AIAA 2008-7389

ADAPTIVE HUBER–BASED FILTERING USING PRO JECTION STATISTICS: APPLICATION TO SPACECRAFT ATTITUDE ESTIMATION

Christopher D. Karlgaard and Hanspeter Schaub

AIAA Guidance, Navigation and Control Conference
Honolulu, Hawaii, August 18–21, 2008

Adaptive Huber–Based Filtering Using Projection Statistics: Application to Spacecraft Attitude Estimation

Christopher D. Karlgaard*
Analytical Mechanics Associates, Inc., Hampton, VA, 23666

Hanspeter Schaub†
University of Colorado, Boulder, CO, 80309-0431

This paper discusses the development of an adaptive discrete-time robust filtering technique. The technique is based on a recursive form of Huber's mixed minimum ℓ_1/ℓ_2 norm approach to estimation, which is robust with respect to deviations from the assumed Gaussian error probability distributions inherent to the Kalman filter. An adaptive scheme is proposed whereby the filter can estimate the process noise and measurement noise covariance matrices along with the state estimate and state estimate error covariance matrix. The adaptation technique also adopts a robust approach to estimating these covariances that can resist the effects of outliers. The adaptive/robust filter is applied to the spacecraft attitude quaternion estimation problem, using a standard multiplicative approach for handling the quaternion norm constraint. Simulation cases involving both Gaussian and non-Gaussian error distributions are provided.

I. Introduction

The Kalman filter is a recursive weighted least–squares or minimum ℓ_2 norm estimation procedure, and is a maximum likelihood technique assuming that the error statistics follow Gaussian probability distributions. It is known that the performance of least–squares and the Kalman filter can rapidly degrade when the true underlying measurement error probability distributions deviate from the assumed Gaussian case. For this reason, it is important to develop filtering methods that are robust with respect to deviations from the assumed Gaussian distribution. One technique is the Huber filter,[1–4] which is a blended minimum ℓ_1 and ℓ_2 norm estimator. This estimator exhibits reduced sensitivity to deviations in the assumed measurement error probability distributions.

In addition to the form of the error probability density function, the standard Kalman filter also assumes that the statistics of the distribution, namely its mean and covariance, are known quantities. If these assumed parameters of the distribution differ greatly from the true parameters, then the filter can exhibit large errors and possibly divergence.[5] This behavior is possible even if the true error distributions are Gaussian. To remedy these problems, adaptive filtering techniques have been introduced in order to automatically tune the Kalman filter by estimating the measurement and process noise covariances.[6]

The Huber–based formulations of the filtering problem also make some assumptions regarding the distribution, namely the Huber approach considers a class of contaminated densities in the neighborhood of the Gaussian density. Essentially the method assumes that the statistics of the main Gaussian density are known, as well as the ratio or percentage of the contamination (although the filter makes no assumption of the nature of the contaminating density other than it be symmetric with finite variance). The technique can be improved upon by the introduction of a method to adaptively estimate the noise statistics along with the state and state error covariance matrix.

One technique in common use for adaptively estimating the noise statistics in real–time filtering applications is known as *covariance matching*. The covariance matching technique is an intuitively appealing approach in which the measurement noise and process noise covariances are determined in such a way that the true residual covariance matches the theoretically predicted covariance. The true residual covariance is approximated in real time using the

*Senior Project Engineer, Analytical Mechanics Associates, Inc., 303 Butler Farm Road, Suite 104A, Hampton VA, 23666. karlgaard@ama-inc.com. Senior Member AIAA.

†Associate Professor, Aerospace Engineering Sciences Department, University of Colorado, Boulder, CO, 80309-0431. hanspeter.schaub@colorado.edu. Associate Fellow AIAA.

sample covariance, over some finite buffer of stored residuals. An explicit solution to the covariance matching problem is proposed by Myers and Tapley in Ref. 7, which uses the sample mean and sample covariance of the stored residuals in order to estimate the measurement and process noise covariance matrices. The drawback to this approach is that the presence of outliers and non-Gaussianity can create problems of robustness with the use of sample covariance. Therefore some additional steps must be taken to identify the outliers before forming the covariance estimates.

The purpose of this paper is to merge the covariance matching technique with the robust Huber–based Kalman filter, including a means by which to identify and properly weight any outliers in the stored residual sequence that are used to compute the measurement and process noise covariance estimates. Two techniques for handling such outliers are introduced. The first method is a classical technique based on the weighted Euclidean distance between individual data points and the sample mean, where the weights are determined by the inverse of the sample covariance matrix. These quantities, known as Mahalanobis distances,[8] are non–robust in nature since they involve the sample mean and covariance.[9–11] In particular they are extremely sensitive to clusters of outliers.[12] Therefore, a second method is introduced that replaces the sample mean and covariance with the sample median and median absolute deviation. This approach yields a Mahalanobis-like quantity, known as the Projection Statistics,[9] that can deal effectively with clustered outliers. The Mahalanobis distance or Projection Statistics can then be used to calculate weights that are used to remove the effect of outliers on the measurement and process noise covariance estimates computed in the adaptive filter. This hybrid adaptive/robust filter is then applied to the spacecraft attitude estimation problem, with numerical examples involving Gaussian and non–Gaussian errors.

It is anticipated that the hybrid adaptive/robust covariance matching technique proposed in this paper can be modified in order to be incorporated into several other state estimation approaches beyond the Kalman filter. For instance, the divided difference filter can be generalized as in Ref. 4 to include a Huber–style measurement update, and can be augmented as in Ref. 13 to use the Myers–Tapley[7] covariance matching technique. It would be relatively straightforward to blend these techniques together into an adaptive, Huber–based divided difference filter. In addition, it may be possible to modify certain classes of particle filtering techniques such as the ensemble Kalman filter[14] to use the robust approach considered in this paper.

II. Huber–Based Filtering Techniques

This section discusses the proposed filtering technique that is robust with respect to deviations in the assumed measurement error probability distributions. The technique is a recursive form of the generalized maximum–likelihood regression theory introduced by Huber.[1] The method is described in detail in Ref. 2, but a short review is provided as follows. The review is concerned with the discrete-time linear system

$$\boldsymbol{x}_{k+1} = \boldsymbol{F}_k \boldsymbol{x}_k + \boldsymbol{G}_k \boldsymbol{v}_k \tag{1}$$

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{w}_k \tag{2}$$

where $\boldsymbol{x}_k$ is the state at time k, $\boldsymbol{y}_k$ is the measurement, $\boldsymbol{v}_k$ is an unbiased random input with covariance $\boldsymbol{Q}_k$, and $\boldsymbol{w}_k$ is an unbiased random measurement error with covariance $\boldsymbol{R}_k$.

A. State Prediction

Given an initial estimate of the state at time k denoted by $\hat{\boldsymbol{x}}_k$ and the associated state estimate error covariance matrix $\hat{\boldsymbol{P}}_k$, then the state estimate and covariance matrix can be propagated forward in time by means of

$$\bar{\boldsymbol{x}}_{k+1} = \boldsymbol{F}_k \hat{\boldsymbol{x}}_k \tag{3}$$

$$\bar{\boldsymbol{P}}_{k+1} = \boldsymbol{F}_k \hat{\boldsymbol{P}}_k \boldsymbol{F}_k^T + \boldsymbol{Q}_k \tag{4}$$

where $\bar{\boldsymbol{x}}_{k+1}$ and $\bar{\boldsymbol{P}}_{k+1}$ are the predicted state estimate and state estimate error covariance at time $k+1$. The predicted measurement follows from the state prediction as $\bar{\boldsymbol{y}}_k = \boldsymbol{H}_k \bar{\boldsymbol{x}}_k$.

B. Measurement Update

Given a measurement $\boldsymbol{y}_k$, the filter update step takes the form of a linear regression problem between the measured quantity and the predicted quantity. This form can be found by first defining the state prediction error at time k as $\boldsymbol{\delta}_k = \boldsymbol{x}_k - \bar{\boldsymbol{x}}_k$, where $\boldsymbol{x}_k$ is the true state at time k. It then follows that the state prediction can be written as $\bar{\boldsymbol{x}}_k = \boldsymbol{x}_k - \boldsymbol{\delta}_k$. The measurement update can then be written as the linear regression problem

$$\begin{Bmatrix} \boldsymbol{y}_k \\ \bar{\boldsymbol{x}}_k \end{Bmatrix} = \begin{bmatrix} \boldsymbol{H}_k \\ \boldsymbol{I} \end{bmatrix} \boldsymbol{x}_k + \begin{Bmatrix} \boldsymbol{w}_k \\ -\boldsymbol{\delta}_k \end{Bmatrix} \tag{5}$$

The regression problem can be further simplified by defining the quantities

$$\boldsymbol{S}_k = \begin{bmatrix} \boldsymbol{R}_k & \boldsymbol{0} \\ \boldsymbol{0} & \bar{\boldsymbol{P}}_k \end{bmatrix} \tag{6}$$

$$\boldsymbol{z}_k = \boldsymbol{S}_k^{-1/2} \begin{Bmatrix} \boldsymbol{y}_k \\ \bar{\boldsymbol{x}}_k \end{Bmatrix} \tag{7}$$

$$\boldsymbol{M}_k = \boldsymbol{S}_k^{-1/2} \begin{bmatrix} \boldsymbol{H}_k \\ \boldsymbol{I} \end{bmatrix} \tag{8}$$

$$\boldsymbol{\xi}_k = \boldsymbol{S}_k^{-1/2} \begin{Bmatrix} \boldsymbol{w}_k \\ -\boldsymbol{\delta}_k \end{Bmatrix} \tag{9}$$

leading to the result

$$\boldsymbol{z}_k = \boldsymbol{M}_k \boldsymbol{x}_k + \boldsymbol{\xi}_k \tag{10}$$

The Huber filter measurement update is found by minimizing the cost function

$$J(\hat{\boldsymbol{x}}_k) = \sum_{i=1}^{n} \rho(\zeta_i) \tag{11}$$

where ζ_i refers to the ith component of the residual vector, $\boldsymbol{\zeta} = \boldsymbol{G}_k \hat{\boldsymbol{x}}_k - \boldsymbol{z}_k$, and $\rho(\zeta_i)$ is known as the "score function." The solution of the minimization problem can be found from the derivative of the cost function, leading to the implicit equation

$$\sum_{i=1}^{n} \phi(\zeta_i) \frac{\partial \zeta_i}{\partial \boldsymbol{x}_k} = 0 \tag{12}$$

where $\phi(\zeta_i) = \rho'(\zeta_i)$. By defining $\psi(\zeta_i) = \phi(\zeta_i)/\zeta_i$, and the matrix $\boldsymbol{\Psi} = \text{diag}[\psi(\zeta_i)]$, then the implicit equation can be written in matrix form as

$$\boldsymbol{M}_k^T \boldsymbol{\Psi} (\boldsymbol{M}_k \boldsymbol{x}_k - \boldsymbol{z}_k) = \boldsymbol{0} \tag{13}$$

The solution of Eq. (13) can be determined by the iteratively reweighted algorithm,[16] given by[4]

$$\boldsymbol{x}_k^{(j+1)} = \left(\boldsymbol{M}_k^T \boldsymbol{\Psi}^{(j)} \boldsymbol{M}_k \right)^{-1} \boldsymbol{M}_k^T \boldsymbol{\Psi}^{(j)} \boldsymbol{y}_k \tag{14}$$

where the superscript (j) refers to the iteration index. The method can be initialized by using the least–squares solution, corresponding to $\boldsymbol{\Psi} = \boldsymbol{I}$. The converged value from the iterative procedure is taken as the state estimate, $\hat{\boldsymbol{x}}_k$. The iteratively reweighted technique converges for non–increasing ψ functions.[8]

In generalized maximum likelihood estimation, the function ρ can be chosen to yield an estimator $\hat{\boldsymbol{x}}$ with certain desirable properties. One desirable property of the solution of the generalized maximum likelihood technique is that of robustness with respect to deviations from the assumed underlying probability distribution. Huber[1] introduces a ρ function of the form

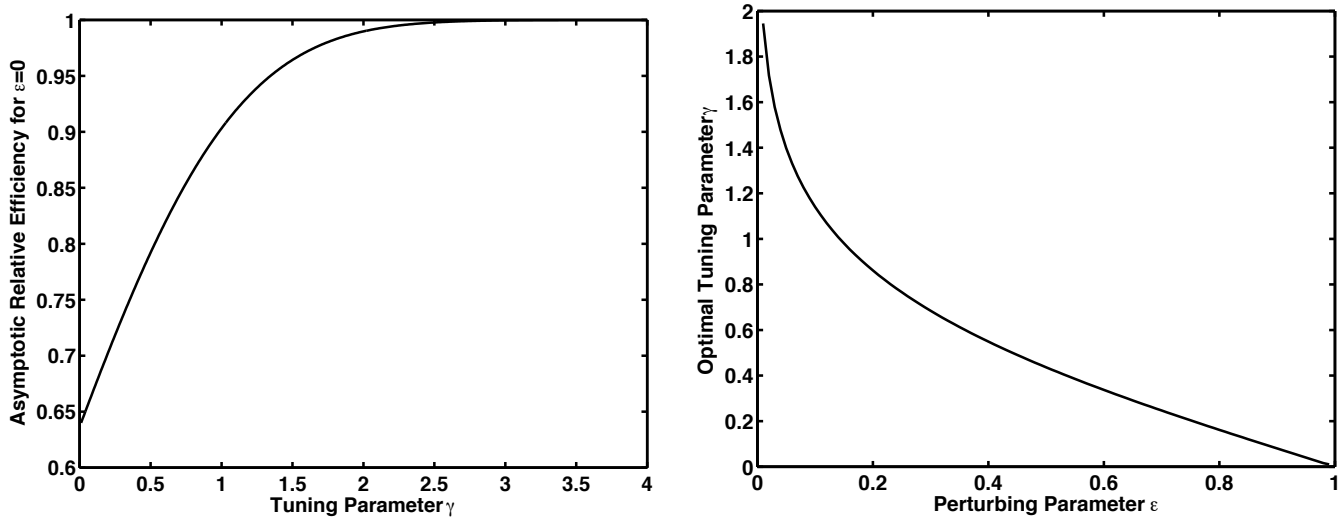
$$\rho(\zeta_i) = \begin{cases} \frac{1}{2}\zeta_i^2 & \text{for } |\zeta_i| < \gamma \\ \\ \gamma|\zeta_i| - \frac{1}{2}\gamma^2 & \text{for } |\zeta_i| \geq \gamma \end{cases} \tag{15}$$

where γ is a tuning parameter. This ρ function is a blend of the minimum ℓ_1 and ℓ_2 norm functions. Note that as $\gamma \to 0$, Eq. (15) approaches the ℓ_1 norm, which is equivalent to the median in the scalar case, and that as $\gamma \to \infty$, Eq. (15) approaches the ℓ_2 norm, which is equivalent to the mean in the scalar case.

Huber[1] shows that the ρ function given in Eq. (15) is asymptotically optimally robust in the ϵ neighborhood of the Gaussian distribution. Explicitly, if the measurement errors follow a perturbed Gaussian density function of the form

$$f(w) = \frac{1-\epsilon}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right) + \epsilon g(w) \tag{16}$$

where ϵ is a perturbing parameter and $g(w)$ is an unknown perturbing density function, then the estimation technique with ρ function given by Eq. (15) minimizes the maximum asymptotic estimation variance in the ϵ neighborhood of the Gaussian distribution. In addition, it can be shown that estimators of this form are asymptotically normal



(a) Huber Asymptotic Relative Efficiency for $\epsilon = 0$

(b) Optimal Tuning Parameter γ

Figure 1. Huber Gaussian Asymptotic Relative Efficiency and Optimal Tuning Parameter

and unbiased.[17] It should also be noted that the ψ function associated with the ρ function in Eq. (15) satisfies the properties discussed previously regarding the convergence of the iteratively reweighted least squares solution methodology.

Due to the particular structure of the matrix $\boldsymbol{M}_k$, the discrete time dynamic state estimation technique can be simplified considerably from the static state estimation technique by application of the matrix inversion lemma.[18] By first decomposing the $\boldsymbol{\Psi}$ matrix into two portions $\boldsymbol{\Psi}_x$ and $\boldsymbol{\Psi}_y$ corresponding to the state prediction and measurement prediction residuals so that

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_y & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_x \end{bmatrix} \tag{17}$$

and by defining the gain matrix

$$\boldsymbol{K}_k = \bar{\boldsymbol{P}}_k^{1/2} \boldsymbol{\Psi}_x^{-1} \bar{\boldsymbol{P}}_k^{1/2} \boldsymbol{H}_k^T \left(\boldsymbol{H}_k \bar{\boldsymbol{P}}_k^{1/2} \boldsymbol{\Psi}_x^{-1} \bar{\boldsymbol{P}}_k^{1/2} \boldsymbol{H}_k^T + \boldsymbol{R}_k^{1/2} \boldsymbol{\Psi}_y^{-1} \boldsymbol{R}_k^{1/2} \right)^{-1} \tag{18}$$

then the state estimate and covariance corrections can be written as

$$\hat{\boldsymbol{x}}_k^{(j+1)} = \bar{\boldsymbol{x}}_k + \boldsymbol{K}_k^{(j)} \left(\boldsymbol{y}_k - \boldsymbol{H}_k \bar{\boldsymbol{x}}_k \right) \tag{19}$$

$$\hat{\boldsymbol{P}}_k = \left(\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k \right) \bar{\boldsymbol{P}}_k^{1/2} \boldsymbol{\Psi}_x^{-1} \bar{\boldsymbol{P}}_k^{1/2} \tag{20}$$

Note that as the tuning parameter $\gamma \to \infty$, the matrix $\boldsymbol{\Psi} \to \mathbf{I}$ and the Huber recursive estimation technique reduces to the familiar Kalman filter solution.

C. Choice of the Tuning Parameter

Figure 1(a) shows the influence of the choice of the tuning parameter γ on the asymptotic relative efficiency of the Huber technique for the Gaussian case ($\epsilon = 0$) for a scalar estimation problem.[2] The efficiency can vary between one for $\gamma = \infty$ (the ℓ_2 case) and $2/\pi$ for $\gamma = 0$ (the ℓ_1 case). If the value of the parameter ϵ is known (even though the density function $g(w)$ of the perturbing distribution is unknown), then the best choice of γ has been determined by Huber.[1, 15] If $\epsilon = \epsilon_0$ is the known perturbing parameter, then the optimal choice of γ, denoted by $\gamma^\star$, is given by the implicit equation[1, 15]

$$\frac{1}{1-\epsilon_0} = \frac{1}{\gamma^\star} \sqrt{\frac{2}{\pi}} \exp\left[-\frac{(\gamma^\star)^2}{2} \right] + \operatorname{erf}\left(\frac{\gamma^\star}{\sqrt{2}} \right) \tag{21}$$

For a given value of ϵ, estimates computed using $\gamma = \gamma^\star$ maximize the asymptotic relative efficiency across the range of all possible values of γ.[1, 15] The optimal value of the tuning parameter $\gamma^\star$ is shown as a function of ϵ_0 in Fig. 1(b).

If the perturbing parameter ϵ is unknown, then the choice of γ is typically motivated by a desired variance at the model distribution. One common choice is $\gamma = 1.345$; for this value the Huber filter exhibits estimation error variances that are 5% larger than that of the least squares method when the measurement error distributions are truly Gaussian.[19] Generally any value of γ between 1.0 and 2.0 is suggested.[1]

III. Adaptive Tuning for Unknown Noise Statistics

Although the Huber technique has been shown to be beneficial in the presence of contaminated Gaussian probability distributions, the approach can still be improved upon by including a method by which the covariance of the main Gaussian distribution can be estimated along with the state and state error covariance matrix. A technique such as this allows the filter to adaptively tune the gain matrix to suite either slow changes in the error statistics, or statistics that are not well known.

A. Myers–Tapley Method

An intuitive approach to adaptive state estimation is proposed by Myers and Tapley.[7] In this approach, the measurement residual sequence is mined to produce estimates of the measurement noise statistics, and state prediction residuals are mined to compute estimates of the process noise statistics. The estimators make use of a sliding buffer of N stored measurement noise and process noise residuals to compute the noise statistics, assuming the noise is stationary over the the duration of the stored noise samples. In batch form, the estimator for the measurement noise covariance can be derived by first defining the empirical covariance matrix

$$\boldsymbol{C}_\zeta = \frac{1}{N-1}\sum_{j=1}^{N}\left(\boldsymbol{\zeta}_j - \bar{\boldsymbol{\zeta}}\right)\left(\boldsymbol{\zeta}_j - \bar{\boldsymbol{\zeta}}\right)^T \tag{22}$$

where $\boldsymbol{\zeta}_j$ is the jth stored measurement residual, and $\bar{\boldsymbol{\zeta}}$ is the sample mean of the residuals, $\bar{\boldsymbol{\zeta}} = \sum_{j=1}^{N}\boldsymbol{\zeta}_j/N$.

The expected value of $\boldsymbol{C}_\zeta$ is[7]

$$\mathrm{E}\left[\boldsymbol{C}_\zeta\right] = \boldsymbol{R} + \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{H}_j\bar{\boldsymbol{P}}_j\boldsymbol{H}_j^T \tag{23}$$

By substituting Eq. (22) into Eq. (23), an estimate for the measurement noise covariance matrix is

$$\hat{\boldsymbol{R}} = \frac{1}{N-1}\sum_{j=1}^{N}\left[\left(\boldsymbol{\zeta}_j - \bar{\boldsymbol{\zeta}}\right)\left(\boldsymbol{\zeta}_j - \bar{\boldsymbol{\zeta}}\right)^T - \left(\frac{N-1}{N}\right)\boldsymbol{H}_j\bar{\boldsymbol{P}}_j\boldsymbol{H}_j^T\right] \tag{24}$$

In order to form estimates for the process noise statistics, the process noise sample is defined as $\boldsymbol{\lambda}_j = \hat{\boldsymbol{x}}_j - \bar{\boldsymbol{x}}_j$. Then, the empirical covariance matrix for $\boldsymbol{\lambda}$ is

$$\boldsymbol{C}_\lambda = \frac{1}{N-1}\sum_{j=1}^{N}\left(\boldsymbol{\lambda}_j - \bar{\boldsymbol{\lambda}}\right)\left(\boldsymbol{\lambda}_j - \bar{\boldsymbol{\lambda}}\right)^T \tag{25}$$

where $\bar{\boldsymbol{\lambda}} = \sum_{j=1}^{N}\boldsymbol{\lambda}_j/N$.

The expected value of $\boldsymbol{C}_\lambda$ is[20] $\mathrm{E}\left[\boldsymbol{C}_\lambda\right] = \boldsymbol{K}_k\boldsymbol{H}_k\bar{\boldsymbol{P}}_k = \boldsymbol{F}_k\hat{\boldsymbol{P}}_{k-1}\boldsymbol{F}_k^T - \hat{\boldsymbol{P}}_k$. Therefore, an estimator for the process noise covariance matrix is

$$\hat{\boldsymbol{Q}} = \frac{1}{N-1}\sum_{j=1}^{N}\left[\left(\boldsymbol{\lambda}_j - \bar{\boldsymbol{\lambda}}\right)\left(\boldsymbol{\lambda}_j - \bar{\boldsymbol{\lambda}}\right)^T - \left(\frac{N-1}{N}\right)\left(\boldsymbol{F}_j\hat{\boldsymbol{P}}_{j-1}\boldsymbol{F}_j^T - \hat{\boldsymbol{P}}_j\right)\right] \tag{26}$$

An adaptive filter can function by using some initial guess of the measurement noise and process noise matrices, storing the residuals for the first N frames, and then updating the covariance estimates based on Eqs.(24) and (26) at each subsequent frame.

It is important to note that the Myers–Tapley method for adaptively estimating the measurement noise and process noise covariance matrices make use of the sample mean and covariance of the stored residuals, which are non–robust estimators. This lack of robustness implies that the performance Myers–Tapley adaptive method can degrade in the presence of non–Gaussianity. Therefore it is of interest to develop a modification of the Myers–Tapley approach that is robust with respect to non–Gaussian distributions, which is the subject of the following section.

B. Modified Myers–Tapley Method

This section discusses a robust modification of the Myers–Tapley adaptive filter in order to protect against the possible presence of non–Gaussian distributions in the measurement and/or process noise. The main issue with the Myers–Tapley method is the use of the sample covariance and the sample mean in forming the noise covariance estimates from the stored residuals. Fortunately, the technique can be modified to use a robust form of the covariance estimate in computing the measurement and process noise statistics. The proposed technique for computing the robust estimates of the mean and covariance from the batch of stored residuals is based on a weighted form of the mean and covariance, where the weights depend on statistics formed in the course of an outlier identification scheme.

In this section, two outlier identification methods are introduced. The first method is a classical technique that is based on the weighted Euclidean norm of the separation between a possible outlier and the sample mean, known as Mahalanobis Distances.[8] The weighting in this method is based on the sample covariance matrix. The second method, known as Projection Statistics, is a robust approach in which the sample mean and covariance are replaced by the sample median and the median absolute deviation, respectively.

1. The Mahalanobis Distances

Given a cloud of m points in n dimensions represented by the vectors $\boldsymbol{h}_i$ for $i = 1, \cdots, m$, the Mahalanobis distances are defined as

$$\mathcal{M}_i = \sqrt{\left(\boldsymbol{h}_i - \bar{\boldsymbol{h}}\right)^T \boldsymbol{C}^{-1} \left(\boldsymbol{h}_i - \bar{\boldsymbol{h}}\right)} \tag{27}$$

where $\bar{\boldsymbol{h}}$ is the sample mean and $\boldsymbol{C}$ is the sample covariance matrix, given by the equations

$$\bar{\boldsymbol{h}} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{h}_i \tag{28}$$

$$\boldsymbol{C} = \frac{1}{m-1} \sum_{i=1}^{m} \left(\boldsymbol{h}_i - \bar{\boldsymbol{h}}\right) \left(\boldsymbol{h}_i - \bar{\boldsymbol{h}}\right)^T \tag{29}$$

respectively.

The Mahalanobis distances can also be expressed as the solution to a maximization problem of the form[9]

$$\mathcal{M}_i = \max_{\|\boldsymbol{v}\|=1} \left[\frac{\left\| \boldsymbol{h}_i^T \boldsymbol{v} - \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{h}_j^T \boldsymbol{v} \right\|}{\sqrt{\frac{1}{m-1} \sum_{k=1}^{m} \left(\boldsymbol{h}_k^T \boldsymbol{v} - \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{h}_j^T \boldsymbol{v} \right)^2}} \right] \tag{30}$$

The Mahalanobis Distances represent the surface of an n-dimensional ellipsoid centered at the sample mean. The square of the Mahalanobis distances follow a χ^2 distribution with n degrees of freedom, assuming that the input data are Gaussian. Therefore, an outlier identification method is to consider all points satisfying

$$\mathcal{M}_i > \sqrt{\chi^2_{n,\alpha}} \tag{31}$$

to be outliers, where α is the probability that a value falls inside the ellipse (for example, $\alpha = 0.95$).

While the Mahalanobis distances are simple to conceptualize and easy to compute, the method suffers from sensitivity to clusters of outliers. This sensitivity is due to the so–called *masking effect*, in which it is possible to find groups of points with nonzero errors that can sum to produce very small residuals. The masking effect in the Mahalanobis distances is related to the sensitivity of the sample mean and covariance, which are not robust estimators. In the case of clustered outliers the sample mean can be pulled toward their direction and away from the main cluster of data, leading to an increase in the size of the sample covariance. In effect, these sensitivities serve to hide or mask the cluster of outliers since their associated Mahalanobis distances are no larger than those related to the main group of data. See Refs. 10 and 11 for further information on the masking effect and sensitivity of the Mahalanobis distances.

2. Projection Statistics

A robust approach to the problem of outlier identification is to simply replace the sample mean and covariance in the equation for the Mahalanobis distances, Eq. 30, with the sample median and the median absolute deviation from the median.[9] These estimators of location and scale are known to be robust with respect to outliers and therefore

one can expect that the computation of a Mahalanobis–like quantity based on these parameters is also robust with respect to outliers. These new quantities are known as *Projection Statistics*, and are defined as the solution to the maximization problem

$$\mathcal{P}_i = \max_{\|\boldsymbol{v}\|=1} \left[\frac{\left\| \boldsymbol{h}_i^T \boldsymbol{v} - \text{median}\left(\boldsymbol{h}_j^T \boldsymbol{v} \right) \right\|}{c \cdot \text{median}\left(\left\| \boldsymbol{h}_k^T \boldsymbol{v} - \text{median}\left(\boldsymbol{h}_j^T \boldsymbol{v} \right) \right\| \right)} \right] \tag{32}$$

where $c = 1.4826$ in the denominator is a correction factor to ensure unbiasedness.[22] The maximization problem can be approximated by considering only the directions that correspond to the unit vectors of the individual data points relative to the median of the point cloud. An algorithm for computing the approximate projection statistics is given in Ref. 21.

3. Robust Covariance Matrix Estimation

A robust form of the covariance matrix of the point cloud by using a weighting scheme that depends on the statistics $\mathcal{M}_i$ or $\mathcal{P}_i$. Specifically, the proposed mean and covariance formulation is

$$\bar{\boldsymbol{h}}_r = \left[\sum_{i=1}^{m} w_i \right]^{-1} \cdot \left[\sum_{i=1}^{m} w_i \boldsymbol{h}_i \right] \tag{33}$$

$$\mathbf{C}_r = \left[\sum_{i=1}^{m} w_i - 1 \right]^{-1} \cdot \left[\sum_{i=1}^{m} \left(w_i \boldsymbol{h}_i - \bar{\boldsymbol{h}}_r \right) \left(w_i \boldsymbol{h}_i - \bar{\boldsymbol{h}}_r \right)^T \right] \tag{34}$$

where w_i are weights computed from the statistics by means of

$$w_i = \min \left[\; 1, \quad \left(\chi^2_{n,\alpha} / \mathcal{M}_i^2 \right) \; \right] \tag{35}$$

or

$$w_i = \min \left[\; 1, \quad \left(\chi^2_{n,\alpha} / \mathcal{P}_i^2 \right) \; \right] \tag{36}$$

for some specified probability α.

4. Modified Myers–Tapley Method Using Projection Statistics

The Myers–Tapley adaptive tuning method can be modified in order to account for non–Gaussianity by means of using the robust covariance estimates based on the projection statistics of the stored residuals. In particular, the measurement and process noise covariance estimates can be written as

$$\hat{\boldsymbol{R}}^{1/2} \bar{\boldsymbol{\Psi}}_y^{-1} \hat{\boldsymbol{R}}^{1/2} = \left[\sum_{i=1}^{N} w_{\zeta_i} - 1 \right]^{-1} \cdot \left[\sum_{i=1}^{N} \left(w_{\zeta_i} \boldsymbol{\zeta}_i - \bar{\boldsymbol{\zeta}}_r \right) \left(w_{\zeta_i} \boldsymbol{\zeta}_i - \bar{\boldsymbol{\zeta}}_r \right)^T \right] - \text{median}_{j \in N} \left(\boldsymbol{H}_j \bar{\boldsymbol{P}}_j^{1/2} \bar{\boldsymbol{\Psi}}_x^{-1} \bar{\boldsymbol{P}}_j^{1/2} \boldsymbol{H}_j^T \right) \tag{37}$$

$$\hat{\boldsymbol{Q}} = \left[\sum_{i=1}^{N} w_{\lambda_i} - 1 \right]^{-1} \cdot \left[\sum_{i=1}^{N} \left(w_{\lambda_i} \boldsymbol{\lambda}_i - \bar{\boldsymbol{\lambda}}_r \right) \left(w_{\lambda_i} \boldsymbol{\lambda}_i - \bar{\boldsymbol{\lambda}}_r \right)^T \right] - \text{median}_{j \in N} \left(\boldsymbol{F}_j \hat{\boldsymbol{P}}_{j-1} \boldsymbol{F}_j^T - \hat{\boldsymbol{P}}_j \right) \tag{38}$$

$$\tag{39}$$

where w_{ζ_i} and w_{λ_i} are the weights based on the Mahalanobis distances or projection statistics of the measurement and process noise residuals, respectively, and

$$\bar{\boldsymbol{\zeta}}_r = \left[\sum_{i=1}^{N} w_{\zeta_i} \right]^{-1} \cdot \left[\sum_{i=1}^{N} w_{\zeta_i} \zeta_i \right] \tag{40}$$

$$\bar{\boldsymbol{\lambda}}_r = \left[\sum_{i=1}^{N} w_{\lambda_i} \right]^{-1} \cdot \left[\sum_{i=1}^{N} w_{\lambda_i} \lambda_i \right] \tag{41}$$

In summary, the proposed modified Myers–Tapley method makes use of a technique for outlier identification and weighting for the purpose of forming the estimates of the measurement and process noise covariance matrices.

5. Estimation of the Contamination Parameter

The modified Myers–Tapley approach discussed in the previous section can also offer a crude scheme for estimating the contamination parameter ϵ by using the weighting parameters relating to the stored residual data. In particular, a crude estimate of the contamination parameter is

$$\hat{\epsilon}_k = 1 - \frac{1}{N}\sum_{i=1}^{N} w_{\zeta_i} \tag{42}$$

At each frame where the measurement and process noise covariances are computed, the contamination parameter can be estimated directly from the weighting parameters. Then, the optimal tuning parameter $\gamma^\star$ can be calculated from Eq. (21), which is then used within the Huber filter at each measurement update. It is expected that the estimated contamination parameter is biased, since in cases of large contamination there may be some portion of errors drawn from the contaminating distribution that appear to be drawn from the nominal distribution, in other words false–negatives or the so-called Type II errors in detection theory. Likewise in cases of small contamination there may be some nonzero quantity of data that appear as outliers when they are in fact perfectly valid, in other words the Type I error in detection theory.

The bias of the contamination parameter is not necessarily problematic, so long as upper and lower bounds are set on the value of the tuning parameter γ used in the Huber measurement update. Clearly, the Huber technique is by nature a sub-optimal filter, since the purpose is to find a filter that is consistent across a range of distributions but may not necessarily be optimal at any one in particular. The performance of the technique can improve by having good estimates of the contamination parameter, but so long as the tuning parameter is bounded by some reasonable value then the robustness properties of the estimator are not be compromised the bias.

6. Fading Memory Filter

In the course of computing the estimates of the measurement and process noise covariance matrices, as well as the contamination parameter estimates, it is useful to introduction a "forgetting" factor, k_f, in order to smooth the estimate histories. The filter can be implemented as

$$\tilde{\boldsymbol{R}}_k = k_f \tilde{\boldsymbol{R}}_{k-1} + (1-k_f)\,\hat{\boldsymbol{R}}_k \tag{43}$$

$$\tilde{\boldsymbol{Q}}_k = k_f \tilde{\boldsymbol{Q}}_{k-1} + (1-k_f)\,\hat{\boldsymbol{Q}}_k \tag{44}$$

$$\tilde{\epsilon}_k = k_f \tilde{\epsilon}_{k-1} + (1-k_f)\,\hat{\epsilon}_k \tag{45}$$

In this approach, the estimates based on the current set of stored residuals is averaged with the previous estimate, with k_f as a weighting parameter.

IV. Application to Spacecraft Attitude Estimation

Spacecraft must make use of various navigation sensors in order to estimate the attitude of the vehicle at any instant in time. One such sensor is the gyroscope, which senses the inertial angular velocity of the spacecraft. The modelling of such a system can be quite complicated, and in many cases approximations are used for in–flight software development and ground–based simulations. One such approximation in common use for gyroscope systems is known as Farrenkopf's model,[23] which considers the measured angular velocity to be the true angular velocity, corrupted by an additive bias and a Gaussian white noise term. The bias dynamics are in turn considered to be driven by a Gaussian white noise process. This model is commonly used with the assumption that the bias term can account for the net effect of many other systematic error sources such as scale factor errors, non-orthogonality, misalignment, and others.

It is well known that the numerical integration of gyroscope measurements to determine the vehicle orientation time history leads to a drift of the estimates away from the truth. This drift is due to both the effect of systematic error and numerical error during the integration. Therefore, it is required to consider data from additional sensors, such as a star tracker, that is incorporated into the attitude estimation procedure in some optimal or nearly optimal fashion. This section describes the application of the adaptive Huber–based filters to the problem of estimating the attitude of a spacecraft by processing gyroscope and star tracker data. The attitude estimation problem is an interesting and relevant example of the application of robust and adaptive filtering techniques, since the typical sensor systems can exhibit non–Gaussian characteristics[24, 25] as well as uncertain noise statistics.[26]

The point should be made that the sensor referred to as the *star tracker* throughout the rest of this paper is in fact more general, and could just as well refer to any system of sensors that can provide an attitude measurement (or

pseudo–measurement) in the form of a quaternion. Therefore the methodology described in the following sections is relatively general in nature. A short review of attitude kinematics, gyroscope and star tracker sensor models and observation equations is provided in the following sections. For additional details on attitude dynamics and estimation, see Refs. 18 and 27.

A. Attitude Kinematics

The attitude dynamics of spacecraft is well understood and can be represented in many forms. Perhaps the most attractive form for attitude estimation is the quaternion representation, which is the smallest globally non-singular attitude parameterization. The quaternion can be expressed in terms of the Euler axis, $\boldsymbol{e} = [e_1 \ e_2 \ e_3]^T$, and the Euler angle, θ, of the transformation between a reference frame fixed in the body and the inertial frame. The quaternion elements are defined as $q_1 = e_1 \sin(\theta/2), q_2 = e_2 \sin(\theta/2), q_3 = e_3 \sin(\theta/2), q_4 = \cos(\theta/2)$. By defining $\boldsymbol{q} = \begin{bmatrix} q_1 & q_2 & q_3 & q_4 \end{bmatrix}^T$ it can be seen that the quaternion representation must satisfy the constraint equation $\boldsymbol{q}^T\boldsymbol{q} = 1$. Quaternions obey the multiplication rules

$$\boldsymbol{q}' \otimes \boldsymbol{q} = \boldsymbol{\Upsilon}(\boldsymbol{q}')\,\boldsymbol{q} = \boldsymbol{\Xi}(\boldsymbol{q})\,\boldsymbol{q}' \tag{46}$$

where

$$\boldsymbol{\Upsilon}(\mathbf{q}') = \begin{bmatrix} q_4' & q_3' & -q_2' & q_1' \\ -q_3' & q_4' & q_1' & q_2' \\ q_2' & -q_1' & q_4' & q_3' \\ -q_1' & -q_2' & -q_3' & q_4' \end{bmatrix}, \quad \boldsymbol{\Xi}(\mathbf{q}) = \begin{bmatrix} q_4 & -q_3 & q_2 & q_1 \\ q_3 & q_4 & -q_3 & q_2 \\ -q_2 & q_1 & q_4 & q_3 \\ -q_1 & -q_2 & -q_3 & q_4 \end{bmatrix} \tag{47}$$

and the inversion rule $\boldsymbol{q}^{-1} = \begin{bmatrix} -q_1 & -q_2 & -q_3 & q_4 \end{bmatrix}^T$.

The quaternion kinematic equations can be written as[27]

$$\dot{\boldsymbol{q}} = \frac{1}{2}\boldsymbol{\Omega}(\boldsymbol{\omega})\,\boldsymbol{q} \tag{48}$$

where

$$\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} 0 & -\omega_1 & -\omega_2 & -\omega_3 \\ \omega_1 & 0 & -\omega_3 & \omega_2 \\ \omega_2 & \omega_3 & 0 & -\omega_1 \\ \omega_3 & -\omega_2 & \omega_1 & 0 \end{bmatrix} \tag{49}$$

and $\boldsymbol{\omega}$ is the angular velocity of the body frame with respect to the inertial frame.

B. Gyroscope Model

The gyroscope system can be represented mathematically by using Farrenkopf's model. In this model, the sensed angular velocity if expressed as the true angular velocity with an additive bias and white noise. The bias term is itself a slowly varying parameter driven by white noise. The model can be expressed as

$$\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega} + \boldsymbol{\beta} + \boldsymbol{\eta}_\omega \tag{50}$$

$$\dot{\boldsymbol{\beta}} = \boldsymbol{\eta}_\beta \tag{51}$$

where $\tilde{\boldsymbol{\omega}}$ is the sensed inertial angular velocity, $\boldsymbol{\omega}$ is the true inertial angular velocity, $\boldsymbol{\beta}$ is the measurement bias, and $\boldsymbol{\eta}_\omega$ and $\boldsymbol{\eta}_\beta$ are unbiased and uncorrelated random vectors with variances given by σ_ω^2 and σ_β^2, respectively. Discrete–time simulated gyroscope measurements can be generated according to this model by use of the equations[18]

$$\tilde{\boldsymbol{\omega}}_{k+1} = \boldsymbol{\omega}_k + \frac{1}{2}\left(\boldsymbol{\beta}_{k+1} + \boldsymbol{\beta}_k\right) + \left(\frac{\sigma_\omega^2}{\Delta t} + \frac{1}{12}\sigma_\beta^2 \Delta t\right)^{1/2} \boldsymbol{n}_\omega \tag{52}$$

$$\boldsymbol{\beta}_{k+1} = \boldsymbol{\beta}_k + \sigma_\beta (\Delta t)^{1/2}\, \boldsymbol{n}_\beta \tag{53}$$

where k refers to the time increment, $\Delta t = t_{k+1} - t_k$ is the sampling interval, and $\boldsymbol{n}_\omega$ and $\boldsymbol{n}_\beta$ are unbiased, uncorrelated, unit–variance random vectors.

C. Star Tracker Model

It is assumed that a star tracker or some other generic attitude sensor is available to provide corrections to the attitude estimates formed by direct numerical integration of the angular velocity measurements, which are subject to error buildup due to integrating errors in the estimated bias and the random noise. The star tracker is assumed to output an estimated quaternion that relates the orientation of the body to the inertial frame. The quaternion estimates are assumed to be unbiased but with a superimposed random measurement noise. The output from such a sensor can be expressed as $\boldsymbol{q}_m = \delta\boldsymbol{q}\left(\delta\boldsymbol{\alpha}\right) \otimes \boldsymbol{q}$ where $\boldsymbol{q}$ is the quaternion representing the true orientation, $\boldsymbol{q}_m$ is the measured quaternion, and $\delta\boldsymbol{q}\left(\delta\boldsymbol{\alpha}\right)$ is an error quaternion parameterized by a random angular error $\delta\boldsymbol{\alpha}$.

D. Attitude Filter Implementation

The filtering of the attitude quaternion is complicated by the unit norm constraint. One standard approach for handling the constraint is to use a multiplicative method for the measurement update. In this approach, the attitude quaternion is replaced by an error quaternion, $\delta\boldsymbol{q}$, which allows for the processing of the error quaternion as a three component vector of small angles, denoted by $\delta\boldsymbol{\alpha}$. In this approach, the quaternion estimate is propagated to the star tracker measurement at time k by means of Eq. (57). Then an error quaternion relating the star tracker measurement quaternion to the predicted quaternion based on propagation from the previous measurement is defined as $\delta\bar{\boldsymbol{q}}_k\left(\delta\bar{\boldsymbol{\alpha}}_k\right) = \boldsymbol{q}_{m_k} \otimes \bar{\boldsymbol{q}}_k^{-1}$. The first three components of this error quaternion are processed as independent small angles with measurement sensitivity matrix $\boldsymbol{H} = \left[\begin{array}{cc} \boldsymbol{I} & \boldsymbol{0} \end{array}\right]$.

The small angle error and gyroscope bias estimate error dynamics are given by

$$\delta\dot{\boldsymbol{\alpha}} = -\hat{\boldsymbol{\omega}}^{\times}\delta\boldsymbol{\alpha} - \delta\boldsymbol{\beta} + \boldsymbol{\eta}_{\omega} \tag{54}$$

$$\delta\dot{\boldsymbol{\beta}} = \boldsymbol{\eta}_{\beta} \tag{55}$$

where

$$\hat{\boldsymbol{\omega}}^{\times} = \left[\begin{array}{ccc} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{array}\right] \tag{56}$$

By setting $\boldsymbol{x} = \left[\begin{array}{cc} \delta\boldsymbol{\alpha} & \delta\boldsymbol{\beta} \end{array}\right]^T$, then these differential equations are in an appropriate form for filter implementation corresponding to Eq. (1). The quaternion and gyroscope bias estimates following the measurement update are given by $\hat{\boldsymbol{q}}_k = \delta\boldsymbol{q}\left(\delta\hat{\boldsymbol{\alpha}}_k\right) \otimes \bar{\boldsymbol{q}}_k$ and $\hat{\boldsymbol{\beta}}_k = \bar{\boldsymbol{\beta}}_k + \delta\hat{\boldsymbol{\beta}}_k$.

State propagation in between star tracker updates are based on assuming the angular velocity is constant over each gyroscope sample. First, the estimated true angular velocity of the spacecraft at time increment k, denoted by $\hat{\boldsymbol{\omega}}$, can be written as $\hat{\boldsymbol{\omega}}_k = \tilde{\boldsymbol{\omega}}_k - \hat{\boldsymbol{\beta}}_k$. where $\hat{\boldsymbol{\beta}}_k$ is the estimated gyroscope measurement bias at time increment k. If the gyroscope sampling rate is high, then the angular velocity can reasonably be assumed to be constant over the sampling interval. In this case, the quaternion kinematic equations become constant coefficient ordinary differential equations and so the solution can be determined from the matrix exponential. Explicitly, the solution for the predicted quaternion at time $k+1$ can be written as[28]

$$\bar{\boldsymbol{q}}_{k+1} = \left[\mathbf{I}\cos\left(\hat{\omega}_k\Delta t\right) + \frac{\sin\left(\hat{\omega}_k\Delta t/2\right)}{\hat{\omega}_k}\boldsymbol{\Omega}\left(\hat{\boldsymbol{\omega}}_k\right)\right]\hat{\boldsymbol{q}}_k \tag{57}$$

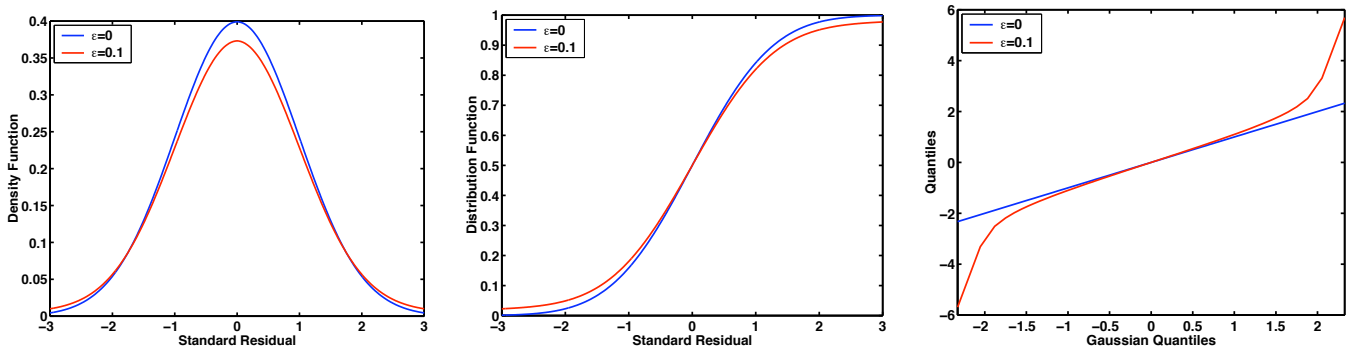
where $\hat{\boldsymbol{q}}_k$ is the estimated quaternion at time k, $\bar{\boldsymbol{q}}_{k+1}$ is the predicted quaternion at time $k+1$, and $\hat{\omega}_k = \|\hat{\boldsymbol{\omega}}_k\|$ is the assumed constant angular velocity between time k and $k+1$.

V. Example Problem

This section describes the results of the application of the filtering techniques discussed in previous sections to the spacecraft attitude estimation problem using gyroscopes and star trackers. The standard Kalman filter, Myers–Tapley adaptive Kalman filter, Huber filter, and the proposed modified Myers–Tapley adaptive Huber filter are applied to the attitude estimation problem. The example includes both Gaussian and non–Gaussian noise densities with unknown statistics.

In this example, the true spacecraft angular velocity is $\boldsymbol{\omega} = \left[\begin{array}{ccc} \omega_0 & \omega_0 & \omega_0 \end{array}\right]^T$, for $\omega_0 = 10^{-3}$ rad/s. The angular velocity is sampled at a rate of 10 Hz, with star tracker updates at a rate of 1 Hz. Noise samples for the gyroscope and star tracker measurements are drawn from a mixture model with a probability density function of the form

$$f(w) = \frac{1-\epsilon}{\sqrt{2\pi}}\exp\left(-\frac{w^2}{2}\right) + \frac{\epsilon}{2b}\exp\left(-\frac{|w|}{b}\right) \tag{58}$$



(a) Density Functions (b) Distribution Functions (c) Quantile–Quantile Plot

Figure 2. Comparison of Probability Density and Distribution Functions for $\epsilon = 0$ and $\epsilon = 0.1$

where $b = 5/\sqrt{2}$.

Note that this density is a mixture of a nominal Gaussian with a Laplacian contaminating density. A comparison of the probability density and distribution functions for the cases $\epsilon = 0$ (no contamination) and $\epsilon = 0.1$ (with 10 % contamination) is shown in Fig. 2. A comparison of pure Gaussian with the contaminated Gaussian densities is shown in Fig. 2(a), the distribution functions are shown in Fig. 2(b), and a quantile–quantile plot (or QQ plot) in Fig. 2(c). From the density function and distribution function plots it is not entirely obvious that the contaminating density is in fact non-Gaussian; it appears to have Gaussian characteristics but with a larger variance. The differences between the two distributions is much more apparent in the QQ plot in Fig. 2(c). A QQ plot is a plot of the scaled ordered data against the quantiles of a comparison distribution. Since the data are scaled prior to creating the plot, the result is independent of the variance, rendering only the shape of the distribution to be important. A linear result with unit slope indicates that the data follows the same distribution as the comparison distribution (the Gaussian, in this case). The QQ plot of the contaminated Gaussian indicates that the data follows a roughly Gaussian distribution in the middle ranges but with a greater thickness in the extremes. This result indicates that the contaminated distribution has a higher probability of generating extreme points.

The sensor uncertainty specifications are $\sigma_\omega^2 = 2.25 \cdot 10^{-13}$ rad^2/s, $\sigma_\beta^2 = 2.25 \cdot 10^{-19}$ rad^2/s^3, and $\sigma_s^2 = 3.81 \cdot 10^{-8}$ rad^2. It is assumed that the known sensor uncertainties differ from the true sensor uncertainties by a factor of 50% of the standard deviation. In addition, the star tracker measurement errors are initially assumed to be uncorrelated in each axis whereas the true errors are correlated with a correlation coefficient of 0.5 in each axis.

The Huber filter and adaptive Huber filter use an initial value for the tuning parameter of $\gamma = 1.5$. The tuning parameter is specified to be bounded such that $1 \leq \gamma \leq 2$ within the adaptive Huber filter. The adaptive filtering techniques are implemented in such a way that they begin estimating the noise statistics after 25 samples and store up to a specified maximum threshold of observations. This maximum threshold is varied in later sections to show the improvements that can be achieved through larger samples.

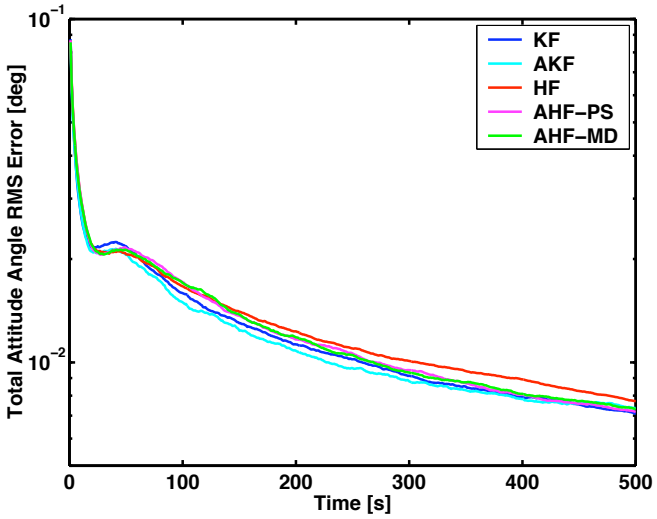
Finally, the adaptive filtering techniques use a value $k_f = 0.9$ to smooth the covariance and contamination parameters estimates.

A. Gaussian Simulation

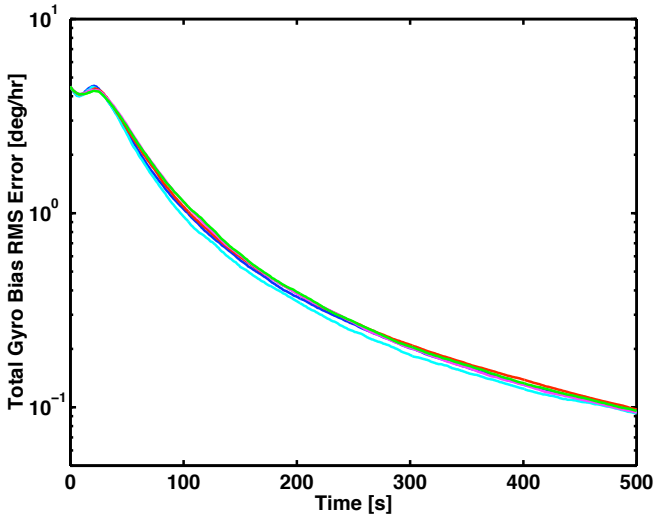
This section describes the results of a 2000 case Monte-Carlo simulation involving pure Gaussian random errors ($\epsilon = 0$) with the specified noise variances discussed in the previous section. The adaptive filtering techniques make use of a maximum buffer size of 100 observations in forming the estimates of the process and measurement noise covariance matrices.

The results in the form of root mean square (RMS) errors are shown in Fig. 3. The total attitude angle RMS error is shown in Fig. 3(a), the vector norm of the gyroscope bias RMS error is shown in Fig. 3(b). The error in the RMS error prediction of the total attitude angle is shown in Fig. 3(c), and that of the gyroscope bias is shown in Fig. 3(d). The (1,1) component of the star tracker variance estimate RMS error is shown in Fig. 3(e), and the RMS error of the contamination parameter estimate is shown in Fig. 3(f). In each case, the Kalman filter results are shown with the blue curve, the adaptive Kalman filter results are shown with the cyan curve, and the Huber filter results are shown with the red curve. For the adaptive Huber filter cases, the filter based on the projection statistics is shown with the magenta curve while that based on the Mahalanobis distances is shown with the green curve.

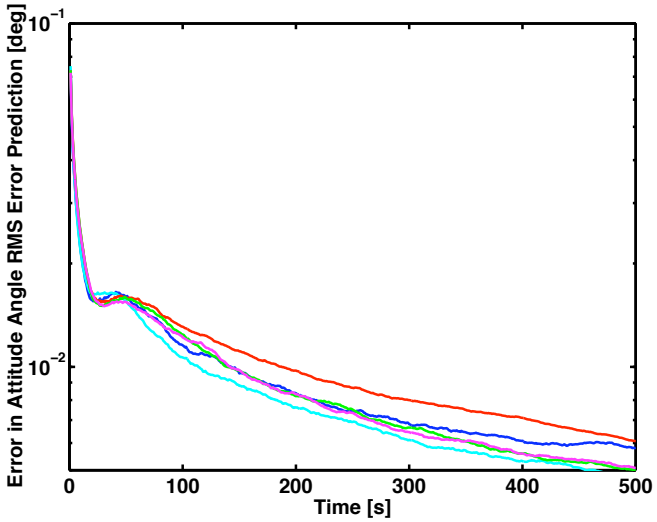
In this case, the attitude angle errors are not significantly different between the various techniques even though



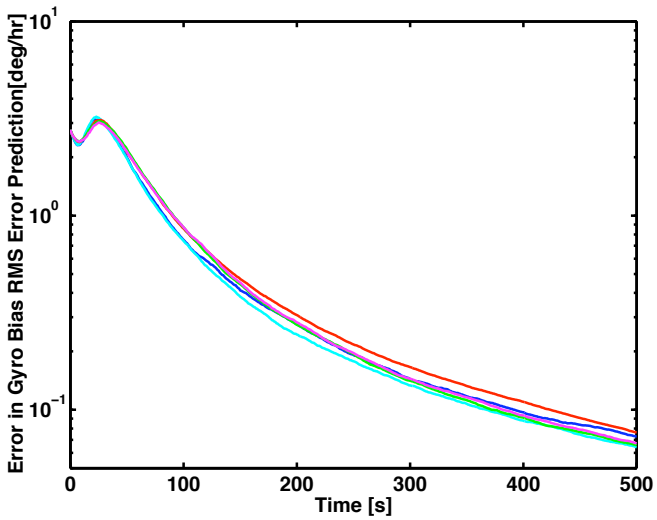
(a) Attitude Angle RMS Error



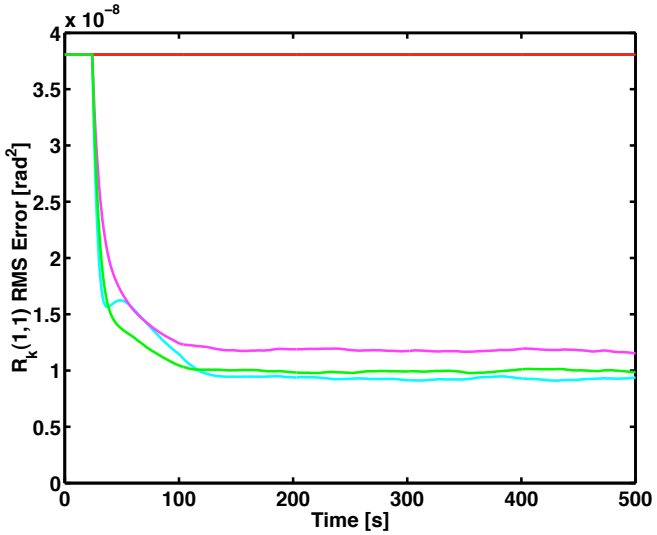
(b) Gyro Bias RMS Error



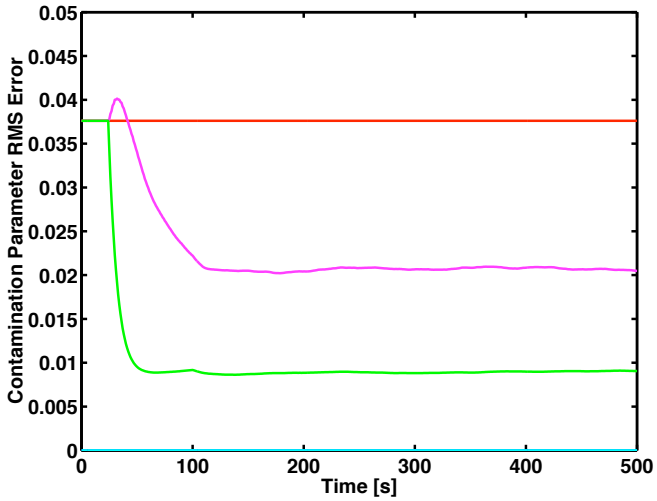
(c) Error in Attitude Angle RMS Error Prediction



(d) Error in Gyro Bias RMS Error Prediction



(e) $\tilde{\boldsymbol{R}}_k(1,1)$ RMS Error



(f) Contamination Parameter RMS Error

Figure 3. Gaussian Simulation RMS Errors

the measurement statistics assumed by the filters are erroneous. The Kalman and adaptive Kalman filter provide similar performance, slightly superior to the Huber approaches, which is to be expected since the errors are perfectly Gaussian. The adaptive Huber filters show slightly better performance than the non–adaptive case. The adaptive Huber filter based on the Mahalanobis distances performs better than that based on the projection statistics, which is to be expected since the sample mean and covariance used in computing the Mahalanobis distances are maximum likelihood estimates for perfectly Gaussian problems. The same trends and relative performance between filters is shown in the gyroscope bias RMS error.

The comparison of the error predictions in Figs. 3(c) and (d) show that the adaptive Kalman filter outperforms the other filters for both the attitude error prediction and rate bias error prediction. The predicted error estimates of the Huber–based adaptive filters are similar, and superior to the non–adaptive Huber and Kalman filters. The improved estimates of the attitude and bias errors for the adaptive filters is to be expected since these filters attain an improved knowledge of the measurement and process noise covariances. Naturally, the adaptive Kalman filter is superior since this problem is Gaussian.

The $\tilde{\boldsymbol{R}}_k(1,1)$ RMS error results in Fig. 3(e) show that the adaptive filters are able to reduce the error in the assumed measurement statistics. The adaptive Kalman filter shows superior performance than the adaptive Huber filter for this case, which is expected since the errors are purely Gaussian in nature. For the same reason, the adaptive Huber filter based on the Mahalanobis distances is superior to that based on the projection statistics. In passing it should be mentioned that the other components of the measurement and process noise covariance estimates exhibit the same behavior as the $\tilde{\boldsymbol{R}}_k(1,1)$ estimate, and therefore only this one particular quantity is shown as an example.

Finally, the contamination parameter RMS error shows that the adaptive Huber filter is able to reduce the error in the assumed contamination parameter by means of the formula based on the robust weighting parameters in Eq. 42. Note that the adaptive Huber filter is the only filter investigated in this paper that can estimate the contamination parameter, which is evident from the fact that the other filter results show a constant RMS error. In this case, both the Kalman and adaptive Kalman filter show no error in the contamination parameter estimate, as these filters implicitly assume the errors are Gaussian and this simulation was conducted with only Gaussian errors. The Huber filter shows a constant RMS error at the value of the contamination parameter corresponding to the tuning parameter $\gamma = 1.5$. Due to Gaussianity, the adaptive Huber filter based on the Mahalanobis distances is superior to the filter based on the projection statistics.

B. Non–Gaussian Simulation

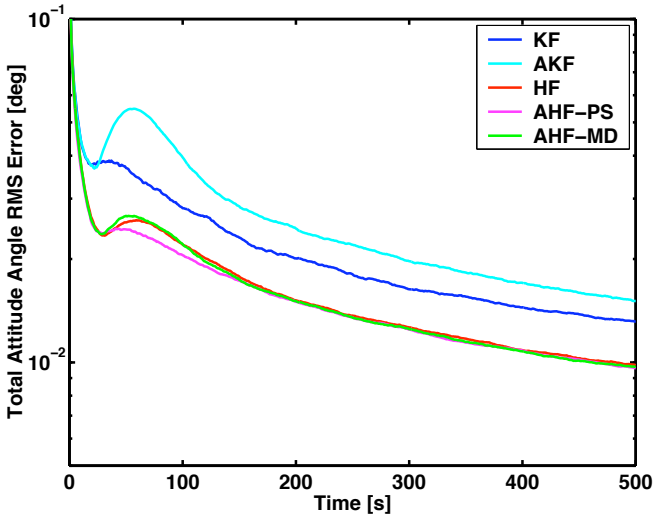
This section describes the results of a 2000 case Monte-Carlo simulation involving non–Gaussian errors, specifically for the mixture model in Eq. (58) with $\epsilon = 0.1$. The RMS error results corresponding to this simulation are shown in Fig. 4. The adaptive filters again use a buffer size of $N = 100$ stored residuals in order to compute the measurement and process noise covariance matrices.

For this case, the total attitude RMS errors clearly show the superiority of the Huber–based filtering methods. The adaptive Huber method exhibits the smallest RMS error, followed by the non–adaptive Huber method. The projection statistics-based adaptive Huber filter is superior to the filter based on the Mahalanobis distances, due to the non–Gaussianity in this example. Interestingly, the standard, non–adaptive Kalman filter exhibits smaller errors than that of the adaptive Kalman filter. This behavior is explored in more detail below. As with the previous case, the same trends seen in the attitude angle error appear in the gyroscope bias estimate RMS error, as expected.

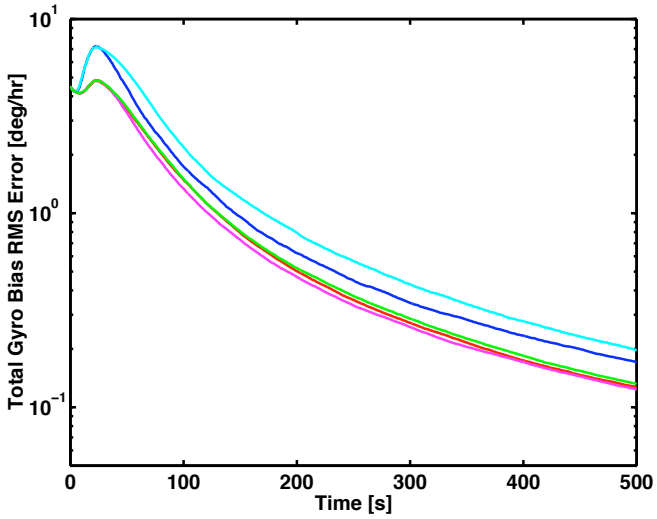
The predicted attitude and bias estimate errors show that the adaptive Huber–based filters are better predictors of the estimation errors than the other filters for the non–Gaussian case. As expected, the filter based on the projection statistics performs the best for both the attitude and bias error prediction. Of the non–adaptive filters, the Huber filter exhibits a better error prediction than the Kalman filter.

The RMS error in the $\tilde{\boldsymbol{R}}_k(1,1)$ estimate is quite interesting. Here, the adaptive Huber technique based on the projection statistics is able to reduce the error associated with the assumed measurement noise covariance, whereas the adaptive Huber filter using the Mahalanobis distances and the adaptive Kalman technique exhibit an *increased* error. This increase in the measurement noise error is the cause of the increased error in the attitude angle estimate as previously noted. Essentially, the adaptive Kalman technique associates the contaminated measurements with an increase in the measurement covariance, therefore causing the assumed covariance to inflate in order to account for the contamination. This covariance inflation is known to cause an increase in the estimation error due to the fact that *all* measurements are processed as if they were outliers, receiving less weight, than a technique that is able to distinguish outlying measurements and weight them accordingly.[29] Likewise, the adaptive Huber technique using the Mahalanobis distances suffers from the inherent limitations of the sample mean and covariance when applied to non-Gaussian problems.

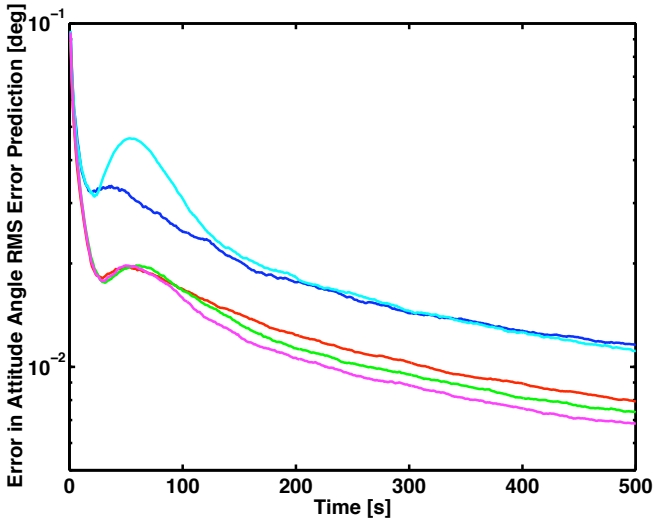
As discussed in the previous section, the only technique described in this paper to adaptively estimate the con-



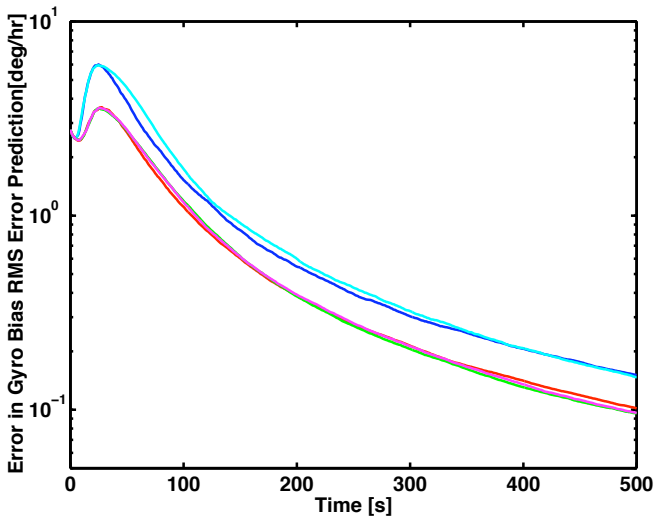
(a) Attitude Angle RMS Error



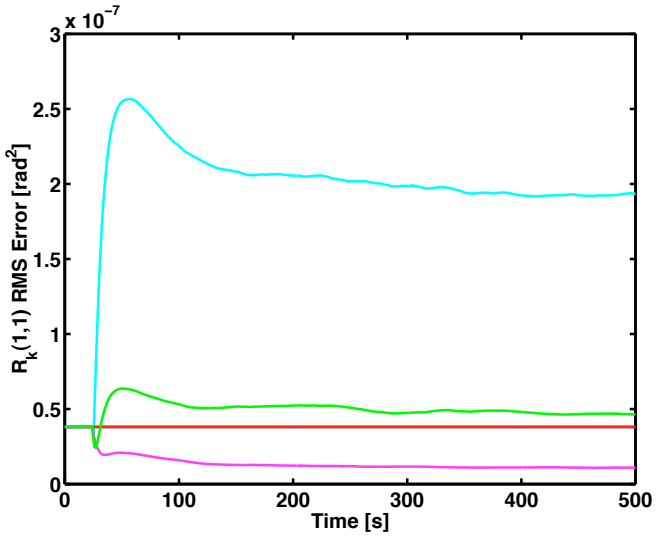
(b) Gyro Bias RMS Error



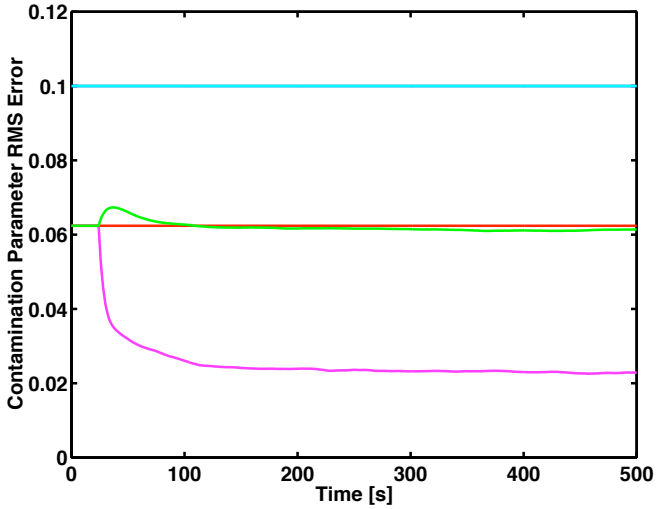
(c) Error in Attitude Angle RMS Error Prediction



(d) Error in Gyro Bias RMS Error Prediction



(e) $\tilde{\boldsymbol{R}}_k(1,1)$ RMS Error



(f) Contamination Parameter RMS Error

Figure 4. Contaminated Gaussian Simulation RMS Errors

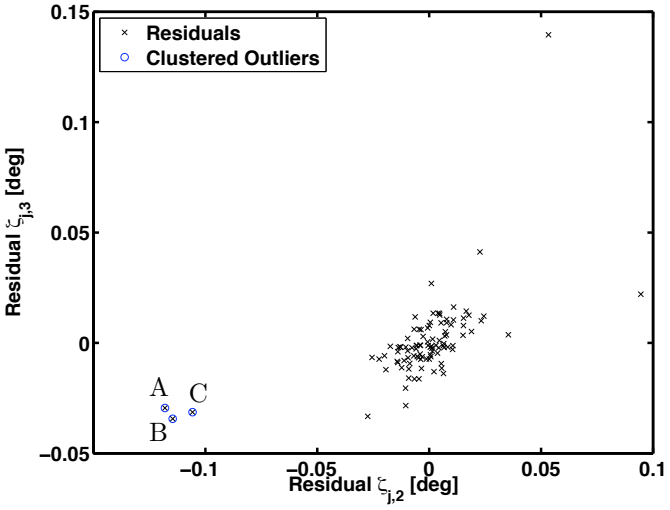


Figure 5. Example of Clustered Outliers in Adaptive Filter

tamination parameter associated with the noise densities are the adaptive Huber-based filters. In the non–Gaussian case, the adaptive Huber filter is successfully able to the reduce the error in the assumed contamination parameter. However, the results of the filter based on the projection statistics is far superior to that of the Mahalanobis distances. Essentially, the filter based on the Mahalanobis distances attempts to find a compromise between the crude covariance inflation of the adaptive Kalman filter and the Huber filter. The result is that the contamination parameter estimate error does not improve substantially over that of the non–adaptive Huber filter, and the covariance estimate inflates slightly but not to the extent of the adaptive Kalman filter. All other filters assume (explicitly or implicitly) a constant contamination parameter. In the case of the adaptive and non–adaptive Kalman filters, the contamination is assumed to be zero, thus the error is constant at 0.1 for this case. The Huber filter error again corresponds to the assumption leading to the tuning parameter $\gamma = 1.5$.

Table 1. Comparison of Identification Techniques for Clustered Outliers

Point	$\mathcal{M}_i$	$\mathcal{P}_i$
A	24.07	166.53
B	22.15	166.96
C	18.22	126.07

Figure 5 shows a scatter plot of the stored measurement residuals from the adaptive Huber technique, from an arbitrary stopping point in a non-Gaussian simulation, projected into the (2,3) plane. This sample includes 100 measurement residuals, with several apparent outliers. In particular, a group of 3 outliers appears in a cluster as indicated in the plot. The Mahalanobis Distances and Projection Statistics have been computed for this set of residuals, and the values corresponding to the clustered outliers is shown in Table 1. Compared with the $\chi^2_{3,0.95} = 7.81$ threshold, both methods are able to detect the clustered outliers in this example. However, the Projection Statistics is not influenced by the clustering and assigns these point much smaller weights than the Mahalanobis Distances. The Projection Statistics provide a better estimate of the contamination parameter, giving a value of 0.104, opposed to the estimate based on the Mahalanobis Distances of 0.051, for this set of residuals.

C. Sensitivity to the Adaptive Filter Buffer Size

The previous two sections have discussed the application of the various filtering techniques to both Gaussian and non–Gaussian cases, using a maximum buffer size of 100 stored residuals for the adaptive filters. This section explores the trade off in performance found by using a smaller buffer size, with the expectation that processing a smaller number of stored residuals reduces the computational costs of the filter. Additional 2000 case Monte–Carlo studies

were conducted for $N = 50$ and $N = 25$ for the cases $\epsilon = 0$ and $\epsilon = 0.1$.

Figure 6 shows the results of the attitude angle RMS error for each adaptive filter investigated in this paper. Each subplot shows the result of a particular filter for a range of N for the given value of ϵ. For instance, Fig. 6(a) shows results for the adaptive Kalman filter for a range of N and for $\epsilon = 0$. These results show that, for this particular problem, the buffer size can be made as low as $N = 25$ without severe degradation in the filter performance for the attitude error estimates. It is also interesting to note that the scales are all the same in each subplot in Fig. 6, and the sensitivity of the adaptive Kalman filter can clearly been detected for the $\epsilon = 0.1$ case compared with the $\epsilon = 0.0$ results. Likewise, the insensitivity of the Huber–based approaches is readily apparent.

Figure 7 shows the $\tilde{\boldsymbol{R}}_k(1,1)$ RMS errors for the adaptive filters for a range of N for both Gaussian and non–Gaussian cases. Differences are apparent in the filter performance depending on the type of filter and the parameter N, which controls the amount of data used in the adaptive filters for estimating the covariance matrices. In all cases, the $\boldsymbol{R}_{11}$ RMS error decreases as the parameter N is increased.

Figure 8 shows the contamination parameter estimates for the Huber filters based on the Mahalanobis Distances and the Projection Statistics. These results show interesting behavior in estimates based on the Mahalanobis distances in that they are actually more accurate for smaller sample sizes in the case of no contamination. This effect implies that the Mahalanobis distances tend to be optimistic regarding the contamination in small samples. This phenomenon is another well known limitation of the Mahalanobis distances: the small sample Mahalanobis distances tend to regard the entire sample as Gaussian, thus setting a lower contamination ratio, which in the Gaussian case leads to a lower error level.[10] In the non–Gaussian case, the opposite trend occurs and the error increases since the true distribution has nonzero contamination. The estimates based on the projection statistics are not as accurate for Gaussian cases but they do follow the expected trends of increasing accuracy with an increasing buffer size. In the small sample case (N=25) without contamination, the projection statistics tend to over predict the ratio of contamination.

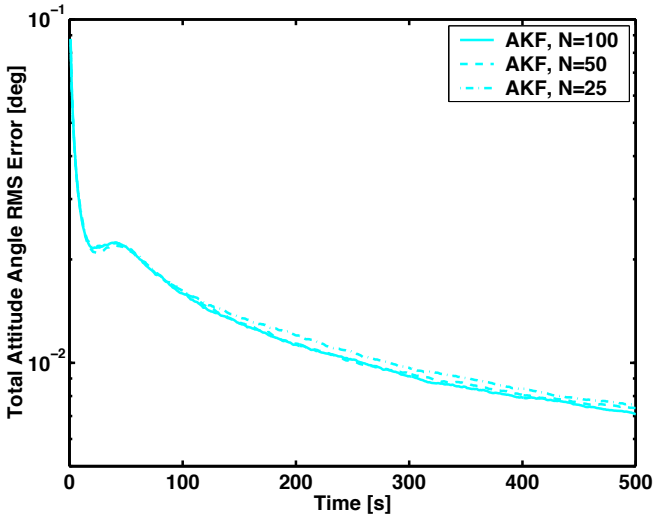
D. Computational Costs

Based on the results shown in the previous section, the Huber–based filtering methods show a clear advantage over traditional Kalman filter methods when applied to contaminated Gaussian measurement distributions. In the Gaussian case, the Huber–based methods do not suffer appreciably from any kind of mistuning associated with incorrect assumptions regarding the true underlying statistics of the problem. This section compares the computational costs associated with the Huber–based approaches compared with the Kalman filtering techniques. Table 2 shows the computational costs for each filter discussed in this paper. Here, the computational time is divided by that corresponding to the standard Kalman filter in order to provide a relative comparison of the computational burden associated with the particular filter.

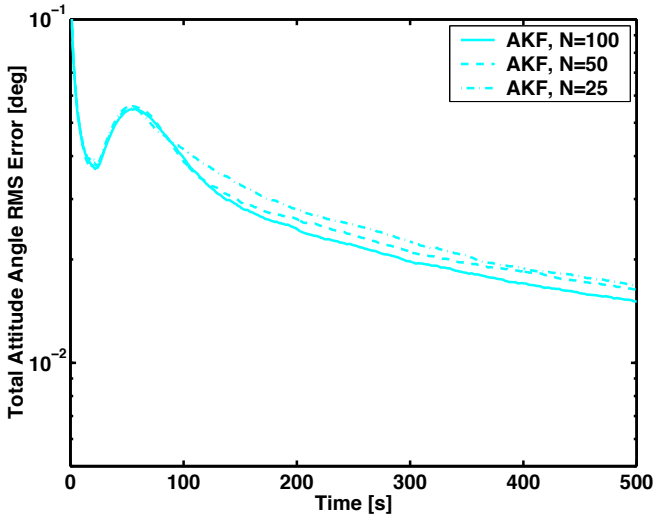
Table 2. Relative Computation Ratios

Filter	Nonadaptive	N=25	$N = 50$	$N = 100$
Kalman Filter	1.00	–	–	–
Adaptive Kalman Filter	–	1.13	1.16	1.24
Huber Filter	1.33	–	–	–
Adaptive Huber Filter (Mahalanobis Distances)	–	1.75	1.89	2.15
Adaptive Huber Filter (Projection Statistics)	–	2.86	4.29	7.18

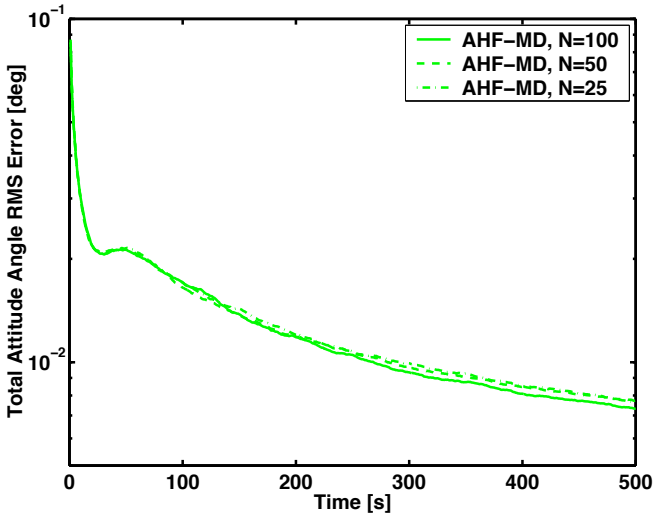
For this problem, the adaptive Kalman filter and the Huber filter are comparable in computational cost relative to the standard Kalman filter. The adaptive Huber filters proposed in this paper require more computation, well beyond that of the adaptive Kalman filter or the non-adaptive Huber filter as can be seen from Table 2. As expected, the computation associated with the adaptive filters is reduced as the buffer size reduced. Note that the computational cost of the filter based on the projection statistics is reduced at a higher rate than the filter based on the Mahalanobis distances. The adaptive Huber filter based on the projection statistics has good properties when applied to contaminated Gaussian distributions, however, one must trade off the computational costs to determine the feasibility of implementing the technique for any particular application.



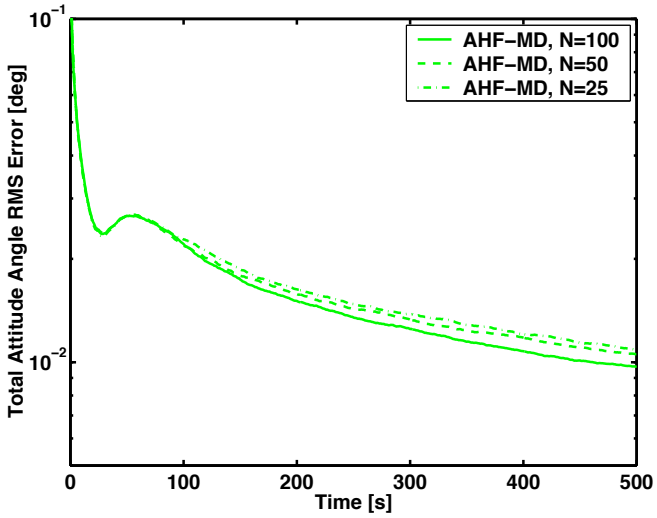
(a) Attitude Angle RMS Error: AKF ($\epsilon = 0$)



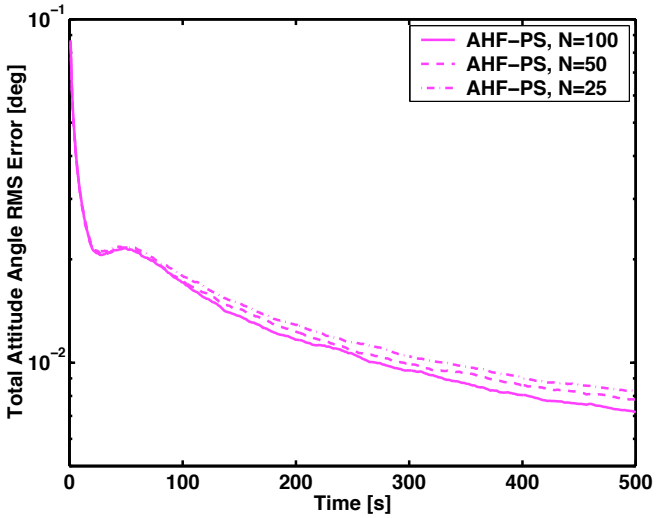
(b) Attitude Angle RMS Error: AKF ($\epsilon = 0.1$)



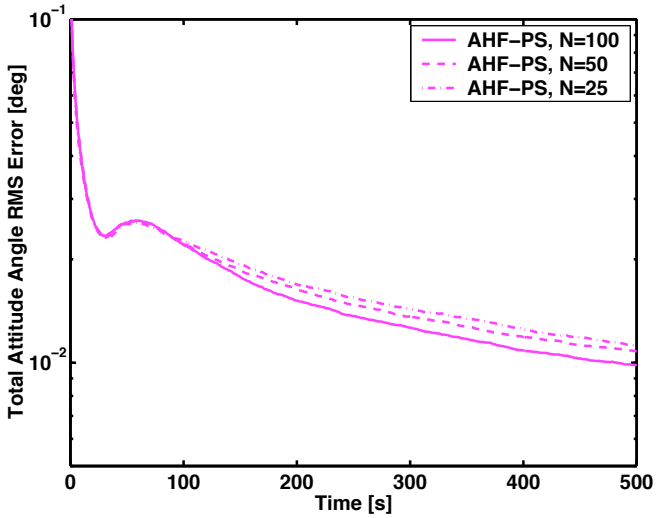
(c) Attitude Angle RMS Error: AHF–MD ($\epsilon = 0$)



(d) Attitude Angle RMS Error: AHF–MD ($\epsilon = 0.1$)

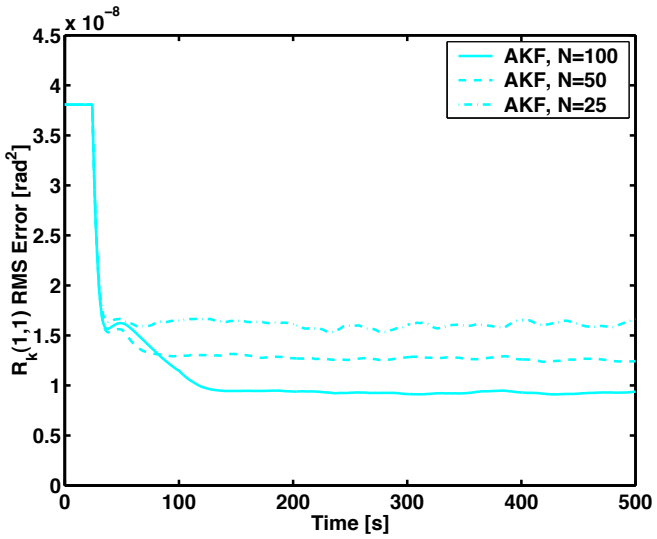


(e) Attitude Angle RMS Error: AHF–PS ($\epsilon = 0$)

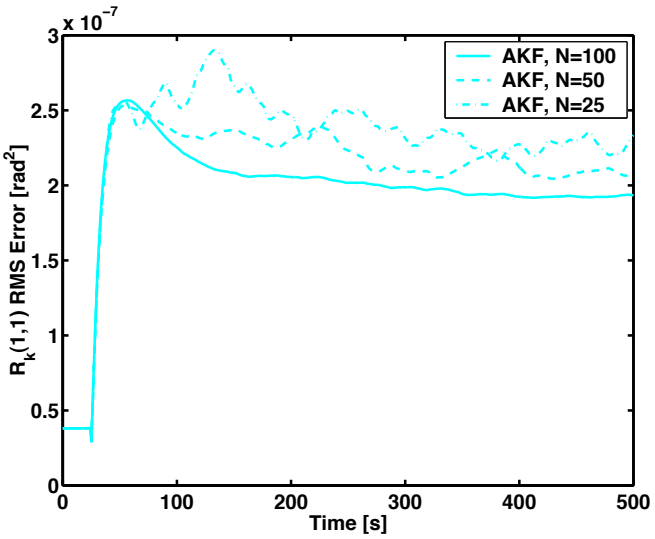


(f) Attitude Angle RMS Error: AHF–PS ($\epsilon = 0.1$)

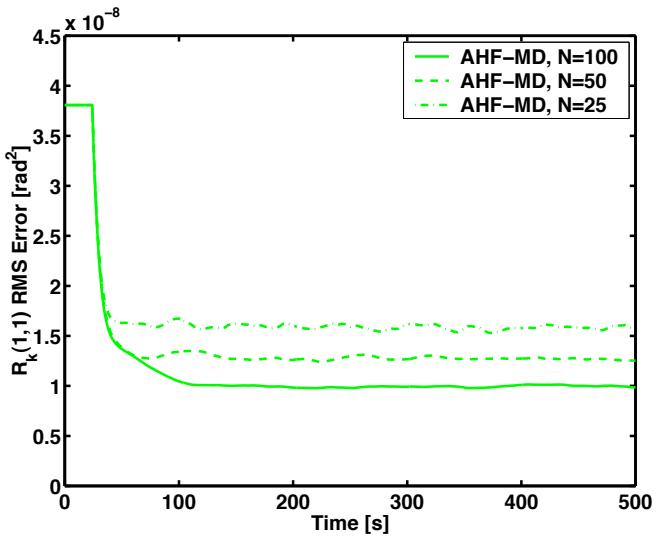
Figure 6. Attitude Angle RMS Errors for Varying Buffer Size



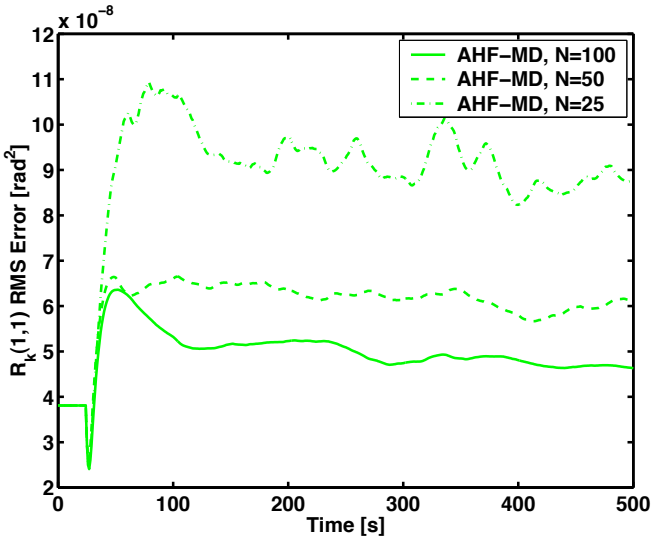
(a) $\tilde{\boldsymbol{R}}_k(1,1)$ RMS Error: AKF ($\epsilon = 0$)



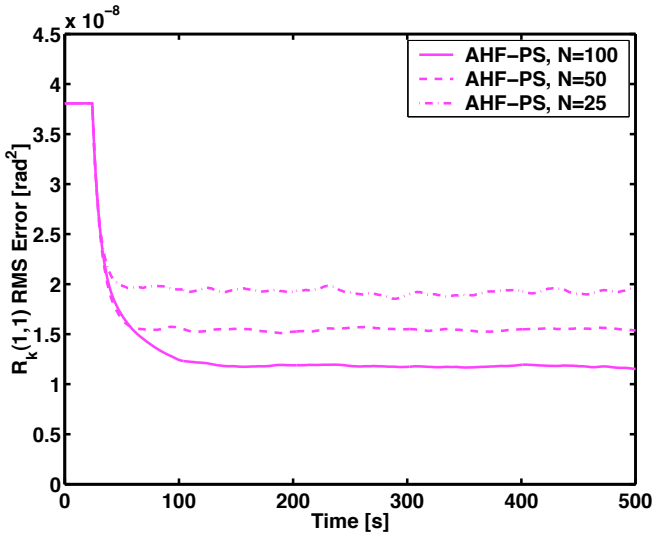
(b) $\tilde{\boldsymbol{R}}_k(1,1)$ RMS Error: AKF ($\epsilon = 0.1$)



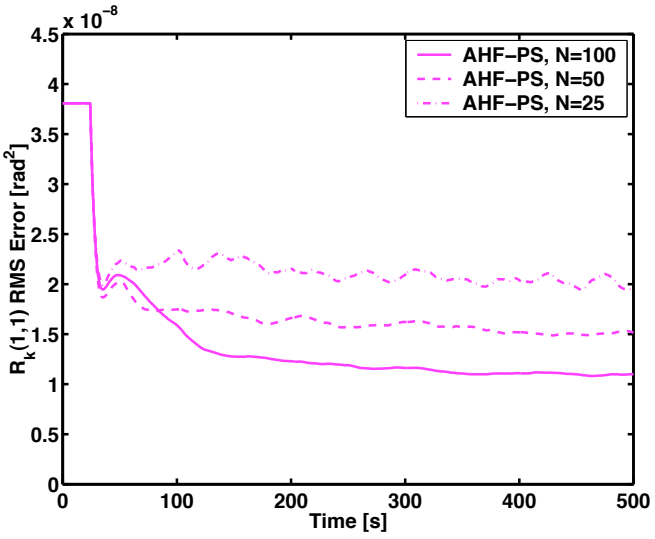
(c) $\tilde{\boldsymbol{R}}_k(1,1)$ RMS Error: AHF–MD ($\epsilon = 0$)



(d) $\tilde{\boldsymbol{R}}_k(1,1)$ RMS Error: AHF–MD ($\epsilon = 0.1$)

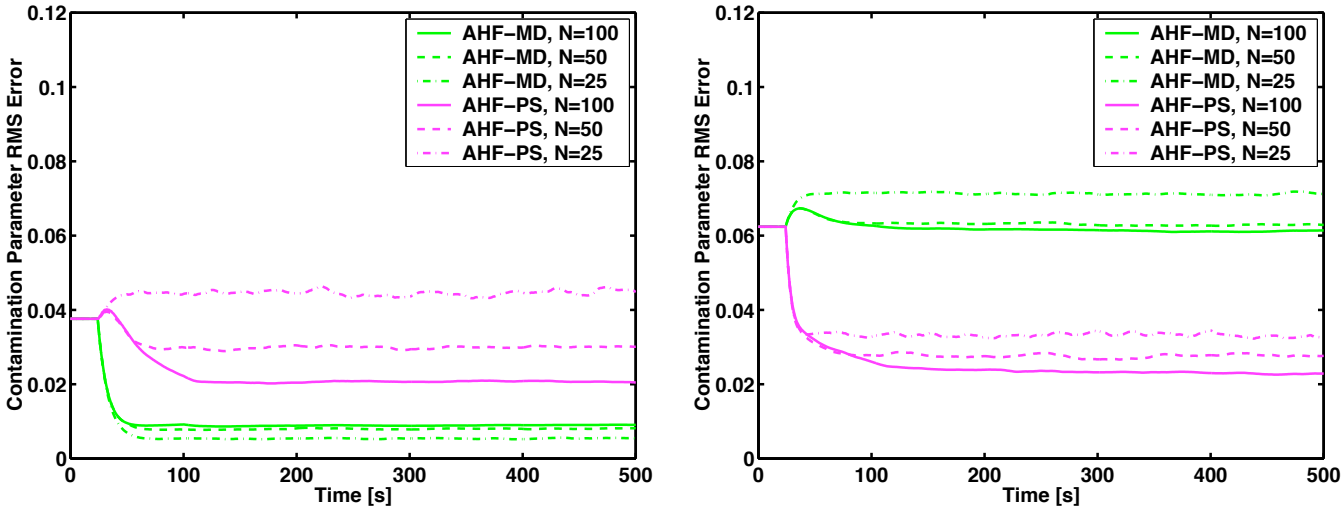


(e) $\tilde{\boldsymbol{R}}_k(1,1)$ RMS Error: AHF–PS ($\epsilon = 0$)



(f) $\tilde{\boldsymbol{R}}_k(1,1)$ RMS Error: AHF–PS ($\epsilon = 0.1$)

Figure 7. R_{11} RMS Errors for Varying Buffer Size



(a) Contamination Parameter RMS Error ($\epsilon = 0$)

(b) Contamination Parameter RMS Error ($\epsilon = 0.1$)

Figure 8. Contamination Parameter RMS Errors for Varying Buffer Size

VI. Conclusions

This paper discusses the development of an adaptive discrete-time robust filtering technique. The technique is based on a recursive form of Huber's mixed minimum ℓ_1/ℓ_2 norm approach, which is robust with respect to deviations from the assumed Gaussian error probability distributions inherent to the Kalman filter. An adaptive scheme based on covariance matching is proposed whereby the filter can estimate the process noise and measurement noise covariance matrices along with the state estimate and state estimate error covariance matrix using a buffer of stored residuals. The adaptation technique also adopts a robust approach to estimating these covariances that can resist the effects of outliers, based on the use of projection statistics. The projection statistics also afford a means for producing an estimate of the contamination parameter of the distribution in order to adaptively tune the filter.

The adaptive/robust filter was applied to the spacecraft attitude quaternion estimation problem, using the standard multiplicative approach for handling the quaternion norm constraint. Monte-Carlo results were shown for both Gaussian and non-Gaussian cases. Computational costs associated with the filtering techniques were provided. The results show that the proposed adaptive Huber filer can achieve greater accuracies than the Kalman filter in situations where unexpected non–Gaussian contaminating noise is present. The amount of improvement depends on the ratio of contamination but for the case of the 10% contamination, the adaptive Huber technique can improve accuracies of the Kalman filter by 50% or better. Furthermore, this adaptive Huber filter can successfully estimate the noise covariances in the non-Gaussian case as well as the contamination ratio whereas the adaptive Kalman filter is not capable of doing so. Therefore, the adaptive Huber filter has better consistency and is self–tuning. The drawbacks to the proposed adaptive Huber filter are that it is slightly less accurate (several percent) than the Kalman filter in instances of pure Gaussian noise, and that the computation involved can reach up to 7 times that of the Kalman filter depending on the buffer size.

References

[1]Huber, P. J., "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, Vol. 35, No. 2, 1964, pp. 73–101.

[2]Karlgaard, C. D. and Schaub, H., "Comparison of Several Nonlinear Filters for a Benchmark Tracking Problem," American Institute of Aeronautics and Astronautics, AIAA Paper 2006-6243.

[3]Karlgaard, C. D., "Robust Rendezvous Navigation in Elliptical Orbit," *Journal of Guidance, Control, and Dynamics*, Vol. 29, No. 2, 2006, pp. 495–499.

[4]Karlgaard, C. D. and Schaub, H., "Huber–Based Divided Difference Filtering," *Journal of Guidance, Control, and Dynamics*, Vol. 30, No. 3, 2007, pp. 885–891.

[5]Fitzgerald, R. J., "Divergence of the Kalman Filter," *IEEE Transactions on Automatic Control*, Vol. 16, No. 6, 1971, pp. 736–747.

[6]Mehra, R. K., "Approaches to Adaptive Filtering," *IEEE Transactions on Automatic Control*, Vol. 17, No. 5, 1972, pp. 693–698.

[7]Myers, K. A. and Tapley, B. D., "Adaptive Sequential Estimation with Unknown Noise Statistics," *IEEE Transactions on Automatic Control,* Vol. 21, No. 4, 1976, pp. 520–523.

[8]Maronna, R. A., Martin, R. D., and Yohai, V. J., *Robust Statistics: Theory and Methods*, Wiley, New York, 2006, pp. 104–105, 178–179.

[9]Gasko, M. and Donoho, D. L., "Influential Observations in Data Analysis," *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, 1982, pp. 104-110.

[10]Rousseeuw, P. J. and van Zomeren, B. C., "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association*, Vol. 85, No. 411, 1990, pp. 633–639.

[11]Rousseeuw, P. J. and van Zomeren, B. C., "Robust Distances: Simulations and Cutoff Values," In Stahel, W. and Weisberg, S. (Eds.), *Directions in Robust Statistics and Diagnostics, Part II*, Vol. 34, The IMA Volumes in Mathematics and its Applications, Springer-Verlag, New York, 1991, pp. 195–203.

[12]Mili, L., Cheniae, M. G., Vichare, N. S., and Rousseeuw, P. J., "Robust State Estimation Based on Projection Statistics," *IEEE Transactions on Power Systems*, Vol. 11, No. 2, 1996, pp. 1118–1125.

[13]Li., W.. Wei, P., and Xiao, X., "An Adaptive Nonlinear Filter of Discrete–Time System with Uncertain Covariance Using Unscented Kalman Filter," International Symposium on Communications and Information Technology, October 2005.

[14]Evensen, G., "The Ensemble Kalman Filter: Theoretical Formulation and Practicle Implementation," *Ocean Dynamics*, Vol. 53, No. 4, 2003, pp. 343–367.

[15]Huber, P. J., "Robust Statistics: A Review," *Annals of Mathematical Statistics*, Vol. 43, No. 4, 1972, pp. 1041–1067.

[16]Beaton, A. E. and Tukey, J. W., "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band–Spectroscopic Data," *Technometrics*, Vol. 16, No. 2, 1974, pp. 147–185.

[17]Huber, P. J., "The Behavior of Maximum–Likelihood Estimates Under Nonstandard Conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, LeCam, L. M. and Neyman, J. (Eds.), University of California Press, Berkeley, CA, Vol. 1, 1967, pp. 221–233.

[18]Crassidis, J. L. and Junkins, J. L., *Optimal Estimation of Dynamic Systems*, Chapman & Hall/CRC, Boca Raton, FL, 2004.

[19]Holland, P. W. and Welsch, R. E., "Robust Regression Using Iteratively Reweighted Least Squares," *Communications in Statistics: Theory and Methods*, Vol. 6, No. 9, 1977, pp. 813–827.

[20]Maybeck, P. S., Jensen, R. L., Harnly, D. A., "An Adaptive Extended Kalman Filter for Target Image Tracking," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 17, No. 2, 1981, pp. 173–180.

[21]Des Rosiers, A. P., Schoenig, G. N., and Mili, L., "Robust Space–Time Adaptive Processing Using Projection Statistics," International Conference on Radar Systems, Toulouse, France, October 2004.

[22]Rousseeuw, P. J. and Leroy, A. M., *Robust Regression and Outlier Detection*, John Wiley and Sons, New York, NY, 1987, pp. 158–174.

[23]Farrenkopf, R. L., "Analytic Steady–State Accuracy Solutions for Two Common Spacecraft Attitude Estimators," *Journal of Guidance and Control*, Vol. 1, No. 4, 1978, pp. 282–284.

[24]Kuhl, C. T. F., "Combined Earth–/Star Sensor for Attitude and Orbit Determination of Geostationary Satellites," Ph.D. Dissertation, Institute fur Flugmechanik und Flugregelung, Universitat Stuttgart, March 2005.

[25]Oshman, Y. and Carmi, A., "Attitude Estimation from Vector Observations Using Genetic-Algorithm Embedded Quaternion Particle Filter," *Journal of Guidance, Control, and Dynamics*, Vol. 29, No. 4, 2006, pp. 879–891.

[26]Lam, Q., Woodruff, C., Ashton, S., and Martin, D., "Noise Estimation for Star Tracker Calibration and Enhanced Precision Attitude Determination," *Proceedings of the Fifth International Conference on Information Fusion*, Vol. 1, 2002, pp. 235–242.

[27]Schaub, H. and Junkins, J. L., *Analytical Mechanics of Space Systems*, American Institute of Aeronautics and Astronautics, AIAA Education Series, Reston, VA, 2003, pp. 95–101.

[28]Kane, T. R., "Solution of Kinematical Differential Equations for a Rigid Body," *Journal of Applied Mechanics*, Vol. 40, No. 1, 1973, pp. 109–113.

[29]Schick, I. C. and Mitter, S. K., "Robust Recursive Estimation in the Presence of Heavy–Tailed Observation Noise," *Annals of Statistics*, Vol. 22, No. 2, 1994, pp. 105–1080.