mathbf{b}_1^T + \sigma_1^2\mathbf{b}_2\mathbf{b}_2^T + \sigma_{tot}^2(\mathbf{b}_1 \times \mathbf{b}_2)(\mathbf{b}_1 \times \mathbf{b}_2)^T] . \qquad (95)$$

This expression is very useful for predicting attitude estimation performance for a given pair of reference vectors. In particular, it explicitly displays the singularity of the covariance matrix for coaligned vector measurements. For nearly coaligned measurements, the eigenvectors of the covariance matrix can be shown with some effort to be $a_1\mathbf{b}_1 + a_2\mathbf{b}_2$, with variance $|\mathbf{b}_1 \times \mathbf{b}_2|^{-2}(\sigma_1^2 + \sigma_2^2) - \sigma_{tot}^2$, and any two axes perpendicular to this line, with variance σ_{tot}^2, up to terms of order $|\mathbf{b}_1 \times \mathbf{b}_2|^2$.

Since the TRIAD algorithm with appropriately chosen vector triads gives the optimal estimate in the special cases $\sigma_1 = \sigma_2$, $\sigma_1 \ll \sigma_2$, and $\sigma_1 \gg \sigma_2$, as discussed above, equation (95) also gives the TRIAD covariance in these cases.

Conclusions

A new algorithm for minimizing Wahba's loss function has been found, which solves for the optimal attitude matrix directly, without the intermediate computation of a quaternion or other parameterization of the attitude. Since the attitude matrix is inherently nonsingular, there are no problems with special cases like 180 degree rotations, and no special procedures are needed to deal with such cases. Two variants of the new algorithm are given; one is completely analytic, while the other requires an iterative calculation of a scalar coefficient in the solution. The method with the iterative calculation is as fast as any existing method even if the computation of a quaternion from the attitude matrix is included. The completely analytic variant is slower and no more accurate, so its use is not recommended. The new algorithm appears to be more robust than the existing fast optimal quaternion estimation method in some difficult cases, and it requires fewer control parameters. The covariance of the attitude error angles can be computed very efficiently, since it makes use of the same scalar and matrix quantities needed for the optimal attitude computation.

Intuitively appealing closed-form solutions for the optimal attitude matrix and the covariance matrix are presented for the special case of two observations. The optimal attitude estimate is compared with the well known non-optimal estimate computed using orthonormal triads formed from the observation and reference vectors. When the angle between the two reference vectors is equal to the angle between the two observation vectors, all triad choices give the optimal estimate, which is independent of the weights in the loss function. Except for this case, the optimal and triad-based attitude estimates agree only when the two vector measurements are given equal weights in the loss function or when the weight given to one vector measurement is negligible compared to the weight given to the other.

Acknowledgments

This paper has benefitted greatly from comments by Malcolm D. Shuster, Gregory A. Natanson, and two anonymous reviewers.

References

[1] WAHBA, G. "A Least Squares Estimate of Spacecraft Attitude," *SIAM Review,* Vol. 7, No. 3, July 1965, p. 409.

[2] FARRELL, J. L., STUELPNAGEL, J. C., WESSNER, R. H., VELMAN, J. R., and BROCK, J. E. "A Least Squares Estimate of Spacecraft Attitude," *SIAM Review,* Vol. 8, No. 3, July 1966, pp. 384–386.

[3] DAVENPORT, P. B. "A Vector Approach to the Algebra of Rotations with Applications," NASA X-546-65-437, November 1965.

[4] KEAT, J. "Analysis of Least Squares Attitude Determination Routine DOAOP," CSC/TM-77/6034, Computer Sciences Corporation, Lanham-Seabrook, Maryland, February 1977.

[5] LERNER, G. M. "Three-Axis Attitude Determination," *Spacecraft Attitude Determination and Control,* J. R. Wertz (editor), D. Reidel, Dordrecht, Holland, 1978.

[6] 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.

[7] BLACK, H. D. "A Passive System for Determining the Attitude of a Satellite," *AIAA Journal,* Vol. 2, No. 7, July 1964, pp. 1350–1351.

[8] GOLUB, G. H., and VAN LOAN, C. F. *Matrix Computations,* Johns Hopkins University Press, Baltimore, Maryland, 1983.

[9] HORN, R. A., and JOHNSON, C. R. *Matrix Analysis,* Cambridge University Press, Cambridge, United Kingdom, 1985.

[10] DAVENPORT, P. B. "Attitude Determination and Sensor Alignment via Weighted Least Squares Affine Transformations," Paper No. 71-396, AAS/AIAA Astrodynamics Specialists Conference, Fort Lauderdale, Florida, August 1971.

[11] MARKLEY, F. L. "Attitude Determination Using Vector Observations and the Singular Value Decomposition," *Journal of the Astronautical Sciences,* Vol. 36, No. 3, July–September 1988, pp. 245–258.

[12] DANBY, J. M. A. *Fundamentals of Celestial Mechanics,* 2nd edition, Willman-Bell, Richmond, Virginia, 1988.

[13] BEYER, W. H. *CRC Standard Mathematical Tables,* 2nd Edition, CRC Press, Boca Raton, Florida, 1987.

[14] SHUSTER, M. D. "Maximum Likelihood Estimation of Spacecraft Attitude," *Journal of the Astronautical Sciences,* Vol. 37, No. 1, January-March 1989, pp. 79–88.

[15] SORENSON, H.W. *Parameter Estimation: Principles and Problems,* Marcel Dekker, New York, 1980.

[16] MARKLEY, F. L. "Attitude Determination and Parameter Estimation Using Vector Observations: Theory," *Journal of the Astronautical Sciences,* Vol. 37, No. 1, January–March 1989, pp. 41–58.

[17] MARKLEY, F. L. "Attitude Determination and Parameter Estimation Using Vector Observations: Application," *Journal of the Astronautical Sciences,* Vol. 39, No. 3, July–September 1991, pp. 367–381.

[18] *IEEE Standard for Binary Floating-Point Arithmetic,* ANSI/IEEE Standard 754–1985, The Institute of Electrical and Electronics Engineers, New York, 1985.

[19] *Apple Numerics Manual,* 2nd Edition, Addison-Wesley, Reading Massachusetts, 1988.

[20] *VAX-11 FORTRAN Language Reference Manual,* Digital Equipment Corporation, Maynard, Massachusetts, 1982.

[21] SHEPPERD, S.W. "Quaternion from Rotation Matrix," *Journal of Guidance and Control,* Vol. 1, No. 3, May–June 1978, pp. 223–224.