Angles Only Three-Axis Attitude Determination

Sergei Tanygin[1]
Analytical Graphics, Inc., Exton, PA, 19341

The general problem of determining the attitude from measurements of arc lengths or equivalently from measurements of dot products of vectors is examined. The measurement model explicitly quadratic in quaternions is used to develop a maximum likelihood quaternion estimator. The optimal estimate is obtained using the Gauss-Newton estimation sequence inspired by the homotopy continuation methods. The algorithm is tested numerically and is found to be more robust than other iterative local linearization techniques.

Nomenclature

A	=	attitude 3×3 direction cosine matrix
$\hat{\mathbf{q}}$	=	4×1 unit quaternion equivalent to A
$\mathbf{q}$	=	3×1 partition of $\hat{\mathbf{q}}$ representing its vector part
q_4	=	4-th element $\hat{\mathbf{q}}$ representing its scalar part
$\mathbf{p}$	=	3×1 vector of modified Rodrigues parameters equivalent to $\hat{\mathbf{q}}$
$\Delta\xi$	=	3×1 attitude error vector
$\mathbf{r}$	=	3×1 reference vector in reference frame
$\mathbf{s}$	=	3×1 sensing axis or antenna baseline in body-fixed frame
d	=	scalar effective angular measurement
w	=	scalar effective measurement noise
$\mathbf{d}$	=	column vector of correlated effective angular measurements
$\mathbf{w}$	=	column vector of correlated effective measurement noise
R	=	measurement error covariance matrix
N	=	number of measurements
I	=	identity matrix

I. Introduction

THE overwhelming majority of modern spacecraft determine three-axis attitude using complete vector data. They typically employ one of two types of algorithms: deterministic algorithms which solve for attitude using a minimal set of data and optimal algorithms which find optimal attitude by using additional measurements to minimize an appropriate cost function. These algorithms first introduced several decades ago have been flight tested numerous times and continue to be subjects of active research (see Refs. 1-5). Nowadays, optimal algorithms are employed more frequently than deterministic ones because most modern spacecraft come equipped with sensors able to provide surplus measurements and with computers able to process them at a negligible additional computational cost. Nevertheless, it may still be necessary to perform three-axis attitude determination without complete vector or attitude data in one of the following scenarios: for a low cost spacecraft equipped with simple alternatives to more capable sensors, for a back-up attitude determination system or for an *ad hoc* attitude determination system that needs to operate in a contingency with partially failed sensors. For example, a spacecraft equipped with a three-axis magnetometer and a single axis Sun sensor may be limited to three angle measurements if one axis of the magnetometer becomes defective.

[1] Senior Astrodynamics Specialist, 220 Valley Creek Blvd., Exton PA 19341, Senior Member AIAA.

Another relatively recent approach to attitude determination involves using phase difference measurements from Global Positioning System (GPS) receivers. The measurements in this approach are effectively angles between antenna baselines and directions to GPS satellites.

The angles only deterministic algorithm has been carefully examined in Ref. 6 and various GPS attitude determination algorithms have been introduced by several authors in Refs. 7-9. In this paper, we employ the measurement model explicitly quadratic in quaternions to design a maximum likelihood quaternion estimator. We find the optimal estimate using the Gauss-Newton estimation sequence inspired by the homotopy continuation methods. The algorithm is tested numerically and is found to be more robust than other iterative local linearization techniques.

II. Measurement Model

The effective measurements for many sensors types are derived from various angle and dihedral angle arcs. These are measured between different reference vectors and sensing axes known respectively in the reference and body-fixed frames. It has been shown in Refs. 6 and 8 that various measurements can be formulated using the same mathematical model: a dot product of two vectors related via the attitude matrix. Let A denote the attitude matrix which is a 3×3 direction cosine matrix mapping vector components from the reference frame to the body-fixed frame and let $\mathbf{s}$ and $\mathbf{r}$ be 3×1 column-vectors. The former defines a sensing axis or an antenna baseline in the body-fixed frame. The latter typically represents one of the following in the reference frame: the Sun direction for single-axis Sun sensors, the Earth direction for Earth horizon scanners, the magnetic field vector for single-axis magnetometers, a cross product of the Sun and the Earth directions for Earth-Sun dihedral measurements (see Refs. 6 and 10 for further details) or a direction to a GPS satellite (see Refs. 7-9 for further details). In all cases we can adopt the following measurement model for uncorrelated measurements:[2]

$$d = \mathbf{s}^{\mathrm{T}}\mathsf{A}\mathbf{r} + w \quad \text{with} \quad w \sim N(0,\sigma^2)\,. \tag{1}$$

This model is quadratic in $\hat{\mathbf{q}}$, a quaternion equivalent of A, and can be transformed into[3]

$$d = \hat{\mathbf{q}}^{\mathrm{T}}\mathsf{K}(\mathbf{r},\mathbf{s})\hat{\mathbf{q}} + w \quad \text{or} \quad d' = d - \mathbf{r}^{\mathrm{T}}\mathbf{s} = \hat{\mathbf{q}}^{\mathrm{T}}\mathsf{L}(\mathbf{r},\mathbf{s})\hat{\mathbf{q}} + w \tag{2ab}$$

where K and L are symmetric 4×4 matrices constructed as follows

$$\mathsf{K}(\mathbf{r},\mathbf{s}) = \begin{bmatrix} \mathbf{s}\mathbf{r}^{\mathrm{T}} + \mathbf{r}\mathbf{s}^{\mathrm{T}} - (\mathbf{r}^{\mathrm{T}}\mathbf{s})\mathsf{I} & -(\mathbf{r}\times\mathbf{s}) \\ -(\mathbf{r}\times\mathbf{s})^{\mathrm{T}} & \mathbf{r}^{\mathrm{T}}\mathbf{s} \end{bmatrix}, \tag{3}$$

$$\mathsf{L}(\mathbf{r},\mathbf{s}) = \begin{bmatrix} [\mathbf{r}\times][\mathbf{s}\times] + [\mathbf{s}\times][\mathbf{r}\times] & -(\mathbf{r}\times\mathbf{s}) \\ -(\mathbf{r}\times\mathbf{s})^{\mathrm{T}} & 0 \end{bmatrix}. \tag{4}$$

Here and in subsequent derivations we define a cross product matrix for any 3×1 column-vector $\mathbf{a}$ as

$$[\mathbf{a}\times] = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}. \tag{5}$$

[2] We discuss correlated measurements later in the paper.

[3] In subsequent derivations we adopt the K-based quadratic model with the understanding that it is trivial to substitute the L-based quadratic model in its place.

III. Maximum Likelihood Quaternion Cost Function

A. Uncorrelated Measurements

We start our development of the maximum likelihood quaternion estimator by selecting a suitable cost function. For a moment we assume that the estimator deals with $N \geq 3$ scalar uncorrelated measurements each described by the measurement model of equation (1). We propose the cost function

$$\phi(\hat{\mathbf{q}}) = \frac{1}{4}\sum_{n=1}^{N} a_n \left(\hat{\mathbf{q}}^{\mathrm{T}}\mathsf{K}_n\hat{\mathbf{q}} - d_n\right)^2 \tag{6}$$

where

$$a_n = \frac{\sigma^2}{\sigma_n^2} \text{ and } \frac{1}{\sigma^2} = \sum_{n=1}^{N}\frac{1}{\sigma_n^2}. \tag{7}$$

If we assume that the measurements are independent, zero-mean and Gaussian then this cost function is proportional to the data-dependent part of the negative-log-likelihood function as shown below

$$J(\hat{\mathbf{q}}) = 2\frac{\phi(\hat{\mathbf{q}})}{\sigma^2}. \tag{8}$$

Therefore, the maximum-likelihood estimate $\hat{\mathbf{q}}^*$ that minimizes $J(\hat{\mathbf{q}})$ also minimizes $\phi(\hat{\mathbf{q}})$ and vice versa:

$$\hat{\mathbf{q}}^* = \arg\min_{|\hat{\mathbf{q}}|=1} J(\hat{\mathbf{q}}) = \arg\min_{|\hat{\mathbf{q}}|=1} \phi(\hat{\mathbf{q}}). \tag{9}$$

Following developments in Refs. 6 and 8 it is easy to show that the maximum-likelihood estimate $\hat{\mathbf{q}}^*$ has the associated attitude error covariance $\mathsf{P}_{\xi\xi}^*$ equal to the inverse of the Fisher information matrix

$$\mathsf{P}_{\xi\xi}^{*\,-1} = \sum_{n=1}^{N}\frac{1}{\sigma_n^2}\left[\mathbf{s}_n \times \left(\mathsf{A}_{(true)}\mathbf{r}_n\right)\right]\left[\mathbf{s}_n \times \left(\mathsf{A}_{(true)}\mathbf{r}_n\right)\right]^{\mathrm{T}}. \tag{10}$$

B. Correlated Measurements

When measurements are correlated, the cost function is somewhat more complicated because correlated measurements are now assembled in column-vectors and their correlations are captured in measurement error covariance matrices. We write

$$\phi(\hat{\mathbf{q}}) = \frac{1}{4}\sum_{n=1}^{N}\left(\boldsymbol{\mu}_n(\hat{\mathbf{q}}) - \mathbf{d}_n\right)^{\mathrm{T}}\bar{\mathsf{R}}_n^{-1}\left(\boldsymbol{\mu}_n(\hat{\mathbf{q}}) - \mathbf{d}_n\right) = \frac{1}{4}\sum_{n=1}^{N}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n}\bar{r}_{ij}^{-1}\left(\hat{\mathbf{q}}^{\mathrm{T}}\mathsf{K}_{ni}\hat{\mathbf{q}} - d_{ni}\right)\left(\hat{\mathbf{q}}^{\mathrm{T}}\mathsf{K}_{nj}\hat{\mathbf{q}} - d_{nj}\right) \tag{11}$$

where

$$\boldsymbol{\mu}_n(\hat{\mathbf{q}}) = \begin{bmatrix} \hat{\mathbf{q}}^{\mathrm{T}}\mathsf{K}_{n1}\hat{\mathbf{q}} \\ \cdots \\ \hat{\mathbf{q}}^{\mathrm{T}}\mathsf{K}_{nm_n}\hat{\mathbf{q}} \end{bmatrix},\ \mathbf{d}_n = \begin{bmatrix} d_{n1} \\ \cdots \\ d_{nm_n} \end{bmatrix},\ \bar{\mathsf{R}}_n^{-1} = \sigma^2\mathsf{R}_n^{-1} \text{ and } \frac{1}{\sigma^2} = tr\left\{\sum_{n=1}^{N}\mathsf{R}_n^{-1}\right\}. \tag{12}$$

Here $\boldsymbol{\mu}_n$, $\mathbf{d}_n$ are the $m_n \times 1$ vectors representing a single group of correlated measurements with the $m_n \times m_n$ measurement error covariance matrix R_n and $\bar{r}_{ij}^{-1}$ is the ijth element of $\bar{\mathsf{R}}_n^{-1}$. In this context N denotes the number of these groups of correlated measurements. If all measurements are uncorrelated, then $m_n = 1$ for all $n = 1, ..., N$, $\boldsymbol{\mu}_n$, $\mathbf{d}_n$ and R_n reduce to scalars and ϕ reduces to the version of the cost function for the uncorrelated scalar measurements shown in equation (6).

The attitude error covariance $\mathsf{P}_{\xi\xi}^*$ for the correlated estimator is developed the same way as the one for the uncorrelated estimator. It is equal to the inverse of the Fisher information matrix given by

$$\mathsf{P}_{\xi\xi}^{*\,-1} = \sum_{n=1}^{N} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} r_{ij}^{-1} \left[\mathbf{s}_{ni} \times \left(\mathsf{A}_{(true)} \mathbf{r}_{ni} \right) \right] \left[\mathbf{s}_{nj} \times \left(\mathsf{A}_{(true)} \mathbf{r}_{nj} \right) \right]^{\mathrm{T}}, \tag{13}$$

where r_{ij}^{-1} is the ijth element of R_n^{-1}.

IV. Cost Function Minimization via Homotopy Continuation

We propose to minimize the cost function ϕ using the Gauss-Newton estimation sequence inspired by the homotopy continuation methods. The sequence is designed using the modified Rodrigues parameters $\mathbf{p}$ in order to avoid dealing with the quaternion norm constraint. Relationships between $\mathbf{p}$ and $\hat{\mathbf{q}}$ are well known (see Ref. 11) and are computationally inexpensive:[4]

$$\mathbf{p} = \frac{\mathbf{q}}{1 + q_4} \text{ and } \hat{\mathbf{q}} = \frac{1}{1 + p^2} \begin{bmatrix} 2\mathbf{p} \\ 1 - p^2 \end{bmatrix}, \tag{14ab}$$

where $p^2 = \mathbf{p}^{\mathrm{T}} \mathbf{p}$, $\mathbf{q}$ is the vector part of the quaternion and q_4 is the scalar part. The homotopy continuation methods solve problems without closed-form solutions by evolving to them from problems with known solutions. Let $\hat{\mathbf{q}}_0$ and $\mathbf{p}_0$ related by equations (14) represent the initial attitude guess. It is trivial to create a new cost function ϕ_0 of the form (6) or (11) with new "measurements" selected so that all measurement residuals vanish at this initial guess, and consequently $\phi_0(\hat{\mathbf{q}}_0) = 0$. Then we can look for an update to $\hat{\mathbf{q}}_0$ and $\mathbf{p}_0$ such that at the updated attitude the actual cost function ϕ is minimized instead of ϕ_0. We consider the necessary conditions for minimization of ϕ_0 and ϕ

$$\phi_{0\mathbf{p}}'^{\mathrm{T}}(\mathbf{p}_0) = \mathbf{0} \text{ and } \phi_{\mathbf{p}}'^{\mathrm{T}}(\mathbf{p}^*) = \mathbf{0} \tag{15ab}$$

and write the following homotopy to first order

$$\phi_{0\mathbf{pp}}''(\mathbf{p}_0)(\mathbf{p} - \mathbf{p}_0) + \phi_{\mathbf{p}}'^{\mathrm{T}}(\mathbf{p}_0) - \phi_{0\mathbf{p}}'^{\mathrm{T}}(\mathbf{p}_0) = \mathbf{0}. \tag{16}$$

Since the last term in the homotopy vanishes by construction as shown in equation (15a), we solve for the updated $\mathbf{p}$ using

[4] Here we present only the positive form of the modified Rodrigues parameters. The negative form is, of course, equally viable.

$$\mathbf{p} = \mathbf{p}_0 - \left[\phi''_{0\mathbf{pp}}(\mathbf{p}_0)\right]^{-1} \phi'^{\mathrm{T}}_{\mathbf{p}}(\mathbf{p}_0). \tag{17}$$

We can differentiate ϕ_0 and ϕ with respect to $\mathbf{p}$ via the chain rule

$$\phi'_{\mathbf{p}} = \phi'_{\hat{\mathbf{q}}}\hat{\mathbf{q}}'_{\mathbf{p}}, \tag{18}$$

where from equation (14b)

$$\hat{\mathbf{q}}'_{\mathbf{p}} = \begin{bmatrix} (1+q_4)\mathbf{I} \\ -\mathbf{q}^{\mathrm{T}} \end{bmatrix} - \hat{\mathbf{q}}\mathbf{q}^{\mathrm{T}}. \tag{19}$$

We differentiate equations (6) and (11) with respect to $\hat{\mathbf{q}}$ and evaluate them at $\hat{\mathbf{q}}_0$ to obtain

$$\phi'_{\hat{\mathbf{q}}}(\hat{\mathbf{q}}_0) = \sum_{n=1}^{N} a_n \left(\hat{\mathbf{q}}_0^{\mathrm{T}}\mathsf{K}_n\hat{\mathbf{q}}_0 - d_n\right)\hat{\mathbf{q}}_0^{\mathrm{T}}\mathsf{K}_n \tag{20}$$

and

$$\phi'_{\hat{\mathbf{q}}}(\hat{\mathbf{q}}_0) = \frac{1}{2}\sum_{n=1}^{N}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n} \overline{r}_{ij}^{-1}\left[\left(\hat{\mathbf{q}}_0^{\mathrm{T}}\mathsf{K}_{ni}\hat{\mathbf{q}}_0 - d_{ni}\right)\hat{\mathbf{q}}_0^{\mathrm{T}}\mathsf{K}_{nj} + \left(\hat{\mathbf{q}}_0^{\mathrm{T}}\mathsf{K}_{nj}\hat{\mathbf{q}}_0 - d_{nj}\right)\hat{\mathbf{q}}_0^{\mathrm{T}}\mathsf{K}_{ni}\right], \tag{21}$$

respectively. In our derivations of the Hessian $\phi''_{0\mathbf{pp}}(\mathbf{p}_0)$, we benefit from the fact that at $\mathbf{p}_0$ (and equivalently $\hat{\mathbf{q}}_0$) all measurement residuals in ϕ_0 vanish by construction. This reduces the chain rule computation for $\phi''_{0\mathbf{pp}}(\mathbf{p}_0)$ to

$$\phi''_{0\mathbf{pp}}(\mathbf{p}_0) = \hat{\mathbf{q}}'^{\mathrm{T}}_{\mathbf{p}}\phi''_{0\hat{\mathbf{q}}\hat{\mathbf{q}}}\hat{\mathbf{q}}'_{\mathbf{p}} \tag{22}$$

and reduces the Hessians $\phi''_{0\hat{\mathbf{q}}\hat{\mathbf{q}}}(\hat{\mathbf{q}}_0)$ derived from equations (6) and (11) to

$$\phi''_{0\hat{\mathbf{q}}\hat{\mathbf{q}}}(\hat{\mathbf{q}}_0) = 2\sum_{n=1}^{N} a_n\mathsf{K}_n\hat{\mathbf{q}}_0\hat{\mathbf{q}}_0^{\mathrm{T}}\mathsf{K}_n. \tag{23}$$

and

$$\phi''_{0\hat{\mathbf{q}}\hat{\mathbf{q}}}(\hat{\mathbf{q}}_0) = \sum_{n=1}^{N}\sum_{i=1}^{m_n}\sum_{j=1}^{m_n} \overline{r}_{ij}^{-1}\left(\mathsf{K}_{ni}\hat{\mathbf{q}}_0\hat{\mathbf{q}}_0^{\mathrm{T}}\mathsf{K}_{nj} + \mathsf{K}_{nj}\hat{\mathbf{q}}_0\hat{\mathbf{q}}_0^{\mathrm{T}}\mathsf{K}_{ni}\right), \tag{24}$$

respectively. Now using equation (17) and aided by equations (18-24), we can establish the Gauss-Newton estimation sequence for $\mathbf{p}$ which should converge to the optimal estimate $\mathbf{p}^*$. In other words, the Gauss-Newton iteration interpreted as an *approximation* of the Newton iteration can also be viewed as *exact* to first order when based on the homotopy continuation. Of course, at any iteration the estimates in $\mathbf{p}$ and in $\hat{\mathbf{q}}$ can be related via equations (14).

It is instructive to verify that the covariance results of equations (10) and (13) are recovered in our algorithm. We can derive the inverse of the error covariance in $\mathbf{p}$ as the Fisher information matrix (see equation (8))

$$\mathsf{P}_{\mathbf{pp}}^{*\,-1} = J''_{\mathbf{pp}(true)} = \frac{2}{\sigma^2}\phi''_{\mathbf{pp}(true)}. \tag{25}$$

Then the attitude error covariance given by

$$\mathsf{P}_{\xi\xi}^{*} = \xi'_{\mathbf{p}(true)}\mathsf{P}_{\mathbf{pp}}^{*}\xi'^{\mathrm{T}}_{\mathbf{p}(true)} = \frac{\sigma^2}{2}\xi'_{\mathbf{p}(true)}\left[\phi''_{\mathbf{pp}(true)}\right]^{-1}\xi'^{\mathrm{T}}_{\mathbf{p}(true)} \tag{26}$$

with

$$\xi'_{\mathbf{p}} = 2\left(1+q_4\right)\frac{\left(1+q_4\right)\mathsf{I}-\left[\mathbf{q}\times\right]}{\left(1+q_4\right)\mathsf{I}+\left[\mathbf{q}\times\right]} \tag{27}$$

is equivalent to equations (10) and (13).

For a numerically stable algorithm we need to ensure that $\phi''_{0\mathbf{pp}}(\mathbf{p}_0)$ in equation (17) remains well-conditioned throughout the Gauss-Newton estimation sequence. There are two conditions that must be satisfied in order to guarantee this: having enough measurements to achieve complete attitude observability and keeping the quaternion estimate away from the negative pole at $\hat{\mathbf{q}}^{\mathrm{T}} = \begin{bmatrix}0 & 0 & 0 & -1\end{bmatrix}$. We postulate that the first condition is an obvious general pre-requisite and turn our attention to the second condition. The fact that the positive form of modified Rodrigues parameters exhibits singularity at the negative pole is well known (see Ref. 11). In this algorithm the issue manifests itself explicitly in equation (19) in which $\hat{\mathbf{q}}'_{\mathbf{p}}$ becomes identically zero at the singularity. We can easily circumvent this issue by taking advantage of the duality of the quaternion representation of attitude. It is straightforward to detect whenever the quaternion scalar part q_4 becomes negative and switch to the opposite quaternion representation by changing signs of all of the quaternion elements. This update retains the same effective attitude estimate but ensures that $\hat{\mathbf{q}}'_{\mathbf{p}}$ does not approach zero and does not make $\phi''_{0\mathbf{pp}}(\mathbf{p}_0)$ ill-conditioned.

V. Homotopy Continuation vs. Local Linearization

The Gauss-Newton estimation sequence developed in the previous section is based on the idea of the homotopy continuation. The resulting iteration in equation (17) is, of course, very similar to iterations in other local linearization techniques. It is instructive to develop these techniques here and to contrast them with the homotopy based approach.

A. Method of Steepest Descent

If we expand the original cost function ϕ in the Taylor series about the initial attitude guess, we obtain to first order

$$\phi(\mathbf{p}) = \phi(\mathbf{p}_0) + \phi'_{\mathbf{p}}(\mathbf{p}_0)\left(\mathbf{p}-\mathbf{p}_0\right) \tag{28}$$

that we can use to update the estimate according to the steepest descent technique:

$$\mathbf{p} = \mathbf{p}_0 - \alpha\phi'^{\mathrm{T}}_{\mathbf{p}}(\mathbf{p}_0). \tag{29}$$

Here α is a positive scalar that controls the step size. It can be set to some fixed value or it can be obtained based on the linear search seeking a step that leads to the largest reduction in the cost function.

B. Newton Method

Another approach is to expand the gradient $\phi'^{\mathrm{T}}_{\mathbf{p}}$ in the Taylor series about the initial guess, which to first order becomes

$$\phi'^{\mathrm{T}}_{\mathbf{p}}(\mathbf{p}) = \phi'^{\mathrm{T}}_{\mathbf{p}}(\mathbf{p}_0) + \phi''_{\mathbf{pp}}(\mathbf{p}_0)\left(\mathbf{p} - \mathbf{p}_0\right), \tag{30}$$

and then look for $\mathbf{p}$ at which the gradient vanishes. Setting the above equation to $\mathbf{0}$, we derive the Newton iteration for $\mathbf{p}$ as

$$\mathbf{p} = \mathbf{p}_0 - \left[\phi''_{\mathbf{pp}}(\mathbf{p}_0)\right]^{-1} \phi'^{\mathrm{T}}_{\mathbf{p}}(\mathbf{p}_0). \tag{31}$$

Equations (31) and (17) are almost identical: the only difference is in the Hessians. Because equation (31) evaluates the Hessian of the original cost function ϕ and not of the specially designed cost function ϕ_0, we cannot take advantage of the convenient cancellations that occur when measurement residuals vanish. It is true that if we are successful at converging toward the optimal estimate then as we get closer to it the measurement residuals in ϕ should become smaller. However, we certainly cannot make that assumption if our initial guess is poor. This means that unlike equation (22) the chain rule derivation here must include additional terms:

$$\phi''_{\mathbf{pp}}(\mathbf{p}_0) = \hat{\mathbf{q}}'^{\mathrm{T}}_{\mathbf{p}} \phi''_{\hat{\mathbf{q}}\hat{\mathbf{q}}} \hat{\mathbf{q}}'_{\mathbf{p}} + \sum_{i=1}^{4} \phi'_{q_i} q''_{i\mathbf{pp}}. \tag{32}$$

where q_i is the ith element of $\hat{\mathbf{q}}$ and ϕ'_{q_i} is simply the ith element of $\phi'_{\hat{\mathbf{q}}}$. The Hessians $q''_{i\mathbf{pp}}$ of individual quaternion elements derived from equation (19) are listed below

$$q''_{1\mathbf{pp}} = 2q_1\mathbf{q}\mathbf{q}^{\mathrm{T}} - \left(1+q_4\right)\begin{bmatrix} 3q_1 & q_2 & q_3 \\ q_2 & q_1 & 0 \\ q_3 & 0 & q_1 \end{bmatrix}, \tag{33a}$$

$$q''_{2\mathbf{pp}} = 2q_2\mathbf{q}\mathbf{q}^{\mathrm{T}} - \left(1+q_4\right)\begin{bmatrix} q_2 & q_1 & 0 \\ q_1 & 3q_2 & q_3 \\ 0 & q_3 & q_2 \end{bmatrix}, \tag{33b}$$

$$q''_{3\mathbf{pp}} = 2q_3\mathbf{q}\mathbf{q}^{\mathrm{T}} - \left(1+q_4\right)\begin{bmatrix} q_3 & 0 & q_1 \\ 0 & q_3 & q_2 \\ q_1 & q_2 & 3q_3 \end{bmatrix}, \tag{33c}$$

$$q''_{4\mathbf{pp}} = \left(1+q_4\right)\left(2\mathbf{q}\mathbf{q}^{\mathrm{T}} - \left(1+q_4\right)\mathrm{I}\right). \tag{33d}$$

We must also include additional terms when computing the Hessian $\phi''_{\hat{\mathbf{q}}\hat{\mathbf{q}}}(\hat{\mathbf{q}}_0)$ in equation (32). Depending on the form of the cost function ϕ, we derive it either from equations (6)

$$\phi''_{\hat{\mathbf{q}}\hat{\mathbf{q}}}(\hat{\mathbf{q}}_0) = \sum_{n=1}^{N} a_n \mathsf{K}_n \left(\hat{\mathbf{q}}_0^{\mathrm{T}} \mathsf{K}_n \hat{\mathbf{q}}_0 - d_n \right) + 2 \sum_{n=1}^{N} a_n \mathsf{K}_n \hat{\mathbf{q}}_0 \hat{\mathbf{q}}_0^{\mathrm{T}} \mathsf{K}_n \tag{34}$$

or from equation (11)

$$\phi''_{\hat{\mathbf{q}}\hat{\mathbf{q}}}(\hat{\mathbf{q}}_0) = \frac{1}{2} \sum_{n=1}^{N} \sum_{i=1}^{m_n} \sum_{j=1}^{m_n} \bar{r}_{ij}^{-1} \left[\mathsf{K}_{nj} \left(\hat{\mathbf{q}}_0^{\mathrm{T}} \mathsf{K}_{ni} \hat{\mathbf{q}}_0 - d_{ni} \right) + \mathsf{K}_{ni} \left(\hat{\mathbf{q}}_0^{\mathrm{T}} \mathsf{K}_{nj} \hat{\mathbf{q}}_0 - d_{nj} \right) + 2 \left(\mathsf{K}_{ni} \hat{\mathbf{q}}_0 \hat{\mathbf{q}}_0^{\mathrm{T}} \mathsf{K}_{nj} + \mathsf{K}_{nj} \hat{\mathbf{q}}_0 \hat{\mathbf{q}}_0^{\mathrm{T}} \mathsf{K}_{ni} \right) \right] \tag{35}$$

C. Comparison to Homotopy Gauss-Newton Method

The only difference between equations (29) and (31), and equation (17) is in the term pre-multiplying $\phi'^{\mathrm{T}}_{\mathbf{p}}(\mathbf{p}_0)$ but this difference is crucial.

The steepest descent is the simplest of the three algorithms. It does not use any second order derivative information relying instead on a single scalar scale to determine the appropriate step size along the negative gradient direction. This apparent simplicity comes at a cost: the biggest challenge in designing this algorithm is in selecting the appropriate step size scale α. Fixing it to a small value will slow down the convergence, fixing it to a value that is too large will make the estimation sequence diverge, searching for it will require additional iterations which still must be bounded by some *a priori* limits. For a nonlinear problem, the algorithm tuned for a particular initial guess may fail or be too slow for a different initial guess or for a different set of measurements. Thus, there are significant disadvantages to using the steepest descent approach compared to either the Newton estimation sequence of equation (31) or the homotopy Gauss-Newton estimation sequence of equation (17). Our numerical tests presented in the next section highlight these disadvantages: the steepest descent algorithm either takes many iterations to converge or fails to consistently converge across various initial guesses and measurement sets.

It may seem that the Newton estimation technique has the advantage over the homotopy continuation approach because the latter uses the Gauss-Newton approximation of the Hessian listed in equation (17) instead of the actual Hessian listed in equation (31). However, when equation (17) is interpreted as the first step along the homotopy continuation path, it can be seen to represent the *departure* from the current guess toward the optimal estimate *more accurately* than equation (31) can be seen to represent the attempt to *arrive* at the optimal estimate from the current guess (especially when the guess is poor). The numerical analysis presented in the next section supports the idea that for this nonlinear estimation problem the homotopy Gauss-Newton method captures the overall topology of the problem better than the Newton method. We can expect that, once estimates enter some small neighborhood of the optimal estimate, the Newton method should achieve faster convergence than the Gauss-Newton method because the Hessian in the Newton method includes additional terms that describe the cost function better in this small neighborhood. However, if the current guess is far from the optimal estimate, the additional terms may hinder rather than improve convergence. This is particularly evident in the following special case. Consider what happens when the current guess $\hat{\mathbf{q}}$ aligns closely with the gradient $\phi'^{\mathrm{T}}_{\hat{\mathbf{q}}}$. Note from equations (18) and (19) that only components of the gradient $\phi'^{\mathrm{T}}_{\hat{\mathbf{q}}}$ that are orthogonal to $\hat{\mathbf{q}}$ can move the next estimate – a direct consequence of the quaternion norm constraint. If the Hessian in the update happens to move the estimate further and further toward alignment with $\phi'^{\mathrm{T}}_{\hat{\mathbf{q}}}$, the estimation sequence "converges" but never reaches the optimal estimate. As shown in the next section, the Newton estimation sequence exhibits this behavior often unlike the homotopy Gauss-Newton estimation sequence which never does. Moreover, the homotopy Gauss-Newton algorithm makes the alignments between $\hat{\mathbf{q}}$ and $\phi'^{\mathrm{T}}_{\hat{\mathbf{q}}}$ unstable: the Gauss-Newton sequence successfully departs from them given even small initial misalignments. Let

$$\lambda = \left| \cos\delta \right| = \left\| \phi'_{\hat{\mathbf{q}}}(\hat{\mathbf{q}}) \right\|^{-1} \left| \phi'_{\hat{\mathbf{q}}}(\hat{\mathbf{q}}) \hat{\mathbf{q}} \right| \tag{36}$$

denote the alignment parameter defined as the absolute value of the cosine of the angle between the gradient $\phi'^{\mathrm{T}}_{\hat{\mathbf{q}}}$ and the estimate $\hat{\mathbf{q}}$. When close to alignment, $\lambda \to 1$ and its increment after one iteration is

$$\Delta\lambda = -\left\|\phi'_{\hat{\mathbf{q}}}(\hat{\mathbf{q}})\right\|^{-1} \phi'_{\hat{\mathbf{q}}}(\hat{\mathbf{q}})\hat{\mathbf{q}}'_{\mathbf{p}}\mathsf{H}^{-1}\hat{\mathbf{q}}'^{\mathrm{T}}_{\mathbf{p}}\phi'^{\mathrm{T}}_{\hat{\mathbf{q}}}(\hat{\mathbf{q}}), \tag{37}$$

where H is a symmetric matrix defined differently for each estimation method. For the steepest descent method, $\mathsf{H} = \alpha^{-1}\mathsf{I}$, for the Newton method, it is defined as the Hessian in equations (34) or (35) and, for the homotopy Gauss-Newton method, it is defined as the Hessian in equations (23) or (24). Note that, unlike the other two methods, the Newton method does not guarantee that its Hessian is positive definite. This means that it is possible for the Newton estimation sequence to approach $\lambda \to 1$ and still produce positive increments $\Delta\lambda$ making the gradient $\phi'^{\mathrm{T}}_{\hat{\mathbf{q}}}$ and the estimate $\hat{\mathbf{q}}$ more and more aligned. This situation is not possible for the other two methods, which, as their estimation sequences approach $\lambda \to 1$, are guaranteed to produce negative increments $\Delta\lambda$ reversing the approach and preventing the sequences from converging to incorrect estimates.

VI. Numerical Examples

We examine performance of the estimators using several numerical examples. In all cases we set the true attitude to

$$\hat{\mathbf{q}}_{true} = \begin{bmatrix} -0.1160 & -0.0429 & 0.1760 & 0.9766 \end{bmatrix}^{\mathrm{T}}. \tag{38}$$

The measurements are based on three body-fixed directions aligned with

$$\mathbf{s}_1 = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^{\mathrm{T}},\ \mathbf{s}_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^{\mathrm{T}} \text{ and } \mathbf{s}_3 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^{\mathrm{T}} \tag{39}$$

and four reference directions

$$\mathbf{r}_1 = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}^{\mathrm{T}},\ \mathbf{r}_2 = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^{\mathrm{T}},\ \mathbf{r}_3 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^{\mathrm{T}} \text{ and } \mathbf{r}_4 = \mathbf{r}_2 \times \mathbf{r}_3. \tag{40}$$

The zero-mean Gaussian noise terms are assumed to be uncorrelated with variances equal to 10^{-5}. In all cases we stop the sequence when the cost function ϕ is reduced below 10^{-8} or when the rotational change in attitude during a single iteration falls below 10^{-5} radians (or after 200 iterations), whichever comes first.

In the first example we include 6 noiseless measurements using pairs $(\mathbf{r}_1, \mathbf{s}_1)$, $(\mathbf{r}_2, \mathbf{s}_1)$, $(\mathbf{r}_3, \mathbf{s}_1)$ and $(\mathbf{r}_1, \mathbf{s}_2)$, $(\mathbf{r}_2, \mathbf{s}_2)$, $(\mathbf{r}_3, \mathbf{s}_2)$. Note that the two body-fixed directions selected in this example are orthogonal. We set the initial attitude guess to

$$\hat{\mathbf{q}}_0 = \begin{bmatrix} 0.6830 & 0 & -0.6830 & 0.2588 \end{bmatrix}^{\mathrm{T}}, \tag{41}$$

which is far from the true attitude and serves our goal of highlighting the differences among the three estimation methods. After fine-tuning with the step size scale set to 0.7854, the steepest descent algorithm based on equation (29) achieves its fastest convergence at 83 iterations (see Figs. 1 and 2) – significantly slower than the homotopy Gauss-Newton method that will be examined later in this section. Smaller step size scales obviously result in slower convergence and larger step size scales in this example cause divergence even in the absence of noise. Note that the fine-tuned steepest descent algorithm still exhibits its characteristic slow convergence clearly visible between iterations 5 and 45 in Fig. 2. The Newton estimation sequence based on equation (31) also performs poorly in this example and simply fails to converge to the correct attitude (see Fig. 3). Figure 4 shows that after the first 4 iterations during which the cost function changes dramatically it settles into a steady growth and converges to a value greater than the initial value. In Fig. 5 it is evident that after the first iteration the Hessian becomes significantly ill-conditioned indicating that it may be inadequate for capturing the nonlinear relationship between the poor attitude guess and the true attitude. Contrast these results with those for the homotopy Gauss-Newton method

shown in Figs. 6-8. The successful convergence of the quaternion estimate is shown in Fig. 6, the successful minimization of the cost function which is steadily decreasing is shown in Fig. 7 and a better behavior of the Hessian condition number is shown in Fig. 8. Note that after the 10th iteration the quaternion estimate is switched to its opposite in order to keep it away from the singularity (see Fig. 6). It is also interesting to note that in this example the Newton method fails because the estimate and the cost function gradient become progressively more aligned (see Figs. 9 and 10); however, when iterations are switched to the homotopy Gauss-Newton method, it successfully recovers from the nearly aligned poor estimate and eventually converges to the correct estimate (again see Figs. 9 and 10). The optimal attitude error covariance matrix for this case of 6 measurements with two orthogonal body-fixed directions is listed below

$$\mathsf{P}_{\xi\xi}^{*} = \begin{bmatrix} 6.4579 & -0.0051 & 6.4198 \\ -0.0051 & 6.5295 & 0.5290 \\ 6.4198 & 0.5290 & 10.3467 \end{bmatrix} \times 10^{-6}. \tag{42}$$

In the second example we include 8 measurements using the same 6 pairs from the first example plus two more pairs using the forth reference direction: $(\mathbf{r}_4, \mathbf{s}_1)$ and $(\mathbf{r}_4, \mathbf{s}_2)$. Our expectation that both the convergence and the optimal covariance should improve due to the addition of new measurements is confirmed in Figs. 11-13 and in equation (43):

$$\mathsf{P}_{\xi\xi}^{*} = \begin{bmatrix} 3.7651 & 0.1383 & 3.4016 \\ 0.1383 & 4.1267 & -0.9355 \\ 3.4016 & -0.9355 & 5.8611 \end{bmatrix} \times 10^{-6}. \tag{43}$$

The homotopy Gauss-Newton estimation sequence converges in 24 iterations with 8 measurements compared to 45 iterations with 6 measurements. The optimal covariance with 8 measurements is about half of the size of the optimal covariance with 6 measurements (compare equations (42) and (43)). We examine the effect of noisy measurements for this example in Fig. 14. It shows the element-by-element difference between quaternion estimates produced with and without noisy measurements. At the end the differences are below 2×10^{-3} which correspond to the attitude rotational difference of about 0.1 degree.

In the third example we keep the same number of 8 measurements but replace one of the body-fixed directions $\mathbf{s}_2$ with $\mathbf{s}_3$. Note that in this case the two body-fixed directions $\mathbf{s}_1$ and $\mathbf{s}_3$ are no longer orthogonal. The results are shown in Figs. 15-17 and in equation (44):

$$\mathsf{P}_{\xi\xi}^{*} = \begin{bmatrix} 7.9247 & 4.1370 & 4.5840 \\ 4.1370 & 4.2214 & 0.9485 \\ 4.5840 & 0.9485 & 6.2933 \end{bmatrix} \times 10^{-6} \tag{44}$$

The results indicate that the homotopy Gauss-Newton method continues to perform successfully. The optimal covariance in this case is somewhat larger and more cross-correlated than the optimal covariance in the case of two orthogonal body-fixed directions (compare equations (43) and (44)).

In the final test we examine robustness of the algorithms by attempting to converge to the true estimate from a number of initial guesses spread on a grid defined as follows: right ascension of the eigen-axis is varied from 0 to 360 degrees in 15 degree increments, declination of the eigen-axis is varied from -90 to 90 degrees in 15 degree increments, and the eigen-angle is varied from 0 to 180 degrees in 15 degree increments. This generates 4225 test cases. The homotopy Gauss-Newton method is successful in all of them converging to the true estimate in no more than 23 iterations and maintaining the Hessian condition number below 35. The Newton method is successful in only 2076 cases or in about 50 % of all test cases. It also fails to maintain the Hessian well-conditioned: its condition number at times reaches above 50000. The steepest descent algorithm, using the step size scale of 0.7854 originally

fine-tuned for the first example, completely fails in this example. It does not converge in 200 iterations for any of the initial guesses on the grid. The new optimum step size scale for which all test cases converge in fewer than 200 iterations is found to be 0.3491. The results are summarized in Figs. 18 and 19. Note that in all test cases the steepest descent algorithm converges slower than the homotopy Gauss-Newton method. Also note that the steepest descent convergence varies widely for different initial guesses and that generally it takes more iterations to converge for poorer initial guesses (see Fig. 18). While in some cases the Newton method is somewhat faster and in other cases it is somewhat slower than the homotopy Gauss-Newton method, when successful it converges on average about as fast as the homotopy Gauss-Newton method. However, as mentioned earlier, the Newton method exhibits a very high rate of failure (about 50 %). These failures are shown in Fig. 18 as iterations maxed out at 200. The method also fails to maintain its Hessian well-conditioned as shown in Fig. 19. Finally, it is instructive to show the largest increase and the largest decrease per iteration experienced by the cost function during minimization by each of the three methods. We collect these results for each of the 4225 test cases and show them for each estimation method in Figs. 20-22. As expected, the largest decreases per iteration tend to occur for large initial errors which correspond to large cost function values. This characteristic is quite evident for the steepest descent method in Fig. 20 and for the homotopy Gauss-Newton method in Fig. 22. It is much less evident for the Newton method in Fig. 21. This figure also shows that the Newton method exhibits the largest increases in the cost function which is not unexpected considering its high rate of failure. The steepest descent method causes occasional small increases in the cost function indicating that its step size scale is fine-tuned for the fastest convergence and that attempting to increase the step size scale any further is going to cause divergence in at least some of the test cases (see Fig. 20). The homotopy Gauss-Newton method demonstrates the most robust behavior by consistently and steadily decreasing the cost function for all test cases (see Fig. 22).

VII. Conclusion

The problem of the optimal attitude determination using only angular measurements was examined in this paper. The measurement model and the maximum likelihood cost function were formulated explicitly in quaternions. The Gauss-Newton estimation sequence based on the homotopy continuation approach was derived and compared to other local linearization techniques. The new approach was found to be faster and more reliable than other methods. The approach can be applied to determine the attitude for a spacecraft that operates with sensors measuring only angular data or for a spacecraft that operates in a contingency situation with sensors nominally capable of measuring complete vector or attitude data failing to produce measurements about certain axes. The approach can also be applied to determine attitude of a vehicle using normalized GPS phase difference measurements.

References

[1]Shuster, M. D., and Oh, S. D., "Three-Axis Attitude Determination from Vector Observations," *Journal of Guidance and Control*, Vol. 4, No. 1, 1981, pp. 70-77.

[2]Markley, F. L., "Attitude Determination Using Vector Observations and the Singular Value Decomposition," *The Journal of the Astronautical Sciences*, Vol. 36, No. 3, 1988, pp. 245-258.

[3]Markley, F. L., "Attitude Determination Using Vector Observations: A Fast Optimal Matrix Algorithm," *The Journal of the Astronautical Sciences*, Vol. 41, No. 2, 1993, pp. 261-280.

[4]Mortari, D., "Second Estimator for the Optimal Quaternion," *Journal of Guidance, Control and Dynamics*, Vol. 23, No. 5, 2000, pp. 885-888.

[5]Markley, F. L., and Mortari, D., "Quaternion Estimation Using Vector Observations," *The Journal of the Astronautical Sciences*, Vol. 48, Nos. 2 and 3, 2000, pp. 359-380.

[6]Shuster, M. D., "Deterministic Three-Axis Attitude Determination," *The Journal of the Astronautical Sciences*, Vol. 52, No. 3, 2004, pp. 405-419.

[7]Cohen, C. E., "Attitude Determination Using GPS," Ph.D. Dissertation, Stanford Univ., 1992.

[8]Crassidis, J. L., and Markley, F. L., "New Algorithm for Attitude Determination Using Global Positioning System Signals," *Journal of Guidance, Control and Dynamics*, Vol. 20, No. 5, 1997, pp. 891-896.

[9]Crassidis, J. L., Lightsey, E. G., and Markley, F. L., "Efficient and Optimal Attitude Determination Using Recursive Global Positioning System Signal Operations," *Journal of Guidance, Control and Dynamics*, Vol. 22, No. 2, 1999, pp. 193-201.

[10]Tanygin, S., and Shuster, M. D., "Spin-Axis Attitude Estimation," *The Journal of the Astronautical Sciences*, Vol. 55, No. 1, 2007, pp. 107-139.

[11]Shuster, M. D., "A Survey of Attitude Representations," *The Journal of the Astronautical Sciences*, Vol. 41, No. 4, 1993, pp. 439-517.

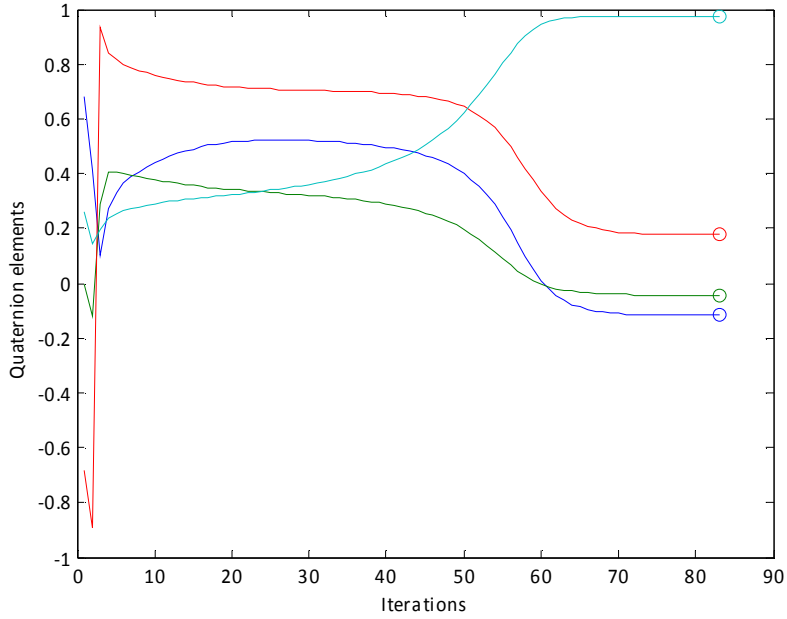

Figure 1. Quaternion estimation using the fine-tuned steepest descent method with 6 noiseless measurements and 2 orthogonal body-fixed directions.

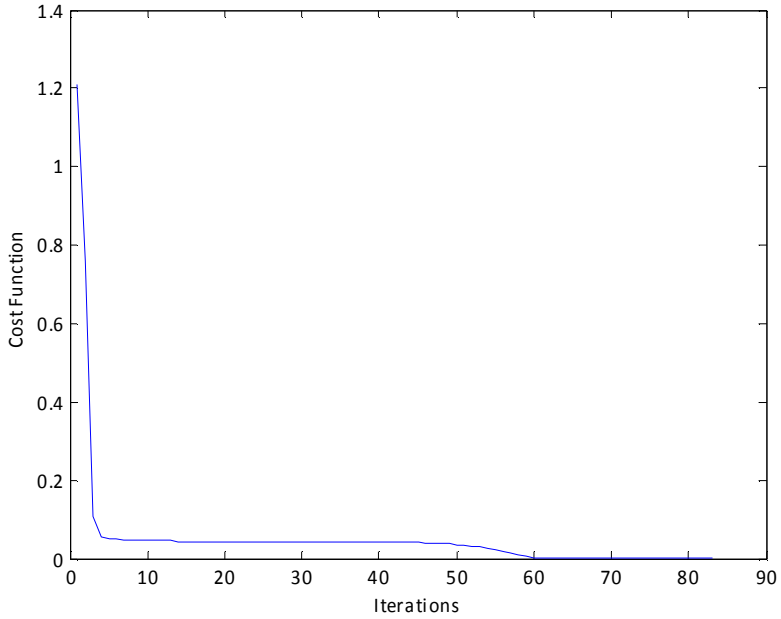

Figure 2. Cost function minimization using the fine-tuned steepest descent method with 6 noiseless measurements and 2 orthogonal body-fixed directions.

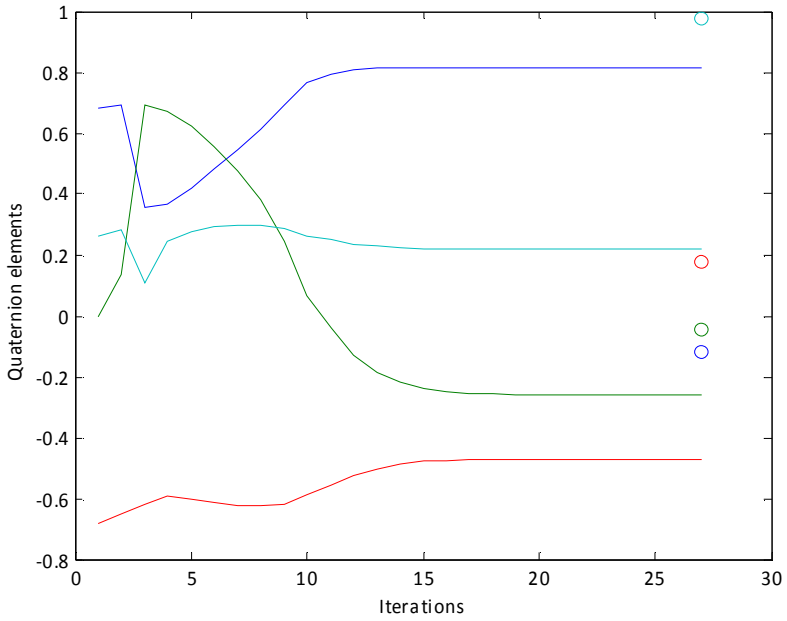

Figure 3. Failed quaternion estimation using the Newton method with 6 noiseless measurements and 2 orthogonal body-fixed directions.

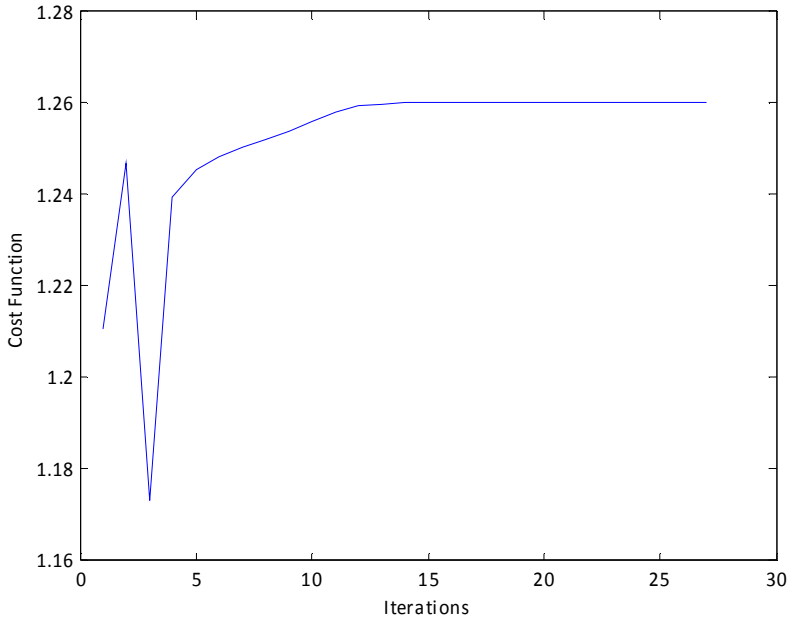

Figure 4. Failed cost function minimization using the Newton method with 6 noiseless measurements and 2 orthogonal body-fixed directions.

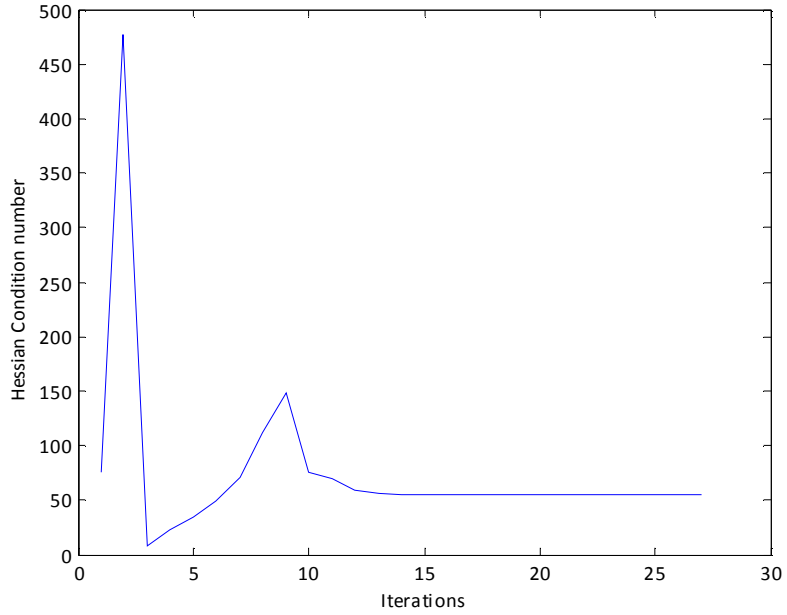

Figure 5. Hessian condition number during estimation using the Newton method with 6 noiseless measurements and 2 orthogonal body-fixed directions.

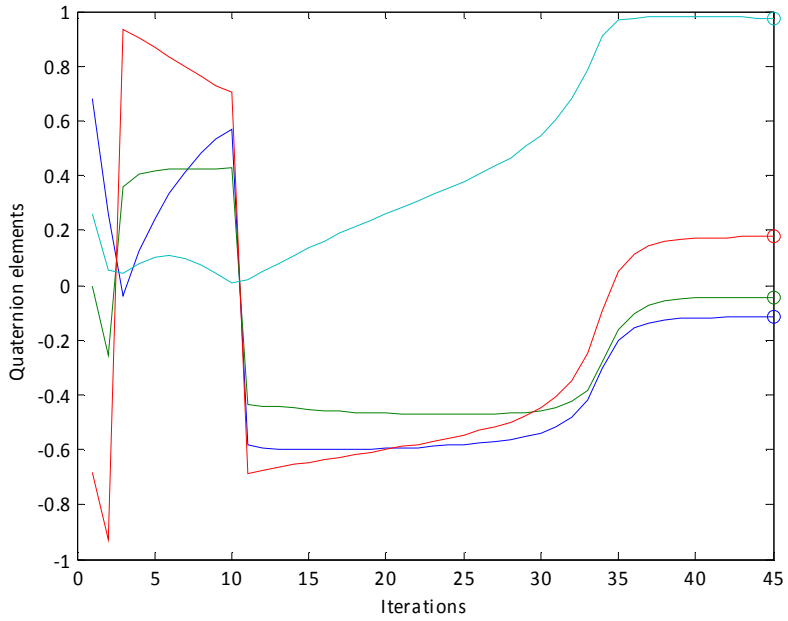

Figure 6. Quaternion estimation using the homotopy Gauss-Newton method with 6 noiseless measurements and 2 orthogonal body-fixed directions.

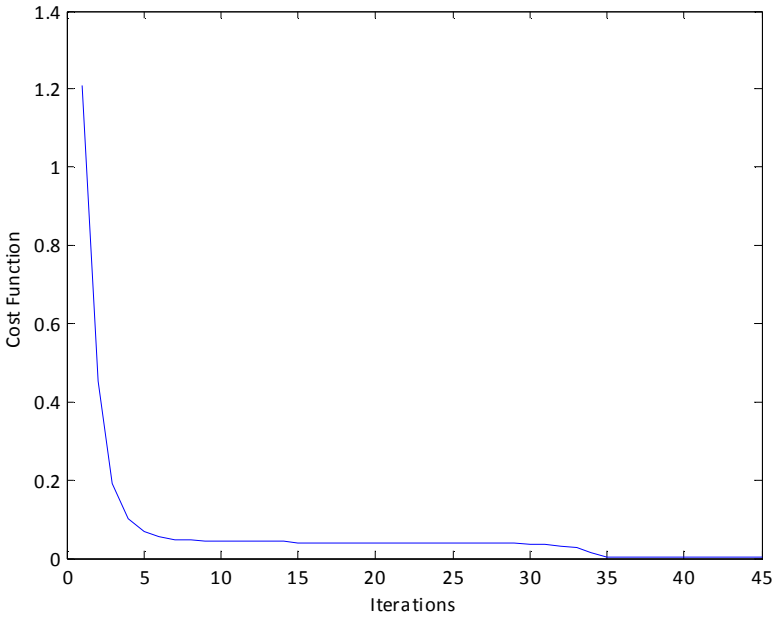

Figure 7. Cost function minimization using the homotopy Gauss-Newton method with 6 noiseless measurements and 2 orthogonal body-fixed directions.

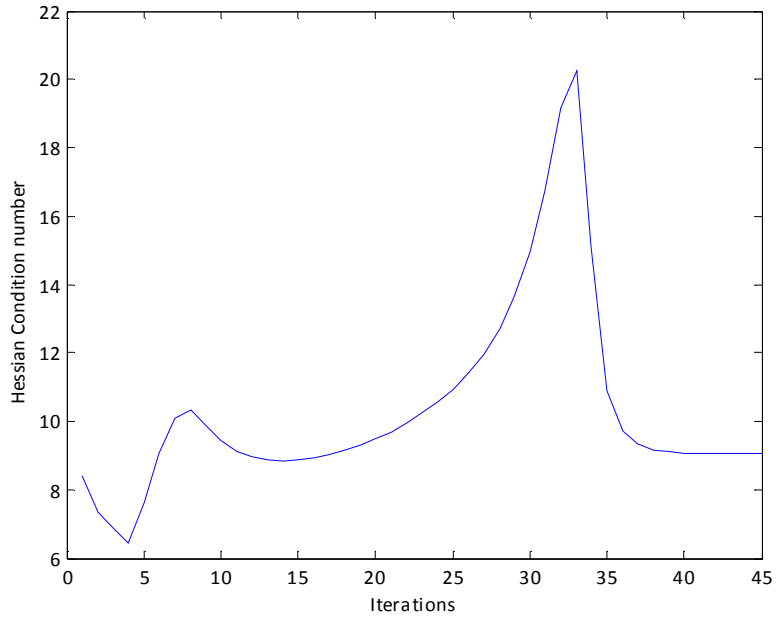

Figure 8. Hessian condition number during estimation using the homotopy Gauss-Newton method with 6 noiseless measurements and 2 orthogonal body-fixed directions.

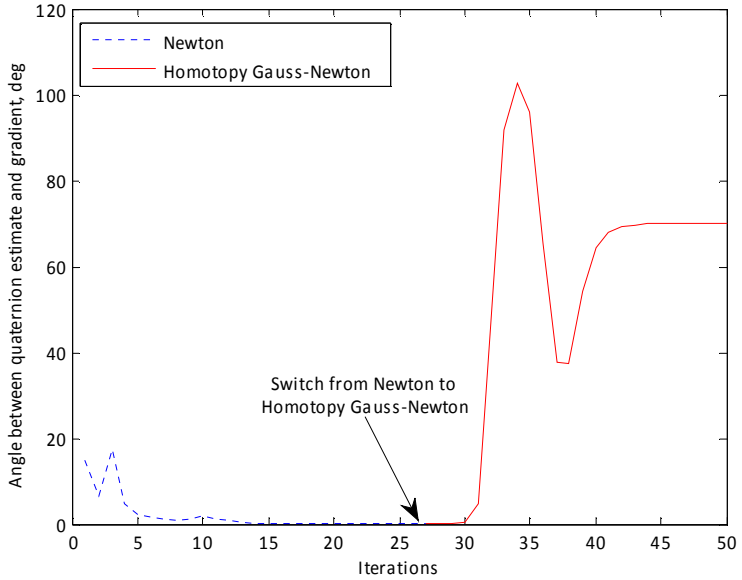

Figure 9. Alignment angle between quaternion estimate and cost function gradient during estimation using first the Newton method and then the homotopy Gauss-Newton method with 6 noiseless measurements and 2 orthogonal body-fixed directions.

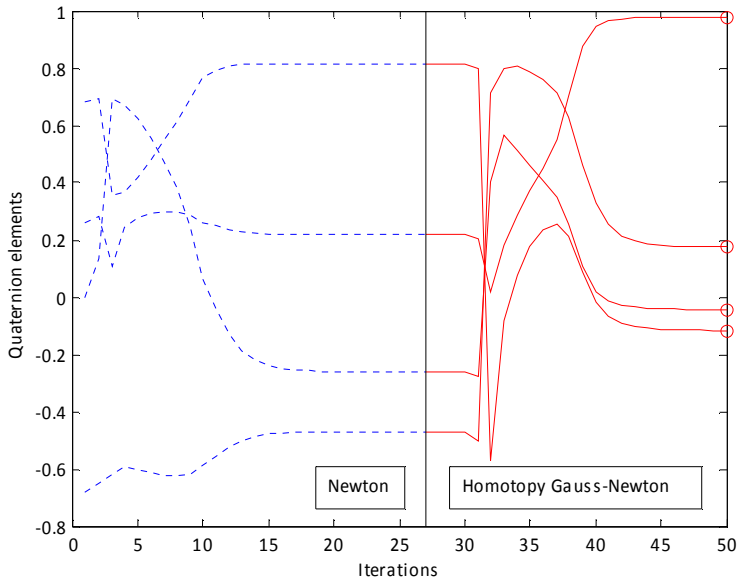

Figure 10. Quaternion estimation using first the Newton method and then the homotopy Gauss-Newton method with 6 noiseless measurements and 2 orthogonal body-fixed directions.

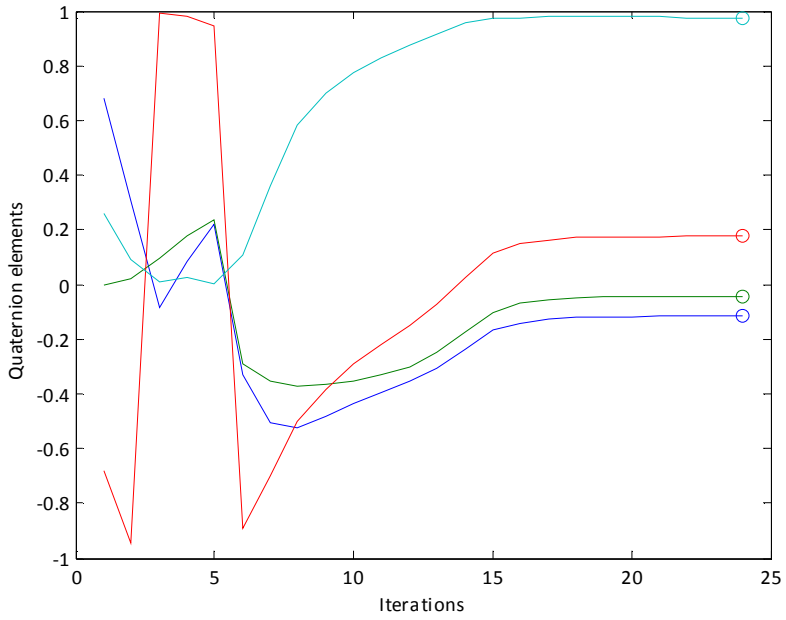

Figure 11. Quaternion estimation using the homotopy Gauss-Newton method with 8 noiseless measurements and 2 orthogonal body-fixed directions.

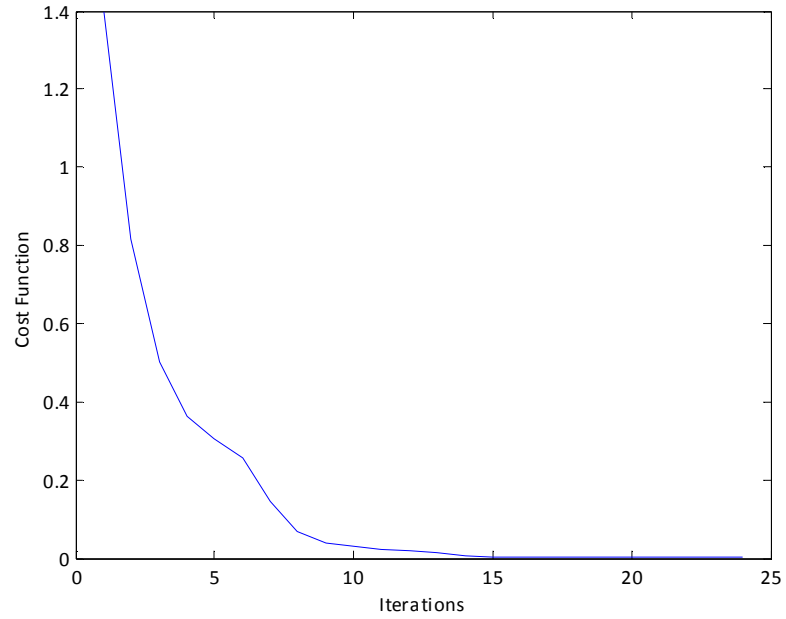

Figure 12. Cost function minimization using the homotopy Gauss-Newton method with 8 noiseless measurements and 2 orthogonal body-fixed directions.

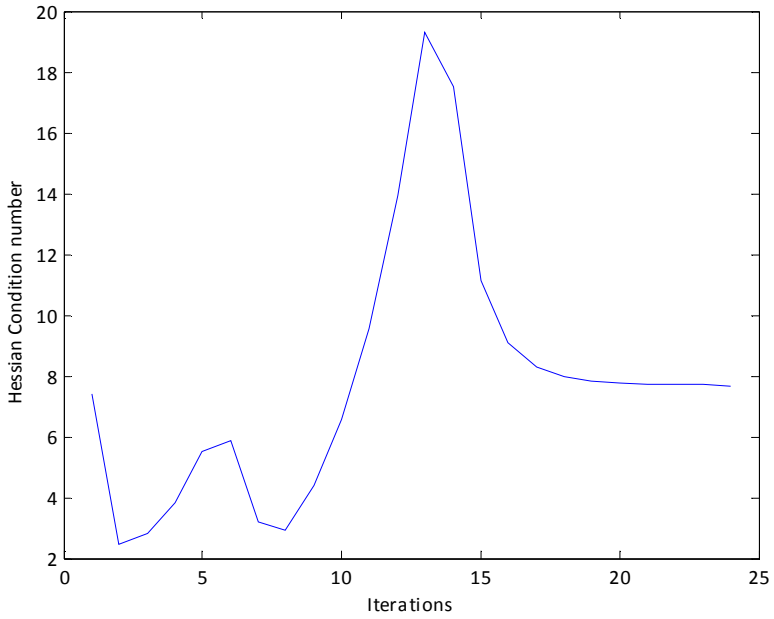

Figure 13. Hessian condition number during estimation using the homotopy Gauss-Newton method with 8 noiseless measurements and 2 orthogonal body-fixed directions.

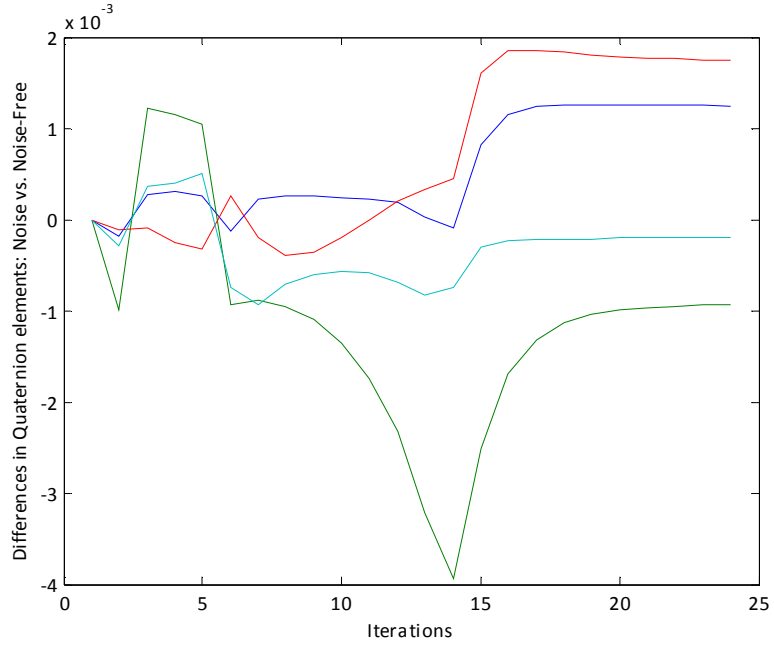

Figure 14. Differences between noisy and noiseless quaternion estimation using the homotopy Gauss-Newton method with 8 measurements and 2 orthogonal body-fixed directions.

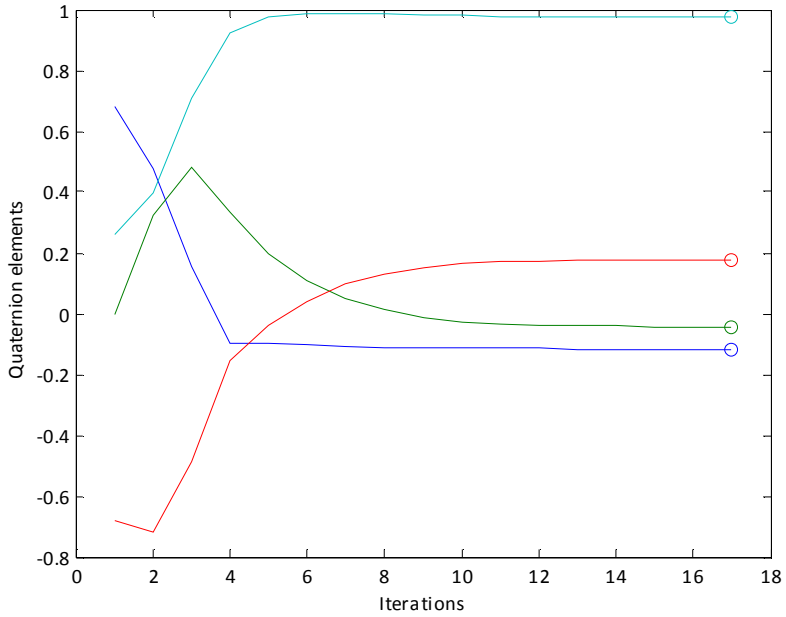

Figure 15. Quaternion estimation using the homotopy Gauss-Newton method with 8 noiseless measurements and 2 non-orthogonal body-fixed directions.

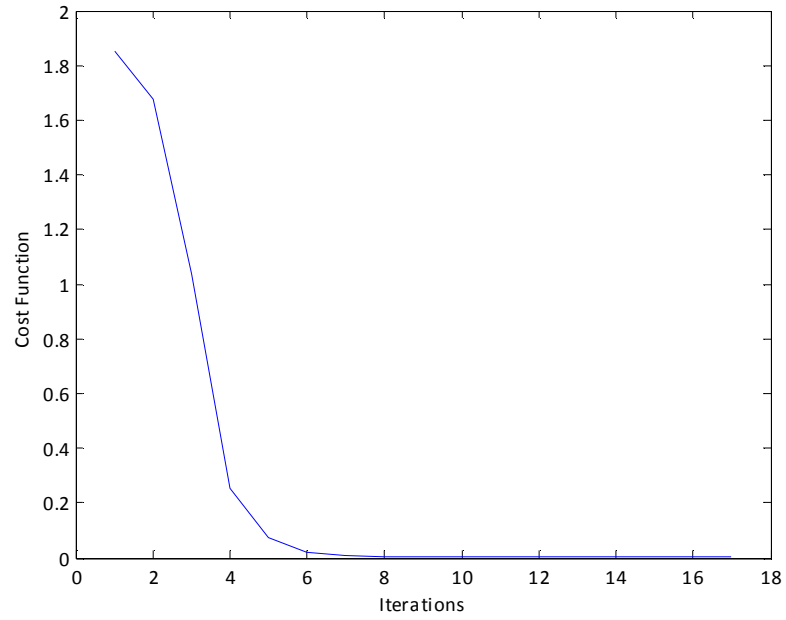

Figure 16. Cost function minimization using the homotopy Gauss-Newton method with 8 noiseless measurements and 2 non-orthogonal body-fixed directions.

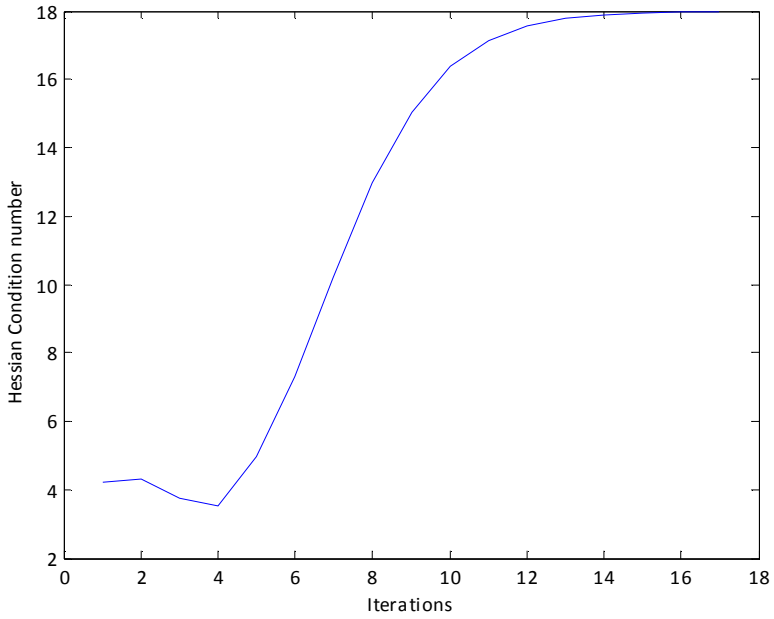

Figure 17. Hessian condition number during estimation using the homotopy Gauss-Newton method with 8 noiseless measurements and 2 non-orthogonal body-fixed directions.

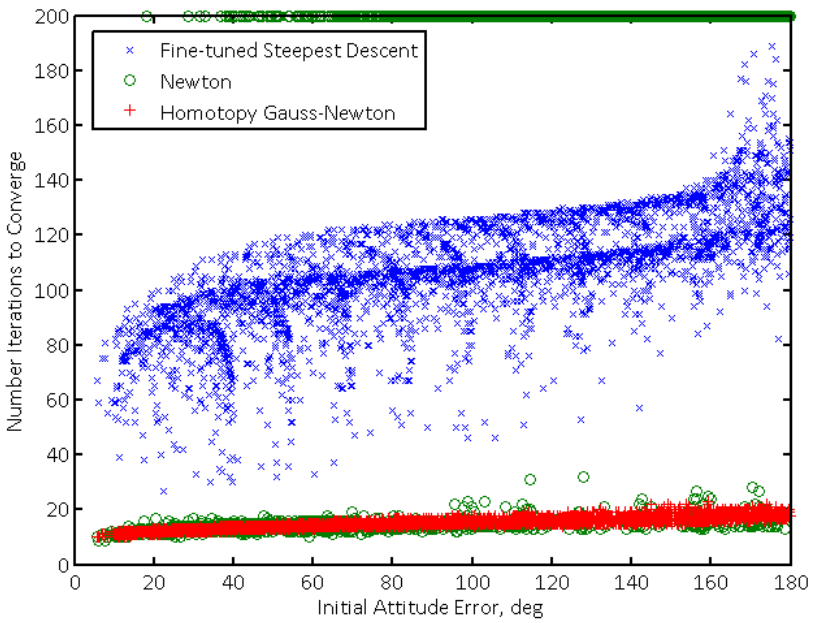

Figure 18. Estimation sequence convergence for various initial guesses with 8 noiseless measurements and 2 non-orthogonal body-fixed directions.

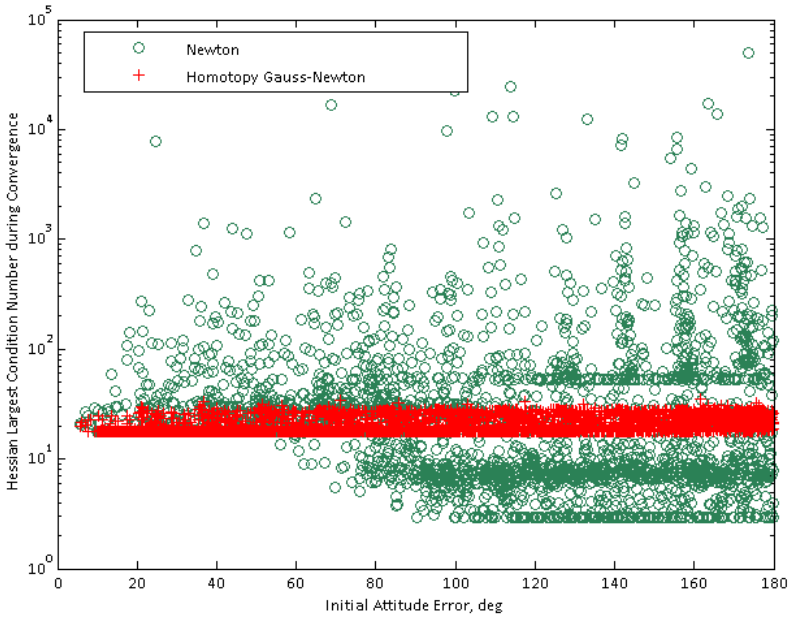

Figure 19. The Hessian largest condition number during estimation for various initial guesses with 8 noiseless measurements and 2 non-orthogonal body-fixed directions.

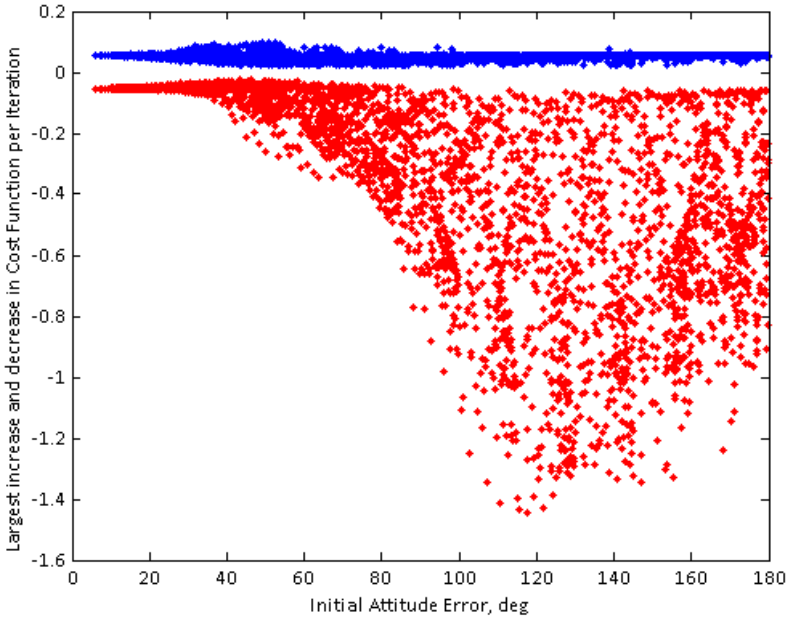

Figure 20. The largest increase and decrease in the cost function per iteration during the steepest descent estimation for various initial guesses with 8 noiseless measurements and 2 non-orthogonal body-fixed directions.

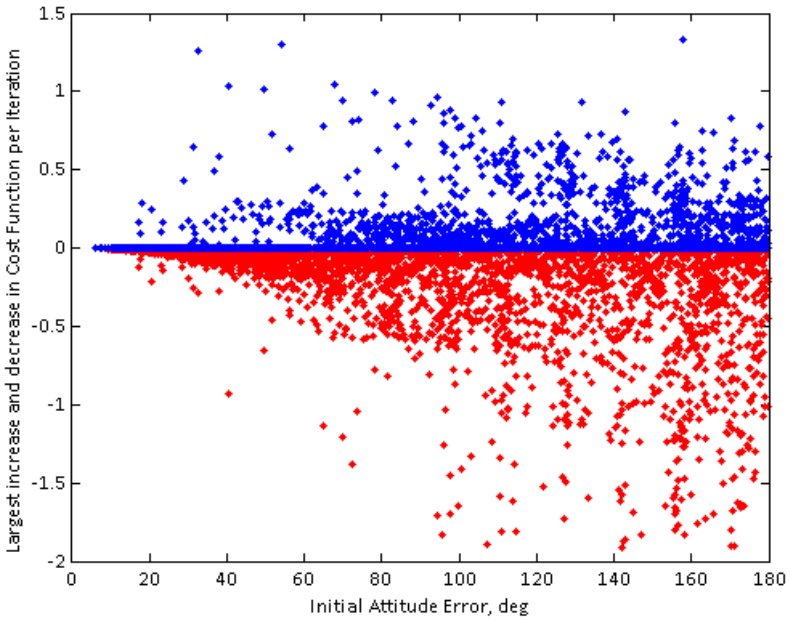

Figure 21. The largest increase and decrease in the cost function per iteration during the Newton method estimation for various initial guesses with 8 noiseless measurements and 2 non-orthogonal body-fixed directions.

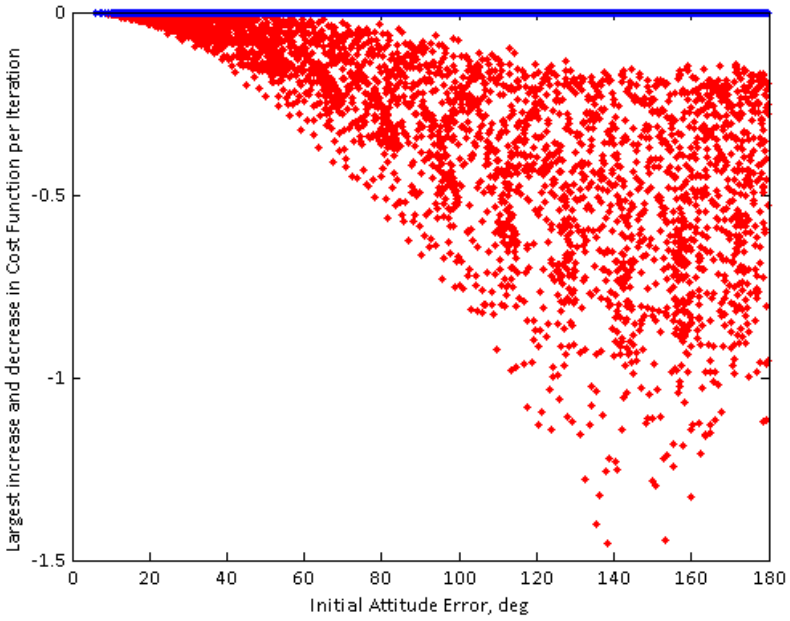

Figure 22. The largest increase and decrease in the cost function per iteration during the homotopy Gauss-Newton method estimation for various initial guesses with 8 noiseless measurements and 2 non-orthogonal body-fixed directions.