# AIAA 2006-6243

## COMPARISON OF SEVERAL NON-LINEAR FILTERS FOR A BENCH-MARK TRACKING PROBLEM

Christopher D. Karlgaard
*Analytical Mechanics Associates, Inc., Hampton, VA 23666*
and Hanspeter Schaub
*Virginia Tech, Blacksburg, VA 24061-0203*

**AIAA Guidance, Navigation and Control Conference**
**Keystone, CO, August 21-24, 2006**

AIAA Guidance, Navigation, and Control Conference and Exhibit
21 - 24 August 2006, Keystone, Colorado

AIAA 2006-6243

# Comparison of Several Nonlinear Filters for a Benchmark Tracking Problem

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

Hanspeter Schaub†
*Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061-0203*

**This paper discusses the application of a robust filtering technique to a benchmark trajectory estimation problem involving radar measurements of an atmospheric entry vehicle. The robust technique is based on the Huber estimator, which is a mixed minimum $\ell_1/\ell_2$-norm linear regression technique. The approach is robust with respect to deviations in the traditionally assumed Gaussian measurement error probability distributions. It is shown how the Huber technique can be applied to the Extended Kalman Filter and to First and Second-Order Divided Difference Filter measurement update equations to provide robustness against non-Gaussian errors. The filters are applied to the benchmark problem using Monte-Carlo techniques, and computational costs associated with the filtering techniques are provided.**

## I. Introduction

THIS paper describes the application of several nonlinear filters to a benchmark problem involving estimating the trajectory of an entry body from discrete–time range data measured by a radar tracking station. The processing of the sensor measurements to estimate the target trajectory can be accomplished by one of several means. Perhaps the most basic estimation technique in this area of application is the Kalman filter.[1] The Kalman filter is a recursive weighted least–squares or minimum $\ell_2$ norm (sample or conditional mean) estimation procedure, and is a maximum likelihood technique assuming that the error statistics follow Gaussian probability distributions.[2, 3]

Unfortunately, the least–squares method is not a robust estimation technique because the procedure can fail to perform adequately when the true error statistics follow non-Gaussian probability distributions, particularly those with much thicker tails than the Gaussian distribution.[4, 5] Thickly–tailed probability distributions are those with high probability of large errors compared with the Gaussian distribution. The term *robustness* in a statistical sense was coined by Box[6] to describe procedures that are insensitive to deviations in the assumed underlying probability distribution from which the data are sampled, although the sensitivity of the least squares technique to the distribution of the data has been known for longer. In fact, in the first publication on least squares, Legendre suggests rejecting data that appear erroneous before using least squares on the remaining observations.[7] Formal definitions of robust estimation can be found in Refs. 9–11. Intuitively, robustness means that the estimation error remains bounded for arbitrarily large observation residuals. As radar systems are known to exhibit thick tailed non–Gaussian random measurement errors,[8] it is important to develop estimation techniques for radar–based tracking that are robust with respect to deviations from the commonly assumed Gaussian measurement error distribution.

One robust technique is the Huber estimator,[12] which is a combined minimum $\ell_1$ and $\ell_2$ norm estimation method. The Huber estimator is a blend of the two estimators that seeks to take the best of both techniques. Specifically, the Huber estimator takes the robustness of the sample median and the efficiency of the sample mean. The Huber technique was originally developed as a generalization of maximum likelihood estimation, applied first to estimating the center of a probability distribution in Ref. 12 and further generalized to multiple linear regression in Refs. 13–15.

The Huber technique has been further extended to dynamic estimation problems. By expressing the discrete-time Kalman filter as a sequence of linear regression problems,[16] Boncelet and Dickinson[17] first proposed to solve the robust filtering problem by means of the Huber technique at each measurement point. The authors do not provide

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

†Assistant Professor, Department of Aerospace and Ocean Engineering, Virginia Polytechnic Institute and State University, 228 Randolph Hall, Blacksburg, VA, 24061-0203. schaub@vt.edu. Senior Member AIAA.

any simulation results to validate the proposed technique. Kovacevic, *et al*[18] follow the work of Ref. 17 and develop a robust filter using the Huber technique applied to a linear regression problem at each measurement update. Refs. 19–21 express the dynamic filtering problem as a sequential linear regression to be solved by the Huber technique, and apply the filter to underwater vehicle tracking, power system state estimation, and spacecraft rendezvous navigation, respectively.

Some other approaches to robust filtering include several ad-hoc techniques for dealing with observation data that do not follow the assumed Gaussian distribution. In particular, that of *censoring* the data such that any observation that differs from the predicted value by some set threshold is discarded entirely. This approach to robust filtering has numerous disadvantages that have been known for quite some time,[22] notably that the non-continuous weighting function can lead to non-robust covariance estimates.[5, 9] Another disadvantage of the censoring method is the increase in estimation variance due to the fact that information contained within the residual is discarded entirely and not processed according to a statistical procedure. Another ad-hoc technique is to simply inflate the measurement error covariance matrix to account for the perturbing density that leads to the non-Gaussianity. This method also suffers from several drawbacks, namely that *all* measurements processed according to this rule receive decreased weights, not only those that are to be considered as outliers. This reduction in overall weighting in turn leads to an increase in the estimation variance.[5]

Other techniques for robust filtering have been proposed. Sorenson and Alspach[23, 24] advocate a Bayesian approach in which the measurement error distributions are adaptively approximated using Gaussian sums. The Gaussian sum methodology yields a robust estimation technique, but at an extreme cost in computation. Specifically, the number of terms kept in the Gaussian sum grows exponentially with time. The computation time can be reduced using filter banks and parallel processing techniques,[25, 26] but the overall computational complexity across the entire processing system remains substantial. For these reasons, the Gaussian sum filters are not feasible for real–time implementation in many applications. Masreliez and Martin[27] develop a robust filter by means of a Bayesian approach in which either the state noise is Gaussian and the measurement noise is non-Gaussian, or the state noise is non-Gaussian and the measurement noise is Gaussian. No consideration is made for the case when both noise sources are non-Gaussian, which is allowed for under the Huber filtering approach.

Recent developments in dynamic filtering are the Sigma Point Kalman Filters (SPKFs). Like the basic Kalman filter, the SPKFs seek to determine a state estimate that minimizes the $\ell_2$–norm of the residuals. The SPKF technique differs from the basic Kalman filter in the estimate and covariance propagation prior to determining the state estimate at the time of the measurement update. The differences lie in that the SPKFs do not linearize the dynamic system for the propagation, but instead propagate a cluster of points centered around the current estimate in order to form improved approximations of the conditional mean and covariance. SPKFs have the additional advantage over the basic Kalman filter in that they can easily be extended to determine second–order solutions to the minimum $\ell_2$–norm measurement update, which increases the estimation accuracy when the system and measurement equations are nonlinear. It is important to note that the SPKFs use a minimum $\ell_2$–norm measurement update and are therefore subject to the same sensitivity to non-Gaussian measurement errors as the Kalman filter.

This paper first discusses the Huber estimation technique as a generalization of the method of maximum likelihood regression. Next, the dynamic state estimation problem is shown to be equivalent to a sequence of regression problems occurring at each measurement update. When written in the sequential regression form, the dynamic estimation problem can be solved by means of the robust Huber estimator. It is further shown how the robust Huber measurement update solution can be applied to the update equations of the first–order (DD1) and second–order Divided Difference Filter (DD2), which is one subclass of the SPKF technique. Finally, the robust techniques are applied to a benchmark tracking problem. The simulation is conducted using Monte-Carlo techniques for both the Gaussian and non-Gaussian cases. The computational cost associated with each filter is discussed. The next section begins the discussion of the Huber estimator as a robust solution of the linear regression problem.

## II. Some Concepts in Regression Theory

The purpose of this section is to review some basic concepts in regression theory. The estimation of a static linear system will later be shown to be easily extended to discrete-time dynamic estimation. In this section the linear regression model of the form

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{1}$$

is considered, where $\mathbf{y}$ represents the measurement data, $\mathbf{x}$ is the variable to be estimated, $\mathbf{H}$ represents the relationship between $\mathbf{x}$ and $\mathbf{y}$ and is usually called the design matrix, and $\mathbf{w}$ represents the random measurement error.

## A. Maximum Likelihood Estimation

### 1. Overview of the Method

The method of maximum likelihood estimation was introduced by Fisher[28] in a series of papers between 1912 and 1922, although some of the basic ideas had been touched on previously by Bernoulli[29] for semi-circular distributions, and Gauss [30] in connection with an effort to determine the probability distribution that made the least squares estimate the most probable estimate.[13, 31] The maximum likelihood method proceeds as follows. Given a sample of $m$ measurement data $\mathbf{y} = [y_1, \cdots, y_m]^T$, let the joint probability density function of the measurement data for a given value of $\mathbf{x}$ be denoted by $f(\mathbf{y}|\mathbf{x})$. The likelihood function $L$ is then defined as

$$L(\mathbf{x}; \mathbf{y}) = f(\mathbf{y}|\mathbf{x}) \tag{2}$$

If the observation data are considered to be independent and identically distributed, then the joint density function is the product of the marginal density functions.[32] Also, the density function reduces to a function only of the residuals between the estimate and the observation data, defined as $\boldsymbol{\zeta} = \mathbf{H}\mathbf{x} - \mathbf{y}$. Under these assumptions, the likelihood function takes the form

$$L(\mathbf{x}; \mathbf{y}) = \prod_{i=1}^{m} f(\boldsymbol{\zeta}) \tag{3}$$

The most probable value of $\mathbf{x}$ (the mode of the joint distribution) is the value of $\mathbf{x}$ that maximizes the likelihood function. In practice it is more convenient to minimize the natural logarithm of the likelihood function. Such a value of $\mathbf{x}$ can be found by minimizing the cost function,

$$J(\mathbf{x}) = -\sum_{i=1}^{m} \ln\left[f\left(\zeta_i\right)\right] \tag{4}$$

The solution to the minimization problem is given implicitly by the derivative of the cost function

$$\sum_{i=1}^{m} \phi(\zeta_i) \frac{\partial \zeta_i}{\partial \mathbf{x}} = \mathbf{0} \tag{5}$$

where $\phi(\zeta_i) = -f'(\zeta_i)/f(\zeta_i)$. By defining the function $\psi\left(\zeta_i\right) = \phi(\zeta_i)/\zeta_i$, and the matrix $\boldsymbol{\Psi} = \mathrm{diag}\left[\psi\left(\zeta_i\right)\right]$, the implicit equation can be written in matrix form as

$$\mathbf{H}^T \boldsymbol{\Psi} \left(\mathbf{H}\mathbf{x} - \mathbf{y}\right) = \mathbf{0} \tag{6}$$

Note that, in general, the implicit likelihood equation (6) is nonlinear, since the matrix $\boldsymbol{\Psi}$ is a function of $\mathbf{x}$. Solutions of the likelihood equation can be determined by the application standard iterative techniques for nonlinear systems. The solution of the likelihood equation is the maximum likelihood estimate for $\mathbf{x}$ and is denoted by $\hat{\mathbf{x}}$.

Once the solution of the maximum likelihood method is obtained, the covariance associated with the estimate can be computed from[9]

$$\hat{\mathbf{P}} = \frac{\mathrm{E}\left(\phi^2\right)}{\left[\mathrm{E}\left(\phi'\right)\right]^2} \left(\mathbf{H}^T\mathbf{H}\right)^{-1} \tag{7}$$

It can be shown that estimators of this form are asymptotically normal with mean $\hat{\mathbf{x}}$ and covariance $\hat{\mathbf{P}}$. Further, the estimator is minimum variance and asymptotically unbiased, when the true measurement error distribution follows the assumed model exactly. See Ref. 33 for further details.

### 2. Correlated Residuals

The development of the maximum likelihood estimation method made the assumption that the residuals were independent and identically distributed. If in fact the residuals are correlated then the maximum likelihood method is still applicable, but a transformation of variables is required. If the measurement error covariance matrix is given by $\mathbf{R}$, then the regression problem can be transformed to uncorrelated coordinates (this transformation is sometimes known as stochastic decoupling) by defining the new variables

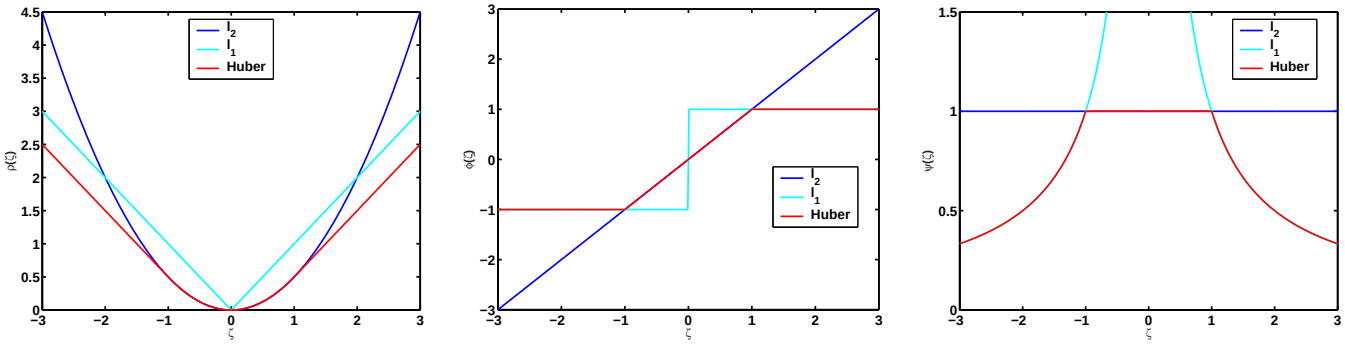
$$\widetilde{\mathbf{w}} = \mathbf{R}^{-1/2}\mathbf{w} \tag{8}$$

$$\widetilde{\mathbf{y}} = \mathbf{R}^{-1/2}\mathbf{y} \tag{9}$$

$$\widetilde{\mathbf{H}} = \mathbf{R}^{-1/2}\mathbf{H} \tag{10}$$

Upon defining the above quantities, the regression problem is transformed into

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{H}}\mathbf{x} + \widetilde{\mathbf{w}} \tag{11}$$




(a) Comparison of $\rho$ Functions (b) Comparison of $\phi$ Functions (c) Comparison of $\psi$ Functions

**Figure 1. Comparison of $\rho$, $\phi$, and $\psi$ Functions**

In the transformed problem, the measurement errors have a covariance matrix given by $\widetilde{\mathbf{R}} = \mathbf{I}$, as can be seen by expanding the expectation $\mathrm{E}\left[\widetilde{\mathbf{w}}\widetilde{\mathbf{w}}^T\right]$. This fact implies that the transformed residuals are linearly independent and hence that the standard maximum likelihood method can be applied to the transformed linear regression problem.

## B. Generalized Maximum Likelihood Estimation

### 1. *Overview of the Method*

Huber in 1964[12] introduced a new class of estimation methods known as the generalized maximum likelihood method, also known as the technique of $M$-estimation. In this method, one seeks to minimize a function of the residuals of the form

$$J\left(\mathbf{x}\right) = \sum_{i=1}^{m} \rho\left(\zeta_i\right) \tag{12}$$

where $\rho$ is an arbitrary function. Note that the generalized maximum likelihood method reduces the standard maximum likelihood method in the special case when $\rho(\zeta_i) = -\ln\left[f\left(\zeta_i\right)\right]$. The solution to the generalized maximum likelihood problem can be found from the implicit equation

$$\sum_{i=1}^{m} \phi(\zeta_i) \frac{\partial \zeta_i}{\partial \mathbf{x}} = \mathbf{0} \tag{13}$$

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

$$\mathbf{H}^T \mathbf{\Psi}\left(\mathbf{H}\mathbf{x} - \mathbf{y}\right) = \mathbf{0} \tag{14}$$

In generalized maximum likelihood estimation, the function $\rho$ can be chosen to yield an estimator $\hat{\mathbf{x}}$ with certain desirable properties. A 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.

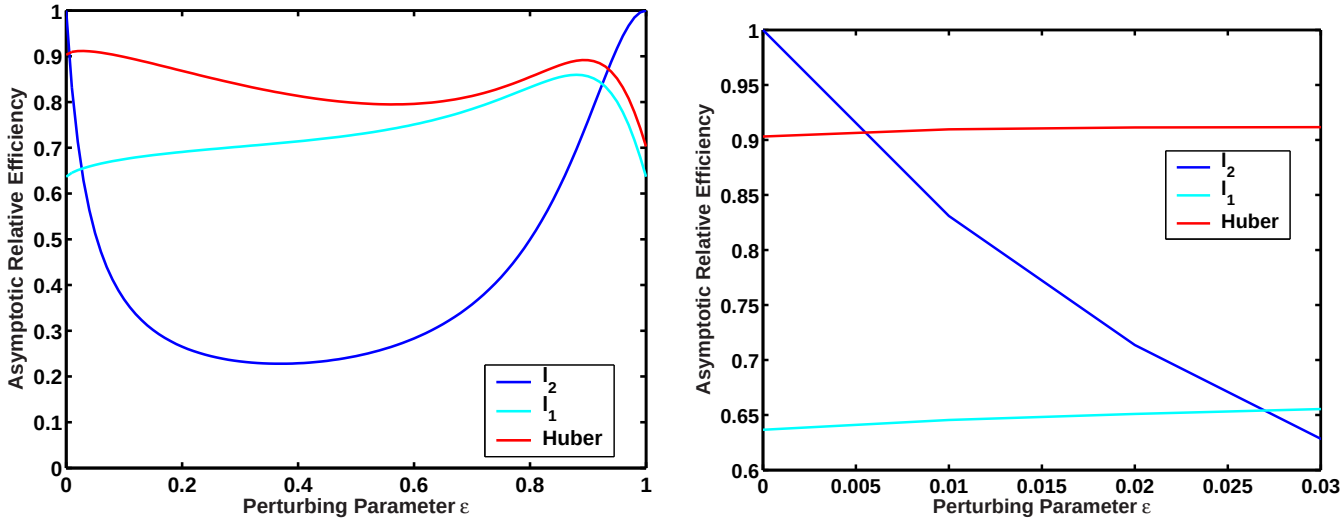
Huber[12] introduces a $\rho$ function of the form

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

where $\gamma$ is a tuning parameter. This $\rho$ function is a blend of the minimum $\ell_1$ and $\ell_2$ norm functions. Note that as $\gamma \to 0$, Eq. (15) approaches the $\ell_1$ norm, which is equivalent to the median in the scalar case, and that as $\gamma \to \infty$, Eq. (15) approaches the $\ell_2$ norm, which is equivalent to the mean in the scalar case. The $\rho$, $\phi$, and $\psi$ functions for the $\ell_1$, $\ell_2$, and Huber cases are shown in Fig. 1 for comparison. Note that the $\rho$, $\phi$, and $\psi$ functions are sometimes known as the score, influence, and weight functions, respectively.

Huber[12] shows that the $\rho$ function given in Eq. (15) is asymptotically optimally robust in the $\epsilon$ neighborhood of the Gaussian distribution. Explicitly, if the measurement errors follow a perturbed Gaussian density function of the form

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




(a) Comparison of Asymptotic Relative Efficiencies

(b) Comparison of Asymptotic Relative Efficiencies (Detail)

**Figure 2. Comparison of Asymptotic Relative Efficiencies for Gaussian Mixture**

where $\epsilon$ is a perturbing parameter and $g(w)$ is an unknown perturbing density function, then the $M$-estimation technique with $\rho$ function given by Eq. (15) minimizes the maximum asymptotic estimation variance in the $\epsilon$ neighborhood of the Gaussian distribution. Further, it can be shown that estimators of this form are asymptotically normal and unbiased.[34]

### 2. *Example of a Gaussian Mixture*

To illustrate the advantages of the generalized maximum likelihood techniques, consider for a moment the scalar estimation problem found by setting $\mathbf{H} = 1$. This problem corresponds with estimating the center of a probability distribution given a sample of measurements. In this example, the measurement data are drawn from a mixture of two Gaussian densities given by the function

$$f(w) = \frac{1-\epsilon}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right) + \frac{\epsilon}{\sqrt{2\pi}a} \exp\left(-\frac{w^2}{2a^2}\right) \tag{17}$$

where $a^2 > 1$ is the variance of the perturbing density. The mixture of Gaussian densities in Eq. 17 was first introduced by Newcomb for the purpose of studying heavy–tailed measurement data.[7, 35] Since $\mathbf{H} = 1$, the estimation variance is a scalar quantity that simplifies from Eq. (7) as $\hat{P} = \mathrm{E}\left(\phi^2\right) / \left[\mathrm{E}\left(\phi'\right)\right]^2$. The $\ell_1$, $\ell_2$, and Huber estimation variances can be found by carrying out the the necessary integrals. If the asymptotic variance of the $\ell_1$, $\ell_2$, and Huber estimators are denoted by $\hat{P}_{\ell_1}$, $\hat{P}_{\ell_2}$, and $\hat{P}_H$, respectively, then the results are

$$\hat{P}_{\ell_1} = \frac{\pi}{2}\left(1 - \epsilon + \frac{\epsilon}{a}\right)^{-2} \tag{18}$$

$$\hat{P}_{\ell_2} = 1 + \epsilon\left(a^2 - 1\right) \tag{19}$$

$$\begin{aligned}\hat{P}_H = &\left\{\gamma^2 + (1-\epsilon)\left[\left(1-\gamma^2\right)\mathrm{erf}\left(\frac{\gamma}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}}\gamma\exp\left(\frac{-\gamma^2}{2}\right)\right]\right. \\ &\left.+\epsilon\left[\left(a^2-\gamma^2\right)\mathrm{erf}\left(\frac{\gamma}{\sqrt{2}a}\right) - \sqrt{\frac{2}{\pi}}a\gamma\exp\left(\frac{-\gamma^2}{2a^2}\right)\right]\right\} \cdot \left[(1-\epsilon)\,\mathrm{erf}\left(\frac{\gamma}{\sqrt{2}}\right) + \epsilon\mathrm{erf}\left(\frac{\gamma}{\sqrt{2}a}\right)\right]^{-2}\end{aligned} \tag{20}$$

A useful metric for comparing estimators is the relative efficiency, which is the ratio of the maximum likelihood asymptotic variance and the asymptotic variance of the estimator. If the asymptotic variance of the maximum likelihood estimator is denoted by $\hat{P}_{MLE}$, then the asymptotic relative efficiency is defined as $e = \hat{P}_{MLE}/\hat{P}$, where $\hat{P}$ is the asymptotic variance of the estimator in question. Note that the best possible asymptotic relative efficiency is $e = 1$, since the maximum likelihood estimator is asymptotically minimum variance at the particular distribution.

The asymptotic relative efficiencies are shown in Fig. 2 as a function of the perturbing parameter $\epsilon$ for fixed $a = 5$ and $\gamma = 1$. The results for the entire range of $\epsilon$ is shown in Fig. 2(a) where just the portion of $\epsilon$ between 0 and 0.03 is shown in Fig. 2(b). From the plots in Fig. 2 it is apparent that when the data are perfectly Gaussian, the

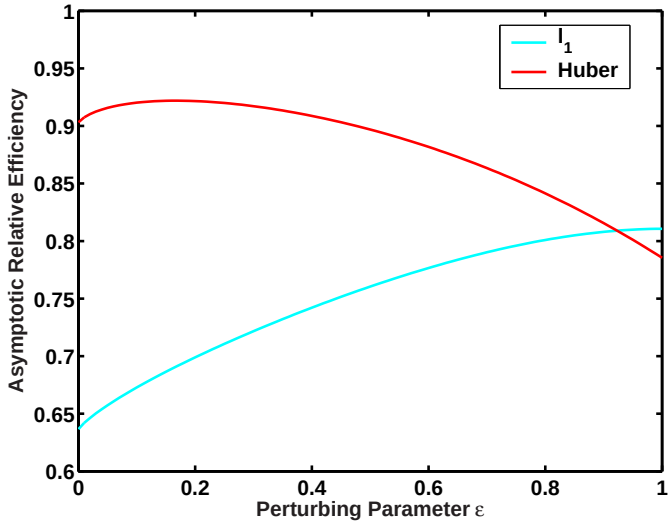



**Figure 3. Asymptotic Relative Efficiencies for Gaussian–Cauchy Mixture**

$\ell_2$ estimator is more efficient than either the $\ell_1$ or the Huber estimator. This behavior is to be expected since the $\ell_2$ estimator is the maximum likelihood estimator at the Gaussian distribution. When the perturbing parameter $\epsilon$ is increased, the $\ell_2$ efficiency decreases sharply, but the $\ell_1$ and Huber efficiencies show a more shallow decrease than the $\ell_2$ case. The Huber and $\ell_1$ estimators rapidly show higher efficiencies than the $\ell_2$ case as $\epsilon$ increases away from zero. The $\ell_1$ estimator becomes superior to the $\ell_2$ estimator at $\epsilon = 0.027$ and the Huber estimator becomes superior to the $\ell_2$ estimator at $\epsilon = 0.005$. As $\epsilon$ increases further, the distribution function again becomes nearly Gaussian for which the $\ell_2$ estimator has the minimum variance. This example illustrates the point that a small contamination of an assumed Gaussian model by another distribution will lead to a rapid loss of efficiency of the least squares method, whereas the Huber estimator exhibits a reduced sensitivity. In fact, the Huber efficiency is nearly constant across a wide range of $\epsilon$, illustrating the robustness of the technique.

### *3. Example of a Gaussian–Cauchy Mixture*

As a slightly more extreme example, consider the Gaussian–Cauchy mixture given by the density

$$f(w) = \frac{1-\epsilon}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right) + \frac{\epsilon}{\pi} \frac{1}{1+w^2} \tag{21}$$

Note that for any $\epsilon > 0$, the variance of this distribution is infinite. Therefore, the asymptotic variance of the $\ell_2$ estimator applied to data drawn from this distribution is also infinite for $\epsilon > 0$. The $\ell_1$ and Huber estimator asymptotic variances are finite and given by

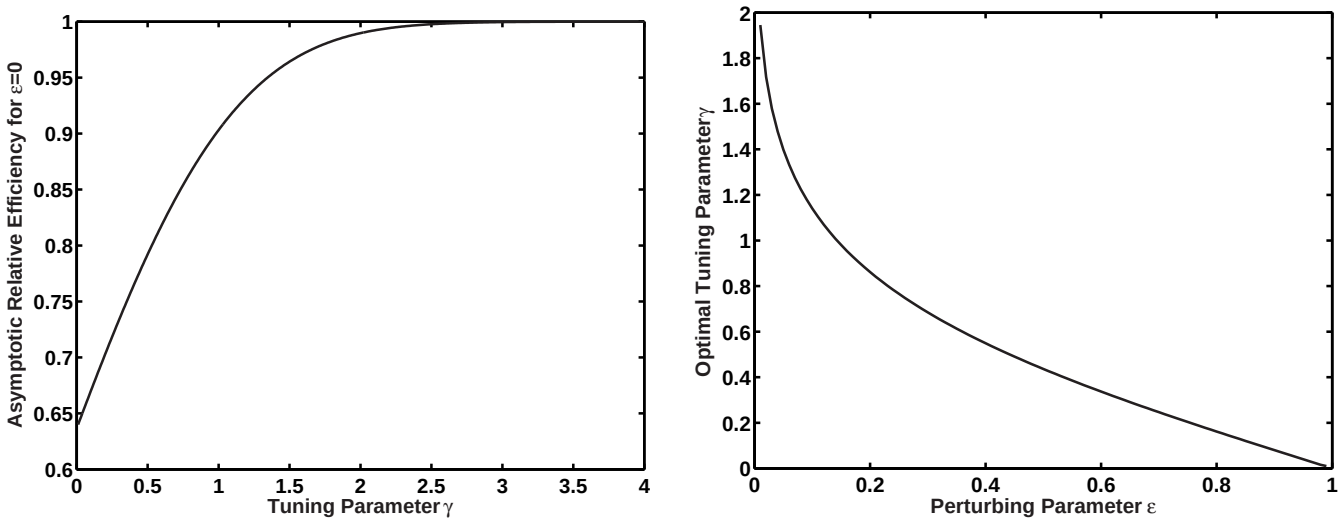
$$\hat{P}_{\ell_1} = \left[\sqrt{\frac{2}{\pi}}(1-\epsilon) + \frac{2\epsilon}{\pi}\right]^{-2} \tag{22}$$

$$\begin{aligned}\hat{P}_H = &\left[\gamma^2 + \frac{2\gamma\epsilon}{\pi} + \gamma\sqrt{\frac{2}{\pi}}(\epsilon-1)\exp\left(\frac{-\gamma^2}{2}\right) - \frac{2\epsilon}{\pi}\left(1+\gamma^2\right)\arctan\gamma + \right.\\ &\left.\left(1-\gamma^2-\epsilon+\gamma^2\epsilon\right)\operatorname{erf}\left(\frac{\gamma}{\sqrt{2}}\right)\right] \cdot \left[\frac{2\epsilon}{\pi}\arctan\gamma + (1-\epsilon)\operatorname{erf}\left(\frac{\gamma}{\sqrt{2}}\right)\right]^{-2}\end{aligned} \tag{23}$$

The $\ell_1$ and Huber asymptotic relative efficiencies are shown in Fig. 3. Here it is apparent that the Huber estimation technique performs quite well, even when applied to a problem where the random measurement errors have infinite variance.

### *4. Choice of Tuning Parameter*

Figure 4(a) shows the influence of the choice of the tuning parameter $\gamma$ on the asymptotic relative efficiency of the Huber technique for the Gaussian case ($\epsilon = 0$). The previous examples have shown that for a penalty in asymptotic




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

(b) Optimal Tuning Parameter $\gamma$

**Figure 4. Huber Gaussian Asymptotic Relative Efficiency and Optimal Tuning Parameter**

relative efficiency at the model distribution (the Gaussian case), the relative efficiency is nearly constant in an $\epsilon$ neighborhood of the model distribution. The efficiency can vary between one for $\gamma = \infty$ (the $\ell_2$ case) and $2/\pi$ for $\gamma = 0$ (the $\ell_1$ case). If the value of the parameter $\epsilon$ is known (even though the density function $g(w)$ of the perturbing distribution is unknown), then the best choice of $\gamma$ has been determined by Huber.[12, 13] If $\epsilon = \epsilon_0$ is the known perturbing parameter, then the optimal choice of $\gamma$, denoted by $\gamma^\star$, is given by the implicit equation[12, 13]

$$\frac{1}{1-\epsilon_0} = \frac{1}{\gamma^\star}\sqrt{\frac{2}{\pi}}\exp\left[-\frac{(\gamma^\star)^2}{2}\right] + \operatorname{erf}\left(\frac{\gamma^\star}{\sqrt{2}}\right) \tag{24}$$

For a given value of $\epsilon$, estimates computed using $\gamma = \gamma^\star$ maximize the asymptotic relative efficiency across the range of all possible values of $\gamma$.[12, 13]

The optimal value of the tuning parameter $\gamma^\star$ is shown as a function of $\epsilon_0$ in Fig. 4(b). When $\epsilon_0 = 0$ the best tuning parameter is infinite. This value is expected because when $\epsilon_0 = 0$, the distribution is Gaussian and least squares estimation ($\gamma = \infty$) is minimum variance. As $\epsilon_0 \to 1$, the tuning parameter $\gamma^\star \to 0$. This behavior suggests that the median is the best choice of estimation technique when the noise density $f(w) = g(w)$ is completely unknown. However, note that if $\epsilon_0 > 1/2$ then the model distribution should no longer be taken as Gaussian since the perturbing density is in fact greater in proportion.

If the perturbing parameter $\epsilon$ is unknown, then the choice of $\gamma$ is typically motivated by a desired variance at the model distribution. One common choice is $\gamma = 1.345$; for this value the Huber filter will exhibit estimation error variances that are 5% larger than that of the least squares method when the measurement error distributions are truly Gaussian. Generally any value of $\gamma$ between 1.0 and 2.0 is suggested.[12]

### *5. Solution of the Generalized Maximum Likelihood Estimator*

This section discusses a numerical technique of solving the implicit equation resulting from application of the generalized maximum likelihood method. The solution of Eq. (14) is typically determined in one of two ways: a Newton–Raphson iteration, or the iteratively reweighted algorithm, which is commonly attributed to Beaton and Tukey.[36] In this paper, the latter method will be used. The algorithm can be derived as follows. Eq. (14) can be expanded to yield $\mathbf{H}^T\mathbf{\Psi}\mathbf{H}\mathbf{x} = \mathbf{H}^T\mathbf{\Psi}\mathbf{y}$, which can be solved for $\mathbf{x}$ to give $\mathbf{x} = \left(\mathbf{H}^T\mathbf{\Psi}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{\Psi}\mathbf{y}$. Since the matrix $\mathbf{\Psi}$ depends on the residuals $\zeta_i$, and hence on $\mathbf{x}$, an iterative solution to Eq. (14) is expressed as

$$\mathbf{x}^{(j+1)} = \left(\mathbf{H}^T\mathbf{\Psi}^{(j)}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{\Psi}^{(j)}\mathbf{y} \tag{25}$$

where the superscript $(j)$ refers to the iteration index. The method can be initialized by using the least–squares solution $\mathbf{x}^{(0)} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{y}$. The converged value from the iterative procedure is taken as the state estimate, $\hat{\mathbf{x}}$.

When the measurement errors are correlated, they can be transformed to uncorrelated measurements with unit variance by means of the transformations given in Sec. A. The resulting solution by means of the iteratively re-

weighted algorithm takes the form

$$\mathbf{x}^{(j+1)} = \left(\mathbf{H}^T\mathbf{R}^{-1/2}\mathbf{\Psi}^{(j)}\mathbf{R}^{-1/2}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1/2}\mathbf{\Psi}^{(j)}\mathbf{R}^{-1/2}\mathbf{y} \tag{26}$$

The algorithm can be iterated until convergence or can be carried out through only one fixed iteration step, as discussed in Refs. 11, 37 and 38. It is shown in Refs. 11, 37 and 38 that the one-step Huber estimates capture all the desired robustness properties of the fully iterated estimator. Note that as $\gamma \to \infty$, the Huber filtering problem reduces to the least squares estimator. Specifically, when $\gamma \to \infty$, the matrix $\mathbf{\Psi} \to \mathbf{I}$, and Eq. (14) can be solved exactly in one iteration step and is equal to the least squares solution.

## III. Application to Dynamic Estimation

This section discusses the problem of estimating the state of the system of ordinary differential equations

$$\dot{\mathbf{x}} = \mathbf{f}\left(\mathbf{x}, \mathbf{u}, \mathbf{v}, t\right) \tag{27}$$

where $\mathbf{x}$ is the state vector, $\mathbf{u}$ are the deterministic inputs to the system, and $\mathbf{v}$ are random inputs to the system. The mean value of $\mathbf{v}$ is $\bar{\mathbf{v}} = \mathbf{0}$ and its covariance is $\mathbf{Q}$. It is assumed that the state of the system can be measured at discrete times in the form of a model given as

$$\mathbf{y}_k = \mathbf{h}\left(\mathbf{x}_k\right) + \mathbf{w}_k \tag{28}$$

where the subscript $k$ refers to the value of the parameter at time $t_k$, $\mathbf{y}_k$ is the measurement at time $t_k$, and $\mathbf{w}_k$ is the measurement noise at time $t_k$. The mean value of $\mathbf{w}_k$ is $\bar{\mathbf{w}}_k = \mathbf{0}$ and its covariance is $\mathbf{R}_k$.

### A. Robust Extended Kalman Filter: Linear Filtering as a Sequence of Linear Regression Problems

The Kalman filter is a well-known technique for estimating the state of systems of differential equations described in the form provided in Eqs. (27–28). The filter is a predictor–corrector approach in which the state predictions are computed by numerical integration of the dynamic model, and state corrections are obtained by a weighted linear combination of the predicted measurements and the actual measurements. In this approach, the state and covariance predictions are given as

$$\bar{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + \int_{t_{k-1}}^{t_k} \mathbf{f}\left(\mathbf{x}, \mathbf{u}, \bar{\mathbf{v}}, t\right) \mathrm{d}t \tag{29}$$

$$\bar{\mathbf{P}}_k = \hat{\mathbf{P}}_{k-1} + \int_{t_{k-1}}^{t_k} \left[\mathbf{A}(\hat{\mathbf{x}}(t), t)\bar{\mathbf{P}}(t) + \bar{\mathbf{P}}(t)^T\mathbf{A}(\hat{\mathbf{x}}(t), t)^T + \mathbf{B}(\hat{\mathbf{x}}(t), t)\mathbf{Q}(t)\mathbf{B}(\hat{\mathbf{x}}(t), t)^T\right] \mathrm{d}t \tag{30}$$

where $\bar{\mathbf{x}}_k$ is the predicted value of the state at time $t_k$, based on the estimated value of the state at time $t_{k-1}$, which is $\hat{\mathbf{x}}_{k-1}$. Similarly, $\bar{\mathbf{P}}_k$ is the predicted state error covariance matrix at time $t_k$ and $\hat{\mathbf{P}}_{k-1}$ is the estimated state error covariance matrix and time $t_{k-1}$. Also, the matrices $\mathbf{A}$ and $\mathbf{B}$ are given by

$$\mathbf{A}(t) = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\bar{\mathbf{x}}(t), \mathbf{v}=\bar{\mathbf{v}}} \tag{31}$$

$$\mathbf{B}(t) = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\right|_{\mathbf{x}=\bar{\mathbf{x}}(t), \mathbf{v}=\bar{\mathbf{v}}} \tag{32}$$

The state correction obtained at the time of the measurement update can be expressed as a linear regression problem between the predicted state and the observed quantity.[16] If the true value of the state is written as $\mathbf{x}_k$ and the state prediction error is written as $\boldsymbol{\delta}_k = \mathbf{x}_k - \bar{\mathbf{x}}_k$, then the state prediction can be expressed as $\bar{\mathbf{x}}_k = \mathbf{x}_k - \boldsymbol{\delta}_k$. By approximating the measurement equation as

$$\mathbf{y}_k \approx \mathbf{h}\left(\bar{\mathbf{x}}_k\right) + \mathbf{H}_k\left(\mathbf{x}_k - \bar{\mathbf{x}}_k\right) \tag{33}$$

the regression problem then takes the form

$$\left\{ \begin{array}{c} \mathbf{y}_k - \mathbf{h}\left(\bar{\mathbf{x}}_k\right) + \mathbf{H}_k\bar{\mathbf{x}}_k \\ \bar{\mathbf{x}}_k \end{array} \right\} = \left[ \begin{array}{c} \mathbf{H}_k \\ \mathbf{I} \end{array} \right] \mathbf{x}_k + \left\{ \begin{array}{c} \mathbf{w}_k \\ -\boldsymbol{\delta}_k \end{array} \right\} \tag{34}$$

where the matrix $\mathbf{H}_k$ is defined as

$$\mathbf{H}_k = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\bar{\mathbf{x}}_k, \mathbf{w}=\bar{\mathbf{w}}} \tag{35}$$

By defining the quantities

$$\mathbf{T}_k = \begin{bmatrix} \mathbf{R}_k & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{P}}_k \end{bmatrix} \tag{36}$$

$$\mathbf{z}_k = \mathbf{T}_k^{-1/2} \begin{Bmatrix} \mathbf{y}_k - \mathbf{h}\left(\bar{\mathbf{x}}_k\right) + \mathbf{H}_k \bar{\mathbf{x}}_k \\ \bar{\mathbf{x}}_k \end{Bmatrix} \tag{37}$$

$$\mathbf{G}_k = \mathbf{T}_k^{-1/2} \begin{bmatrix} \mathbf{H}_k \\ \mathbf{I} \end{bmatrix} \tag{38}$$

$$\boldsymbol{\xi}_k = \mathbf{T}_k^{-1/2} \begin{Bmatrix} \mathbf{w}_k \\ -\boldsymbol{\delta}_k \end{Bmatrix} \tag{39}$$

the linear regression problem is transformed to

$$\mathbf{z}_k = \mathbf{G}_k \mathbf{x}_k + \boldsymbol{\xi}_k \tag{40}$$

Note that Eq. (40) is in precisely the same form as the linear regression problem given in Eq.(11), which can be solved using the robust Huber generalized maximum–likelihood technique. Recall from the previous section that the solution of the Huber estimation problem can be written as the converged solution of the iterative equation

$$\mathbf{x}_k^{(j+1)} = \left(\mathbf{G}_k^T \boldsymbol{\Psi}^{(j)} \mathbf{G}_k\right)^{-1} \mathbf{G}_k^T \boldsymbol{\Psi}^{(j)} \mathbf{z}_k \tag{41}$$

where the superscript $(j)$ refers to the iteration index. The method can be initialized by using the least–squares (Kalman filter) solution $\mathbf{x}_k^{(0)} = \left(\mathbf{G}_k^T \mathbf{G}_k\right)^{-1} \mathbf{G}_k^T \mathbf{z}_k$. The converged value from the iterative procedure is taken as the corrected state estimate following a measurement update, $\hat{\mathbf{x}}_k$.

Finally, the estimated state estimate error covariance matrix following the measurement update is computed from

$$\hat{\mathbf{P}}_k = \left(\mathbf{G}_k^T \boldsymbol{\Psi} \mathbf{G}_k\right)^{-1} \tag{42}$$

using the final value of $\boldsymbol{\Psi}$ corresponding to the converged state estimate.

Due to the particular structure of the matrix $\mathbf{G}_k$, the discrete time dynamic state estimation technique can be simplified considerably from the static state estimation technique by application of the matrix inversion lemma.[39] 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

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_y & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_x \end{bmatrix} \tag{43}$$

and by defining the gain matrix

$$\mathbf{K}_k = \bar{\mathbf{P}}_k^{1/2} \boldsymbol{\Psi}_x^{-1} \bar{\mathbf{P}}_k^{1/2} \mathbf{H}_k^T \left(\mathbf{H}_k \bar{\mathbf{P}}_k^{1/2} \boldsymbol{\Psi}_x^{-1} \bar{\mathbf{P}}_k^{1/2} \mathbf{H}_k^T + \mathbf{R}_k^{1/2} \boldsymbol{\Psi}_y^{-1} \mathbf{R}_k^{1/2}\right)^{-1} \tag{44}$$

then the state estimate and covariance corrections can be written as

$$\hat{\mathbf{x}}_k^{(j+1)} = \bar{\mathbf{x}}_k + \mathbf{K}_k^{(j)} \left[\mathbf{y}_k - \mathbf{h}\left(\bar{\mathbf{x}_k}\right)\right] \tag{45}$$

$$\hat{\mathbf{P}}_k = \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right) \bar{\mathbf{P}}_k^{1/2} \boldsymbol{\Psi}_x^{-1} \bar{\mathbf{P}}_k^{1/2} \tag{46}$$

Again, 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.

### B. Robust Divided Difference Filtering

The Divided Difference Filter is one of several new estimation techniques that are collectively known as Sigma–Point Kalman Filters (SPKF). Like the basic Kalman filter, the SPKFs seek to determine a state estimate that minimizes the $\ell_2$–norm of the residuals. The SPKF technique differs from the basic Kalman filter in that the estimate and

covariance propagation prior to determining the state estimate at the time of the measurement update. The divided difference filter arises from an alternate approach to the nonlinear state estimation and filtering problem than the EKF. Whereas the EKF is based on first–order Taylor series approximations, the divided difference filter relies on multidimensional interpolation formulas to approximate the nonlinear transformations. As a result of this approach, the filter does not require knowledge or existence of the partial derivatives of the system dynamics and measurement equations.

The First–Order (DD1) and Second–Order (DD2) Divided Difference Filters[40, 41] are generalizations of the filter introduced by Schei,[42] and are two examples of SPKF–class estimators; other examples can be found in Refs. 43–45. This section describes the DD1 and DD2 filters and shows how the measurement update equation can be modified to provide robustness against deviations from the assumed Gaussian error model by use of the Huber technique. The filter summary closely follows that given in Refs. 40 and 41. The filter equations rely upon a discrete representation of the system dynamics, in which the differential equation in Eq. (27) is replaced with a difference equation of the form

$$\mathbf{x}_{k+1} = \mathbf{F}\left(\mathbf{x}_k, \mathbf{v}_k, t_k\right) \tag{47}$$

and the measurement equation is given by

$$\mathbf{y}_k = \mathbf{G}\left(\mathbf{x}_k, \mathbf{w}_k, t_k\right) \tag{48}$$

Note that the transformation between the continuous representation and the discrete representation of the system dynamics is essentially an integration between time steps of the measurement data, so that

$$\mathbf{F}\left(\mathbf{x}_k, \mathbf{v}_k, t_k\right) = \mathbf{f}\left(\mathbf{x}_k, \mathbf{v}_k, t_k\right) + \int_{t_k}^{t_{k+1}} \mathbf{f}\left(\mathbf{x}(t), \mathbf{v}(t), t\right) \mathrm{d}t \tag{49}$$

The following square-root decompositions of the covariance, process noise, and measurement noise matrices are defined as

$$\hat{\mathbf{P}}_k = \hat{\mathbf{S}}_{x_k} \hat{\mathbf{S}}_{x_k}^T \tag{50}$$

$$\bar{\mathbf{P}}_k = \bar{\mathbf{S}}_{x_k} \bar{\mathbf{S}}_{x_k}^T \tag{51}$$

$$\mathbf{Q}_k = \mathbf{S}_{v_k} \mathbf{S}_{v_k}^T \tag{52}$$

$$\mathbf{R}_k = \mathbf{S}_{w_k} \mathbf{S}_{w_k}^T \tag{53}$$

Also, the $j$th column of $\bar{\mathbf{S}}_{x_k}$ shall be referred to as $\bar{\mathbf{s}}_{x_{k_j}}$ and likewise for the other matrices.

### 1. *Overview of the DD1 Filter*

The DD1 filter makes use of first–order divided differences to approximate the system and measurement dynamics rather than the first–order Taylor series expansions used in the EKF. The following matrices of first–order divided differences are defined as

$$\mathbf{S}'_{x\hat{x}_{k_{i,j}}} = \frac{1}{2c}\left[\mathbf{F}_i\left(\hat{\mathbf{x}} + c\hat{\mathbf{s}}_{x_j}, \bar{\mathbf{v}}_k, t_k\right) - \mathbf{F}_i\left(\hat{\mathbf{x}}_k - c\hat{\mathbf{s}}_{x_j}, \bar{\mathbf{v}}_k, t_k\right)\right] \tag{54}$$

$$\mathbf{S}'_{xv_{k_{i,j}}} = \frac{1}{2c}\left[\mathbf{F}_i\left(\hat{\mathbf{x}}_k, \bar{\mathbf{v}}_k + c\mathbf{s}_{v_j}, t_k\right) - \mathbf{F}_i\left(\hat{\mathbf{x}}_k, \bar{\mathbf{v}}_k - c\mathbf{s}_{v_j}, t_k\right)\right] \tag{55}$$

$$\mathbf{S}'_{y\bar{x}_{k_{i,j}}} = \frac{1}{2c}\left[\mathbf{G}_i\left(\bar{\mathbf{x}}_k + c\bar{\mathbf{s}}_{x_j}, \bar{\mathbf{w}}_k, t_k\right) - \mathbf{G}_i\left(\bar{\mathbf{x}}_k - c\bar{\mathbf{s}}_{x_j}, \bar{\mathbf{w}}_k, t_k\right)\right] \tag{56}$$

$$\mathbf{S}'_{yw_{k_{i,j}}} = \frac{1}{2c}\left[\mathbf{G}_i\left(\bar{\mathbf{x}}_k, \bar{\mathbf{w}}_k + c\mathbf{s}_{w_j}, t_k\right) - \mathbf{G}_i\left(\bar{\mathbf{x}}_k, \bar{\mathbf{w}}_k - c\mathbf{s}_{w_j}, t_k\right)\right] \tag{57}$$

where $c$ the divided–difference perturbing parameter.

The state, state root–covariance, measurement, and measurement root-covariance predictions are given by

$$\bar{\mathbf{x}}_{k+1} = \mathbf{F}\left(\hat{\mathbf{x}}_k, \bar{\mathbf{v}}_k, t_k\right) \tag{58}$$

$$\bar{\mathbf{S}}_{x_{k+1}} = \mathcal{H}\left(\begin{bmatrix} \mathbf{S}'_{x\hat{x}_k} & \mathbf{S}'_{xv_k} \end{bmatrix}\right) \tag{59}$$

$$\bar{\mathbf{y}}_k = \mathbf{G}\left(\bar{\mathbf{x}}_k, \bar{\mathbf{w}}_k, t_k\right) \tag{60}$$

$$\mathbf{S}_{y_k} = \mathcal{H}\left(\begin{bmatrix} \mathbf{S}'_{y\bar{x}_k} & \mathbf{S}'_{yw_k} \end{bmatrix}\right) \tag{61}$$

where $\mathcal{H}(\cdot)$ represents a Householder transformation of the argument matrix.[40, 41]

The state and root-covariance measurement update equations are given by

$$\hat{\mathbf{x}}_k = \bar{\mathbf{x}}_k + \mathbf{K}_k\left(\mathbf{y}_k - \bar{\mathbf{y}}_k\right) \tag{62}$$

$$\hat{\mathbf{S}}_{x_k} = \mathcal{H}\left(\begin{bmatrix} \bar{\mathbf{S}}_{x_k} - \mathbf{K}_k\mathbf{S}'_{yx_k} & \mathbf{K}_k\mathbf{S}'_{yw_k} \end{bmatrix}\right) \tag{63}$$

where $\mathbf{K}_k = \bar{\mathbf{S}}_{x_k}\mathbf{S}'^T_{y\bar{x}_k}\left(\mathbf{S}_{y_k}\mathbf{S}^T_{y_k}\right)^{-1}$ is the Kalman gain matrix.

*2. Overview of the DD2 Filter*

The DD2 filter makes use of second–order divided differences to approximate nonlinear transformation of the state and covariance. The matrices of second–order divided differences are defined as

$$\mathbf{S}''_{x\hat{x}_{k_{i,j}}} = \frac{\sqrt{c^2-1}}{2c^2}\left[\mathbf{F}_i\left(\hat{\mathbf{x}}_k + c\hat{\mathbf{s}}_{x_j}, \bar{\mathbf{v}}_k, t_k\right) + \mathbf{F}_i\left(\hat{\mathbf{x}}_k - c\hat{\mathbf{s}}_{x_j}, \bar{\mathbf{v}}_k, t_k\right) - 2\mathbf{F}_i\left(\hat{\mathbf{x}}_k, \bar{\mathbf{v}}_k, t_k\right)\right] \tag{64}$$

$$\mathbf{S}''_{xv_{k_{i,j}}} = \frac{\sqrt{c^2-1}}{2c^2}\left[\mathbf{F}_i\left(\hat{\mathbf{x}}_k, \bar{\mathbf{v}}_k + c\mathbf{s}_{w_j}, t_k\right) + \mathbf{F}_i\left(\hat{\mathbf{x}}_k, \bar{\mathbf{v}}_k - c\mathbf{s}_{w_j}, t_k\right) - 2\mathbf{F}_i\left(\hat{\mathbf{x}}_k, \bar{\mathbf{v}}_k, t_k\right)\right] \tag{65}$$

$$\mathbf{S}''_{y\bar{x}_{k_{i,j}}} = \frac{\sqrt{c^2-1}}{2c^2}\left[\mathbf{G}_i\left(\bar{\mathbf{x}} + c\bar{\mathbf{s}}_{x_j}, \bar{\mathbf{w}}_k, t_k\right) + \mathbf{G}_i\left(\bar{\mathbf{x}} - c\bar{\mathbf{s}}_{x_j}, \bar{\mathbf{w}}_k, t_k\right) - 2\mathbf{G}_i\left(\bar{\mathbf{x}}_k, \bar{\mathbf{w}}_k, t_k\right)\right] \tag{66}$$

$$\mathbf{S}''_{yw_{k_{i,j}}} = \frac{\sqrt{c^2-1}}{2c^2}\left[\mathbf{G}_i\left(\bar{\mathbf{x}}_k, \bar{\mathbf{w}}_k + c\mathbf{s}_{w_j}, t_k\right) + \mathbf{G}_i\left(\bar{\mathbf{x}}_k\bar{\mathbf{w}}_k - c\mathbf{s}_{w_j}, t_k\right) - 2\mathbf{G}_i\left(\bar{\mathbf{x}}_k, \bar{\mathbf{w}}_k, t_k\right)\right] \tag{67}$$

The state, state root–covariance, measurement, and measurement covariance predictions are given by

$$\begin{aligned}\bar{\mathbf{x}}_{k+1} = {} & \left(\frac{c^2 - n_x - n_v}{c^2}\right)\mathbf{F}\left(\hat{\mathbf{x}}_k, \bar{\mathbf{v}}_k, t_k\right) \\ & + \frac{1}{2c^2}\sum_{j=1}^{n_x}\left[\mathbf{F}\left(\hat{\mathbf{x}}_k + c\hat{\mathbf{s}}_{x_{k_j}}, \bar{\mathbf{v}}_k, t_k\right) + \mathbf{F}\left(\hat{\mathbf{x}}_k - c\hat{\mathbf{s}}_{x_{k_j}}, \bar{\mathbf{v}}_k, t_k\right)\right] \\ & + \frac{1}{2c^2}\sum_{j=1}^{n_v}\left[\mathbf{F}\left(\hat{\mathbf{x}}_k, \bar{\mathbf{v}}_k + c\mathbf{s}_{v_{k_j}}, t_k\right) + \mathbf{F}\left(\hat{\mathbf{x}}_k, \bar{\mathbf{v}}_k - c\mathbf{s}_{v_{k_j}}, t_k\right)\right]\end{aligned} \tag{68}$$

$$\bar{\mathbf{S}}_{x_{k+1}} = \mathcal{H}\left(\begin{bmatrix} \mathbf{S}'_{x\hat{x}_k} & \mathbf{S}'_{xv_k} & \mathbf{S}''_{x\hat{x}_k} & \mathbf{S}''_{xv_k} \end{bmatrix}\right) \tag{69}$$

$$\begin{aligned}\bar{\mathbf{y}}_k = {} & \left(\frac{c^2 - n_x - n_w}{c^2}\right)\mathbf{G}\left(\bar{\mathbf{x}}_k, \bar{\mathbf{w}}_k, t_k\right) \\ & + \frac{1}{2c^2}\sum_{j=1}^{n_x}\left[\mathbf{G}\left(\bar{\mathbf{x}}_k + c\bar{\mathbf{s}}_{x_{k_j}}, \bar{\mathbf{w}}_k, t_k\right) + \mathbf{G}\left(\bar{\mathbf{x}}_k - c\bar{\mathbf{s}}_{x_{k_j}}, \bar{\mathbf{w}}_k, t_k\right)\right] \\ & + \frac{1}{2c^2}\sum_{j=1}^{n_w}\left[\mathbf{G}\left(\bar{\mathbf{x}}_k, \bar{\mathbf{w}}_k + c\mathbf{s}_{w_{k_j}}, t_k\right) + \mathbf{G}\left(\bar{\mathbf{x}}_k, \bar{\mathbf{w}}_k - c\mathbf{s}_{w_{k_j}}, t_k\right)\right]\end{aligned} \tag{70}$$

$$\mathbf{S}_{y_k} = \mathcal{H}\left(\begin{bmatrix} \mathbf{S}'_{y\bar{x}_k} & \mathbf{S}'_{yw_k} & \mathbf{S}''_{y\bar{x}_k} & \mathbf{S}''_{yw_k} \end{bmatrix}\right) \tag{71}$$

where $n_x$ is the size of the state dimension, $n_v$ is the size of the process noise dimension, and $n_w$ is the size of the measurement noise dimension.

Lastly, the state and root-covariance update equations are given by

$$\hat{\mathbf{x}}_k = \bar{\mathbf{x}}_k + \mathbf{K}_k\left(\mathbf{y}_k - \bar{\mathbf{y}}_k\right) \tag{72}$$

$$\hat{\mathbf{S}}_{x_k} = \mathcal{H}\left(\begin{bmatrix} \bar{\mathbf{S}}_{x_k} - \mathbf{K}_k\mathbf{S}'_{yx_k} & \mathbf{K}_k\mathbf{S}'_{yw_k} & \mathbf{K}_k\mathbf{S}'_{yx_k} & \mathbf{K}_k\mathbf{S}'_{yw_k} \end{bmatrix}\right) \tag{73}$$

where $\mathbf{K}_k = \bar{\mathbf{S}}_{x_k}\mathbf{S}'^T_{y\bar{x}_k}\left(\mathbf{S}_{y_k}\mathbf{S}^T_{y_k}\right)^{-1}$ is the Kalman gain matrix.

Note that many of the same state and noise perturbations used to calculate the first–order divided differences are again used to compute the second–order divided differences. This point has important implications with regard to the computational costs, suggesting that the DD2 filter may not require a great deal more computing time than the DD1 filter.

*3. Modification of Measurement Update Equations*

The DD1 filter measurement update can be modified to yield a robust estimator in a manner similar to that shown for the Kalman filter in the previous section. If the measurement update given in Eq. (62) is taken as the initial guess for the state, similar to how the Kalman filter solution utilized in the previous section, then a one-step Huber update can be written as

$$\mathbf{S}_{y_k}^{(1)} = \mathcal{H}\left(\left[\begin{array}{cc} \mathbf{S}'_{y\bar{x}_k}\boldsymbol{\Psi}_x^{-1/2} & \mathbf{S}'_{yw_k}\boldsymbol{\Psi}_y^{-1/2} \end{array}\right]\right) \tag{74}$$

$$\mathbf{K}_k^{(1)} = \bar{\mathbf{S}}_{x_k}\boldsymbol{\Psi}_x^{-1}\mathbf{S}'^{T}_{y\bar{x}_k}\left(\mathbf{S}_{y_k}^{(1)}\mathbf{S}_{y_k}^{(1)^T}\right)^{-1} \tag{75}$$

$$\hat{\mathbf{x}}_k = \bar{\mathbf{x}}_k + \mathbf{K}_k^{(1)}\left(\mathbf{y}_k - \bar{\mathbf{y}}_k\right) \tag{76}$$

$$\hat{\mathbf{S}}_{x_k} = \mathcal{H}\left(\left[\begin{array}{cc} \bar{\mathbf{S}}_{x_k}\boldsymbol{\Psi}_x^{-1/2} - \mathbf{K}_k^{(1)}\mathbf{S}'_{yx_k}\boldsymbol{\Psi}_x^{-1/2} & \mathbf{K}_k^{(1)}\mathbf{S}'_{yw_k}\boldsymbol{\Psi}_y^{-1/2} \end{array}\right]\right) \tag{77}$$

Similarly for the DD2 filter, if the measurement update given in Eq. (72) is taken as the initial guess for the state, then a one-step Huber update can be written as

$$\mathbf{S}_{y_k}^{(1)} = \mathcal{H}\left(\left[\begin{array}{cccc} \mathbf{S}'_{y\bar{x}_k}\boldsymbol{\Psi}_x^{-1/2} & \mathbf{S}'_{yw_k}\boldsymbol{\Psi}_y^{-1/2} & \mathbf{S}''_{y\bar{x}_k}\boldsymbol{\Psi}_x^{-1/2} & \mathbf{S}''_{yw_k}\boldsymbol{\Psi}_y^{-1/2} \end{array}\right]\right) \tag{78}$$

$$\mathbf{K}_k^{(1)} = \bar{\mathbf{S}}_{x_k}\boldsymbol{\Psi}_x^{-1}\mathbf{S}'^{T}_{y\bar{x}_k}\left(\mathbf{S}_{y_k}^{(1)}\mathbf{S}_{y_k}^{(1)^T}\right)^{-1} \tag{79}$$

$$\hat{\mathbf{x}}_k = \bar{\mathbf{x}}_k + \mathbf{K}_k^{(1)}\left(\mathbf{y}_k - \bar{\mathbf{y}}_k\right) \tag{80}$$

$$\hat{\mathbf{S}}_{x_k} = \mathcal{H}\left(\left[\begin{array}{cccc} \bar{\mathbf{S}}_{x_k}\boldsymbol{\Psi}_x^{-1/2} - \mathbf{K}_k^{(1)}\mathbf{S}'_{yx_k}\boldsymbol{\Psi}_x^{-1/2} & \mathbf{K}_k^{(1)}\mathbf{S}'_{yw_k}\boldsymbol{\Psi}_y^{-1/2} & \mathbf{K}_k^{(1)}\mathbf{S}''_{yx_k}\boldsymbol{\Psi}_x^{-1/2} & \mathbf{K}_k^{(1)}\mathbf{S}''_{yw_k}\boldsymbol{\Psi}_y^{-1/2} \end{array}\right]\right) \tag{81}$$

In each case, $\boldsymbol{\Psi}_x$ and $\boldsymbol{\Psi}_y$ are diagonal matrices computed from the Huber $\psi$ function, with residuals that take the form

$$\boldsymbol{\zeta} = \left[\begin{array}{cc} \mathbf{S}_{w_k}^{-1} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{S}}_{x_k}^{-1} \end{array}\right] \cdot \left\{\begin{array}{c} \mathbf{y}_k - \bar{\mathbf{y}}_k \\ \\ \hat{\mathbf{x}}_k^{(0)} - \bar{\mathbf{x}}_k \end{array}\right\} \tag{82}$$

where the superscript (0) refers to the initial state estimate computed from the standard DD1 or DD2 update.

## IV. Application to a Benchmark Nonlinear Filtering Problem

This section discusses the application of the robust filter to the problem of reconstructing the trajectory of a target using range measurements recorded from a radar ground station. The example problem is to estimate the trajectory of a mass falling through an exponential atmosphere with a constant, yet unknown, drag coefficient. The gravitational acceleration acting on the body is neglected in the dynamic model. This truncated model is valid for high initial velocities that cause the aerodynamic acceleration to dominate over the gravitational acceleration. Measurements of the vehicle state are provided by a radar tracking station. This benchmark nonlinear filtering problem was initially studied in Ref. 46 and has been repeated numerous times. Fig. 5 shows the geometry of the problem.

### A. Dynamic Model and Measurement Equations

The dynamic model for this problem can be derived by writing

$$\dot{x}_1 = -x_2 \tag{83}$$

$$\dot{x}_2 = -\frac{C_D A \rho}{2m} x_2^2 \tag{84}$$

where $x_1$ represents the altitude of the mass, $x_2$ its downward velocity, $C_D$ is the drag coefficient, $A$ is the cross-sectional area, $m$ is the mass, and $\rho$ is the atmospheric density, which is assumed to follow an exponential function of the form

$$\rho = \rho_0 \mathrm{e}^{-\eta x_1} \tag{85}$$

where the quantity $\eta$ is the known constant inverse density scale height. If the aerodynamics of the target are considered to be an unknown parameter to be estimated in real time, then a third state variable can be expressed as

$$x_3 = \sqrt{\frac{C_D A \rho_0}{2m}} \tag{86}$$

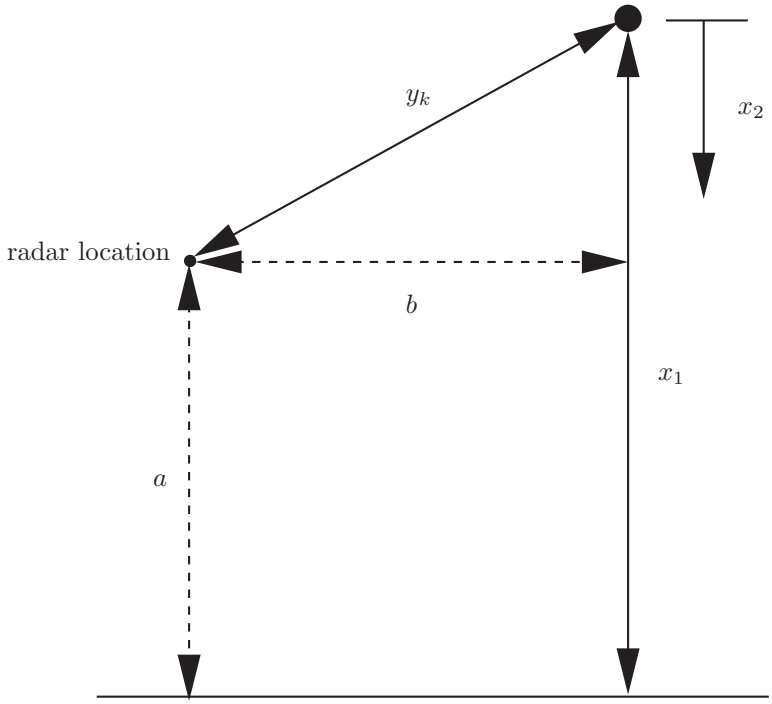



**Figure 5. Geometry of Example Problem**

Here, the square-root of $C_D A\rho_0/2m$ is used since this term must always be positive, in other words since the quantity $x_3^2$ is always non-negative.

In summary, the complete dynamic model for this problem is expressed as

$$\dot{x}_1 = -x_2 \tag{87}$$

$$\dot{x}_2 = -x_3^2 x_2^2 \mathrm{e}^{-\eta x_1} \tag{88}$$

$$\dot{x}_3 = 0 \tag{89}$$

The radar measurement equation is

$$y_k = \sqrt{b^2 + \left[x_1\left(t_k\right) - a\right]^2} + w_k \tag{90}$$

where $w_k$ represents zero-mean random error, with probability density function $f(w_k)$.

The model parameters and initial conditions for the problem are summarized in Table 1. The true trajectory for these initial conditions is shown in Fig. 6.

**Table 1. Simulation Parameters**

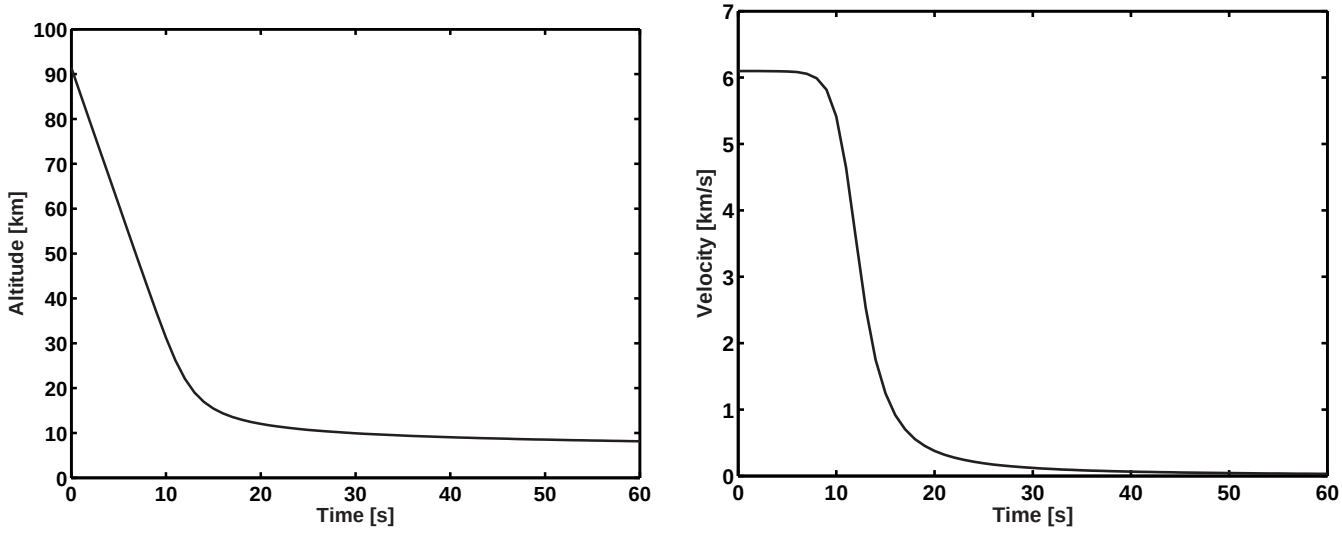
| Parameter | Value |
|---|---|
| $a$, km | 30.5 |
| $b$, km | 30.5 |
| $\eta$, $\mathrm{m}^{-1}$ | $1.64 \cdot 10^{-4}$ |
| $\sigma_1$, m | 30.5 |
| $\gamma$ | 1.345 |
| $c^2$ | 3.0 |

Random measurement errors are drawn from the mixture of zero–mean Gaussian probability distributions, defined by the probability density function

$$f\left(w_k\right) = \left(\frac{1-\epsilon}{\sigma_1\sqrt{2\pi}}\right)\exp\left[-\left(\frac{w_k^2}{2\sigma_1^2}\right)\right] + \left(\frac{\epsilon}{\sigma_2\sqrt{2\pi}}\right)\exp\left[-\left(\frac{w_k^2}{2\sigma_2^2}\right)\right] \tag{91}$$

**Table 2. Initial Conditions**

| Initial State | True Value | Estimated Value | Standard Deviation |
|---|---|---|---|
| $x_1(0)$, km | 91.5 | 91.5 | 0.31 |
| $x_2(0)$, km/s | 6.1 | 6.1 | 0.06 |
| $x_3(0)$, $1/\sqrt{\text{m}}$ | 0.06 | 0.01 | 0.02 |




(a) True Position

(b) True Velocity

**Figure 6. True Trajectory of Target**

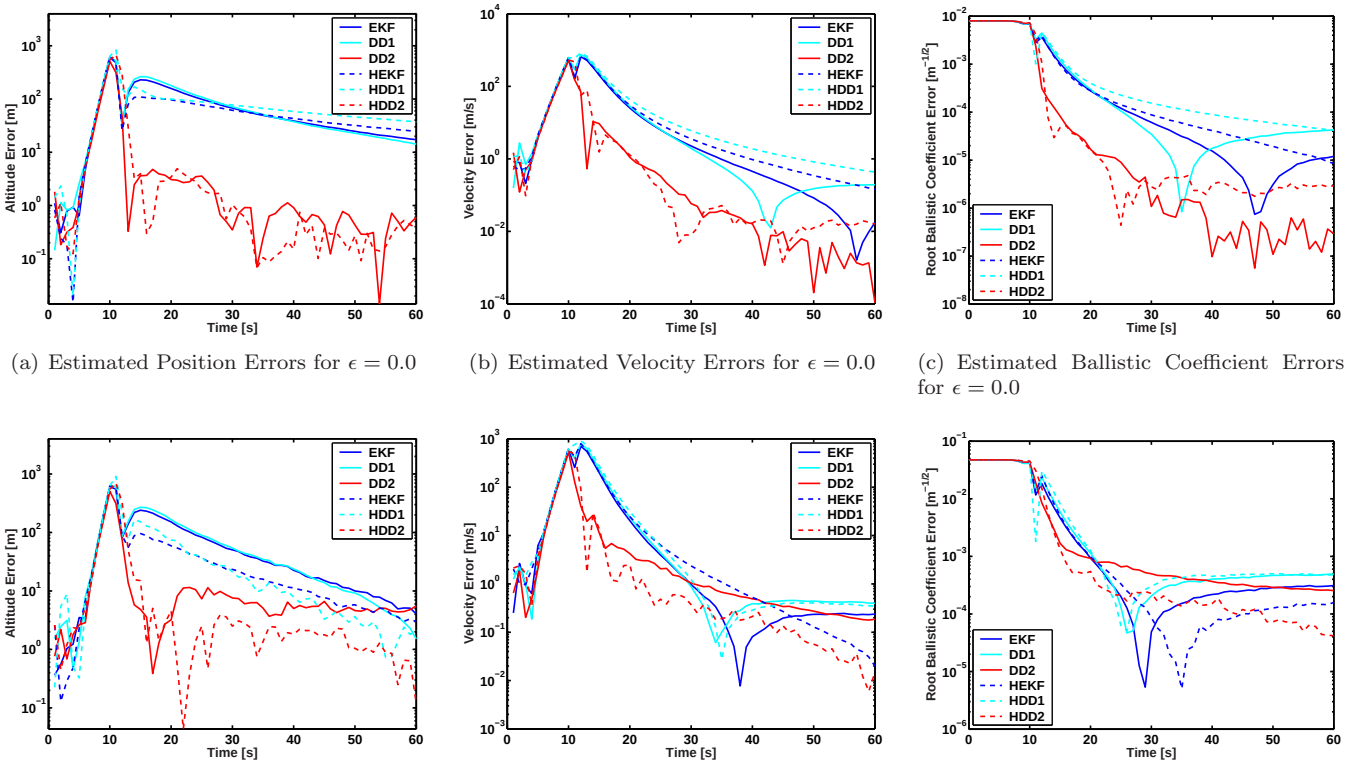
where $\sigma_1$ and $\sigma_2$ are the standard deviations of the individual Gaussian distributions, and $\epsilon$ is a perturbing parameter that represents error model contamination. The standard deviations $\sigma_1$ are chosen according to Table 1 and $\sigma_2$ is chosen as $\sigma_2 = 5\sigma_1$. The measurements are assumed to occur at a frequency of 1 Hz.

## B. Results of Example Problem

This section discusses the results of applying several filters to the benchmark tracking problem. These filters include the EKF, DD1, DD2 and the robust versions of each with the one–step Huber update. The results of a Monte–Carlo simulation are shown in the following figures. In this simulation, 2000 trial cases have been conducted, each case terminating after an elapsed time of 60 s.

The absolute value of the median errors are shown for the case of $\epsilon = 0$ in Figs. 7(a)–(c). In this case the DD2 filter gives a smaller estimation error than the other robust and non-robust filters. The DD1 filter exhibits slightly larger errors than the EKF. These results are not surprising based on the results given in Ref. 40 and 41. The robust filters do not perform as well as their non-robust counterparts in this case, because the Huber update does not minimize the $\ell_2$ norm during the measurement update. The increase in the estimation error for the robust filters is to be expected in a perfectly Gaussian simulation since the minimum $\ell_2$ norm is the maximum likelihood estimator in this case.

The absolute value of the median errors are shown in Figs. 7(d)–(f) for the case $\epsilon = 0.5$, for the case where the measurement errors are highly non–Gaussian. In this case, the Huber–EKF and DD2 filters give comparable results to each other for the position and velocity errors, but the Huber–EKF gives a smaller error in the estimate of the ballistic parameter. Both Huber–EKF and DD2 are superior to the Kalman filter in this case. The DD2 filter with the Huber update exhibits the smallest errors since it captures both nonlinearity and non-Gaussianity. The DD1 and Huber–DD1 filters do not perform as well as the EKF and the Huber–EKF, respectively, in the non–Gaussian case, which follows the behavior from the Gaussian case. In the non-Gaussian case, six cases of the EKF and two cases of the DD1 methods diverged completely, while the robust counterparts, H-EKF and H-DD1, did not exhibit divergence. Divergence problems with the Kalman filter have been noted previously.[47] In this case, the ballistic parameter initialization is poor, which is compounded by the presence of non-Gaussian errors. The DD2 filter, being




(a) Estimated Position Errors for $\epsilon = 0.0$

(b) Estimated Velocity Errors for $\epsilon = 0.0$

(c) Estimated Ballistic Coefficient Errors for $\epsilon = 0.0$

(d) Estimated Position Errors for $\epsilon = 0.5$

(e) Estimated Velocity Errors for $\epsilon = 0.5$

(f) Estimated Ballistic Coefficient Errors for $\epsilon = 0.5$

**Figure 7. Estimated Trajectory Errors**

a second–order filter and therefore not as sensitive to initialization, did not diverge in any of the Monte-Carlo cases, but clearly the median error is reduced by making use of the Huber update method.

**Table 3. EKF-Relative Computation Ratios**

| Filter | Computation Ratio |
|---|---|
| H-EKF | 1.08 |
| DD1 | 2.95 |
| H-DD1 | 3.14 |
| DD2 | 3.02 |
| H-DD2 | 3.19 |

The improvements found by use of the DD2 and Huber filtering techniques in favor of the Kalman filter do not come without a cost. The computational cost associated with implementing the each filter is summarized in Table 3. In this table, the computational costs are divided by the EKF processing time to provide a ratio of the cost relative to that associated with the EKF. These ratios are based on the median computation time for each filter during the Monte-Carlo simulation. The results show that the linear Huber–EKF filter requires the smallest relative computational cost whereas the DD2 and Huber–DD2 filters require a great deal more, at 3.12 and 3.15 times the total computation, respectively. It is not surprising that the Huber–DD2 filter has the largest cost since it has the smallest errors for the non-Gaussian case, but it is interesting to note that the similar levels of accuracy in the non-Gaussian case can be found by use of the Huber filter over that of DD2 filter, for only a fraction of the computation time.

## V. Conclusions

This paper has discussed a robust filtering technique based on Huber's generalized maximum likelihood estimation theory, which provides robustness against deviations from the Gaussian error distributions assumed by standard Kalman filtering techniques. The standard Kalman filtering problem was recast in the form of a sequence of linear regression problems, to be solved at each measurement point using the robust Huber technique. Next it was shown how the update equations of the first–order (DD1) and second–order divided difference filters (DD2), two particular cases of the class of sigma point Kalman filters, can be modified using the Huber technique to provide a robust update. The robust filters were applied to a benchmark tracking problem involving the estimation of the trajectory of an entry body from discrete–time, noisy range measurement data provided by a radar tracking system. The simulation was conducted using Monte–Carlo techniques and both Gaussian and non–Gaussian error distributions in order to asses the performance of the filtering techniques.

The results show that for perfectly Gaussian error distributions the standard DD2 filter exhibits the lowest estimation error time history, which was the expected outcome based on previously published results. The DD1 filter exhibited slightly larger estimation errors than that of the EKF. The filters with the Huber update technique produced larger errors than the standard update, since the standard form of the update is a the maximum likelihood estimate for the perfectly Gaussian case. However, for non–Gaussian error distributions, the modified filters with the Huber update outperformed the standard filters. The modified DD2 filter with the Huber update equation exhibited the smallest errors in the non-Gaussian numerical simulations conducted. The Huber-EKF and the Huber-DD1 filters were able to mitigate divergence problems in their non-robust counterparts.

Comparisons of the computational costs associated with each filter shows that the Huber–EKF filter is able to process data at a rate approximately three times faster than the standard DD2 filter, and produces similar accuracy levels in the non–Gaussian case. Therefore, for non-Gaussian cases, the Huber–EKF filter is superior to the standard DD2 filter. If computation costs are not a concern for the particular application, then the Huber–DD2 filter exhibits the best performance.

Future work in this area could be focused toward an adaptive Huber update, where the value of the perturbing parameter $\epsilon$ is estimated in real time from fixed number of stored residuals, using a fading memory approach. Then the optimal value of the tuning parameter $\gamma$ for the current estimated perturbing parameter $\epsilon$ could be used at each measurement update in order to optimize the filter performance.

## References


[1]Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering,* Vol. 82, No. 1, 1960, pp. 35–50.

[2]Sorenson, H. W., "Least–Squares Estimation: From Gauss to Kalman," *IEEE Spectrum*, Vol. 7, No. 7, 1970, pp. 63–68.

[3]Meinhold, R. J. and Singpurwalla, N. D., "Understanding the Kalman Filter," *American Statistician*, Vol. 37, No. 2, 1983, pp. 123–127.

[4]Tukey, J. W., "A Survey of Sampling from Contaminated Distributions," *Contributions to Probability and Statistics*, Olkin, I., Ghurye, S., Hoeffding, W., Madow, W., and Mann, H. (Eds.), Stanford University Press, Stanford, CA, 1960, pp. 448–485.

[5]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. 1045–1080.

[6]Box, G. E. P., "Non–Normality and Tests on Variances," *Biometrika*, Vol. 40, Nos. 3/4, 1953, pp. 318–335.

[7]Stigler, S. M., "Simon Newcomb, Percy Daniell, and the History of Robust Estimation 1885–1920," *Journal of the American Statistical Association*, Vol. 68, No. 344, 1973, pp. 872–879.

[8]Wu, W., "Maximum Likelihood Identification of Glint Noise," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 32, No. 1, 1996, pp. 41–51.

[9]Huber, P. J., *Robust Statistics*, Wiley, New York, 1981, pp. 43–106, 153–198.

[10]Hampel, F. R., Rousseeuw, P. J., Ronchetti, E., and Stahel, W. A., *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, 1986, pp. 78–186.

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

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

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

[14]Huber, P. J., "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *Annals of Statistics*, Vol. 1, No. 5, 1973, pp. 799–821.

[15]Huber, P. J., "Minimax Aspects of Bounded–Influence Regression," *Journal of the American Statistical Association*, Vol. 78, No. 381, 1983, pp. 66–72.

[16]Duncan, D. B. and Horn, S. D., "Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis," *Journal of the American Statistical Association*, Vol. 67, No. 340, 1972, pp. 815–821.

[17]Boncelet, C. G. and Dickinson, B. W., "An Approach to Robust Kalman Filtering," *22nd IEEE Conference on Decision and Control*, Institute of Electrical and Electronics Engineers, New York, NY, 1983, pp. 304–305.

[18]Kovacevic, B., Durovic, Z., and Glavaski, S., "On Robust Kalman Filtering," *International Journal of Control*, Vol. 56, No. 3, 1992, pp. 547–562.

[19]El-Hawary, F. and Jing, Y., "Robust Regression-Based EKF for Tracking Underwater Targets," *IEEE Journal of Oceanic Engineering*, Vol. 20, No. 1, 1995, pp. 31–41.

[20]Durgaprasad, G. and Thakur, S. S., "Robust Dynamic State Estimation of Power Systems Based on M-Estimation and Realistic Modeling of System Dynamics," *IEEE Transactions on Power Systems*, Vol. 13, No. 4, 1998, pp. 1331–1336.

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

[22]Jeffreys, H., "An Alternative to the Rejection of Observations," *Proceedings of the Royal Society of London, Series A*, Vol. 137, No. 831, 1932, pp. 78–87.

[23]Sorenson, H. W. and Alspach, D. L., "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica*, Vol. 7, No. 4, 1971, pp. 465–479.

[24]Alspach, D. L. and Sorenson, H. W., "Nonlinear Bayesian Estimation Using Gaussian Sum Approximations," *IEEE Transactions on Automatic Control*, Vol. 17, No. 4, 1972, pp. 439–448.

[25]Plantaniotis, K. N., Androutsos, D., and Venetsanopoulos, A. N., "Nonlinear Filtering of Non-Gaussian Noise," *Journal of Intelligent and Robotic Systems*, Vol. 19, No. 2, 1997, pp. 201–231.

[26]Tam, W. I., Plantaniotis, K. N., and Hatzinakos, D., "An Adaptive Gaussian Sum Algorithm for Radar Tracking," *Signal Processing*, Vol. 77, No. 1, 1999, pp. 85–104.

[27]Masreliez, C. J. and Martin, R. D., "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter," *IEEE Transactions on Automatic Control*, Vol. 22, No. 3, 1977, pp. 361–371.

[28]Aldrich, J., "R. A. Fisher and the Making of Maximum Likelihood, 1912-1922," *Statistical Science*, Vol. 12, No. 3, 1997, pp. 162–176.

[29]Kendall, M. G., "Daniel Bernoulli on Maximum Likelihood," *Biometrika*, Vol. 48, Nos. 1/2, 1961, pp. 1–18.

[30]Hald, A., "On the History of Maximum Likelihood in Relation to Inverse Probability and Least Squares," *Statistical Science*, Vol. 14, No. 2, 1999, pp. 214–222.

[31]Eisenhart, C., "Laws of Error II: The Gaussian Distribution," *Encyclopedia of Statistical Science*, Kotz, S. and Johnson, N. L., (Eds.), Wiley, New York, NY, 1983, pp. 547–566.

[32]Leon–Garcia, A., *Probability and Random Processes for Electrical Engineering*, Addison–Wesley, Reading, MA, 1994, pp. 207–231.

[33]Sorenson, H. W., *Parameter Estimation: Principles and Problems*, Marcel Dekker, Inc., New York, 1980, pp. 183–198.

[34]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.

[35]Newcomb, S., "A Generalized Theory of the Combination of Observations so as to Obtain the Best Result," *American Journal of Mathematics*, Vol. 8, No. 4, 1886, pp. 343–366.

[36]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.

[37]Bickel, P. J., "One-Step Huber Estimates in the Linear Model," *Journal of the American Statistical Association*, Vol. 70, No. 350, 1975, pp. 428-434.

[38]Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W., *Robust Estimates of Location: Survey and Advances*, Princeton University Press, Princeton, NJ, 1972.

[39]Burl, J. B., *Linear Optimal Control*, Addison–Wesley, Reading, MA, 1999, pp. 437.

[40]Nørgaard, M., Poulsen, N. K. and Ravn, O., "Advances in Derivative-Free State Estimation for Nonlinear Systems," Technical Report IMM-REP-1998-15, Department of Mathematical Modelling, Technical University of Denmark, revised April 2000.

[41]Nørgaard, M., Poulsen, N. K. and Ravn, O., "New Developments in State Estimation for Nonlinear Systems," *Automatica,* Vol. 36, No. 11, 2000, pp. 1627–1638.

[42]Schei, T. S., "A Finite–Difference Method for Linearization in Nonlinear Estimation Algorithms," *Automatica*, Vol. 33, No. 11, 1997, pp. 2053–2058.

[43]Julier, S., Uhlmann, J., and Durrant–Whyte, H. F., "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," *IEEE Transactions on Automatic Control*, Vol. 45, No. 3, 2000, pp. 477–482.

[44]Julier, S. J. and Uhlmann, J. K., "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, Vol. 92, No. 3, 2004, pp. 401–422.

[45]Ito, K. and Xiong, K., "Gaussian Filters for Nonlinear Filtering Problems," *IEEE Transactions on Automatic Control,* Vol. 45, No. 5, 2000, pp. 910–927.

[46]Athans, M., Wishner, R. P. and Bertolini, A., "Suboptimal State Estimation for Continuous-Time Nonlinear Systems from Discrete Noisy Measurements," *IEEE Transactions on Automatic Control,* Vol. 13, No. 5, 1968, pp. 504–514.

[47]Schlee, F. H., Standish, C. J., and Toda, N. F., "Divergence in the Kalman Filter," *AIAA Journal*, Vol. 5, No. 6, 1967, pp. 1114–1120.