Article

Sequential Mixed Cost-Based Multi-Sensor and Relative Dynamics Robust Fusion for Spacecraft Relative Navigation

Shoupeng Li *[ID] and Weiwei Liu [ID]

College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300350, China; liuweiwei@nankai.edu.cn
* Correspondence: lishoupengnku@nankai.edu.cn

Abstract: The non-redescending convex functions degrade the filtering robustness, whereas the redescending non-convex functions improve filtering robustness, but they tend to converge towards local minima. This work investigates the properties of convex and non-convex cost functions from robustness and stability perspectives, respectively. To improve filtering robustness and stability to the high level of non-Gaussian noise, a sequential mixed convex and non-convex cost strategy is presented. To avoid the matrix singularity induced by applying the non-convex function, the M-estimation type Kalman filter is transformed into its information filtering form. Further, to address the problem of the estimation consistency in the iterated unscented Kalman filter, the iterated sigma point filtering framework is adopted using the statistical linear regression method. The simulation results show that, under different levels of heavy-tailed non-Gaussian noise, the mixed cost strategy can avoid the non-convex function-based filters falling into the local minimum, and further can improve the robustness of the convex function-based filter. Therefore, the mixed cost strategy provides a comprehensive improvement in the efficiency of the robust iterated filter.

Keywords: cost function; local minimum; non-Gaussian noise; M-estimation; Kalman filter; iterative methods; spacecraft relative navigation

1. Introduction

Spacecraft relative navigation techniques serve extensive applications in spacecraft rendezvous and docking, spacecraft formation flight, and active space debris removal. Considering autonomy, complexity, and high reliability requirements of space missions, the navigation system needs to be capable of providing high navigation accuracy and coping with complex noise environment. For spacecraft long-range relative navigation, microwave radar can generally serve as an ideal sensor due to its wide operating range from tens of kilometers to a few meters. Observations provided by microwave radars usually obey the heavy-tailed non-Gaussian distribution, whereas Gaussian filters are not robust to this type of observation noise. To address this problem, the robust filtering method based on maximum likelihood estimation has been developed.

Maximum likelihood estimation-based filters are a group of robust Kalman filters [1–3]. Robust filters can suppress the impact of heavy-tailed non-Gaussian noise on estimation, and different cost functions give the robust filters different robustness to the anomalous observations. Therefore, the work on the properties of the cost functions is particularly important. The most extensively implemented cost function is the Huber function, which was first proposed for application in robust statistics [4]. Subsequently, the Huber function was combined with the linear Kalman filter, the extended Kalman filter, and the sigma point Kalman filter, respectively, from which various types of robust Kalman filters emerged [5,6].

The Huber function is a combination of the L1-norm function and the L2-norm function [5]. Since the tuning parameter of the Huber threshold is 1.345 and greater than 1, residuals above this threshold contribute to the robust estimator as the L1-norm of the

residuals, and the absolute values corresponding to the residuals greater than 1 are less than their L2-norm, which indicates that the impact of the greater residuals on the estimator is bounded, and thus robustness is achieved. Furthermore, the Huber function is convex throughout the domain of definition, which prevents the Huber-based estimation from falling into local minima, hence the Huber function balances robustness and numerical stability. The flaws of the Huber function are also noticeable. Due to its non-redescending property, the great residuals still affect the cost function of the Huber-based estimator, thus the Huber function is not robust to outliers [7].

To improve the filtering robustness to outliers, Chen et al. [8] combined the maximum correntropy criterion (MCC) with the Kalman filter, where the Gaussian function allows the filter to obtain stronger robustness than the Huber-based one. Among the robust filtering theories, the MCC-based filter has also attracted much attention since its introduction [9–11]. The MCC is a novel local similarity metric, but the MCC-based Kalman filter possesses almost the same filtering process as that of the Huber-based one, and the only difference is the adopted cost function. Therefore, the Gaussian function is more robust than the Huber function, as shown in the Gaussian influence function or its weight function, their values corresponding to the great residuals rapidly tend to 0, which renders the great residuals unable to affect the estimator. A similar property can also be found in the G-M function [12]. To further develop the MCC theory, the mixture MCC was presented [13,14], but the kernel widths and mixed ratios are given by simulation and lack theoretical guidance. For the asymmetric distribution of residuals, Chen et al. [15] adopted an asymmetric Gaussian function to replace the original symmetric Gaussian function [16], while the kernel width of the asymmetric Gaussian function is another urgent problem.

The Huber function and the Gaussian function have been investigated relatively more; we introduce a few functions, which are less used in robust filtering but have better performance. The Cauchy function has similar properties to the Gaussian function, and compared with the Huber function, the Cauchy function has an advantage in robustness to larger outliers [17]. Huang and Zhang [18] replaced the Gaussian function in the MCC-based filtering framework with the Student's t function. which can better model the heavy-tailed non-Gaussian distribution using more tuning parameters, and thus Student's t-based filters achieve higher estimation accuracy compared to the MCC-based filters, but more tuning parameters imply more complicated tuning procedures. Li et al. [19] instead used the dynamic-covariance-scaling (DCS) function [20,21], which is incorporated into the information filtering framework and obtains strong robustness. Similar functions are the Hampel function and the Tukey function [22].

Another important part of robust estimation is the iterative strategy, but there are relatively few studies have developed the robust and iterative strategy. The extensively implemented iterative strategy is the fixed-point iteration [5,23]. To build a more accurate linearized regression model, several linearization methods are adopted, including the first-order Taylor polynomial expansion, the statistical linearization method, and the statistical linear regression (SLR) method. However, these strategies fail to build an accurate linear regression model; the reason is that the linearized model is built based on the prior estimates rather than the more accurate posterior estimates. To incorporate the posterior estimates, Chang et al. [6] built a nonlinear regression model and then followed the iterated unscented Kalman filter (IUKF) to derive a robust IUKF [24]. As for the nonlinear regression model, Karlgaard [25] used the Gauss–Newton method to iteratively solve the problem, and the method mainly follows the iterated extended Kalman filter (EKF). Wang et al. [26] used the Gauss–Newton and Levenberg–Marquardt methods to derive two robust iterated filters, respectively, where the SLR method is adopted in linearizing the nonlinear function. Li et al. [19] followed the iterated unscented Kalman filter (IUKF) to derive a different robust IUKF.

This work focuses on the properties of different cost functions in M-estimation. The contributions are summarized as follows:

1. This work gives a detailed investigation of convex and non-convex functions from robustness and stability perspectives, respectively.
2. A sequential mixed convex and non-convex cost strategy is presented to combine their properties, and further the switching strategy from the convex function to the non-convex one is proposed.
3. The analytical determination method of the DCS and Gaussian tuning parameters is given.
4. The iterated sigma point Kalman filter is incorporated into robust estimation, and further the information filtering form is given to address the matrix singularity problem.

The remainder is outlined as follows. Section 2 gives the problem statement. Section 3 analyzes the properties of several cost functions and further presents a mixed cost strategy. Section 4 introduces a robust iterated sigma point information filter. Section 5 gives evaluations of the proposed algorithm. Section 6 discusses the simulation results. Section 7 gives conclusions.

2. Problem Statement

Regarding the first problem, robust cost functions can be roughly classified into two groups: convex functions (e.g., the Huber function) and non-convex functions (e.g., the Gaussian, Cauchy, and DCS functions). The non-redescending property of the convex functions limits their robustness, whereas the non-convex functions are redescending and hence completely eliminate the impact of high level non-Gaussian noise. However, the non-convex functions tend to induce estimation to fall into local minima and further results in the filtering divergence. Therefore, the non-convex function is not entirely advantageous over the Huber function. This paper first analyzes the properties of the cost function in detail from both robustness and stability perspectives.

Regarding the second problem, the application of non-convex functions induces the matrix singularity problem, and iteration is also required in the robust filters, hence an iterated information filtering framework is needed to incorporate robust filtering. From the results given in [27], we find that the iterated unscented Kalman filer is not a consistent estimator, and further that the combination of different iterative strategies and the cost functions yields different properties. Therefore, we detail the properties of the robust and iterative strategy.

Through the above analysis, when the robust cost function and the information filter are used to cope with the high level of non-Gaussian noise, there are still many problems to be addressed in this work.

3. Cost Functions for M-Estimation

3.1. Robustness of Different Cost Functions

The property of the cost function affects an estimator's robustness to outliers, thereby affecting filtering efficiency. It is therefore essential to select an appropriate cost function. This section will detail the properties of several extensively implemented cost functions given in Tables 1 and 2, including the L2-norm criterion, the Huber function, the Cauchy function, the Gaussian function, and the DCS function, as shown in Figure 1.

To analyze the robustness of the cost functions, the residuals are divided into three ranges: the Gaussian noise, the non-Gaussian noise, and the outlier. The Gaussian noise refers to the range where the residuals are less than the tuning parameter, the non-Gaussian noise refers to the range where the residuals are greater, but not much greater than, the tuning parameter, and the range where the residuals are much greater than the tuning parameter is considered as the outlier. Note that the three ranges do not have exact bounds, but are relative concepts. Before giving the property of these cost functions, we need to define several concepts, including the cost function, the influence function, and the weight function. The cost function is defined as the function that evaluates the discrepancy between the observed quantity and the predicted quantity. This discrepancy is denoted by $\rho(\xi)$, where $\xi = x - \hat{x}$, ξ is the residual, x is the observed quantity using sensors, and $\hat{x}$ is

the predicted quantity using the system model. The influence function is defined by the following equation: $\psi(\xi) = d\rho/d\xi$, i.e., the influence function is a first-order derivative of the cost function. The weight function is defined by the equation $\phi(\xi) = \psi(\xi)/\xi$, which can be expressed as the ratio of the influence function to the residual.
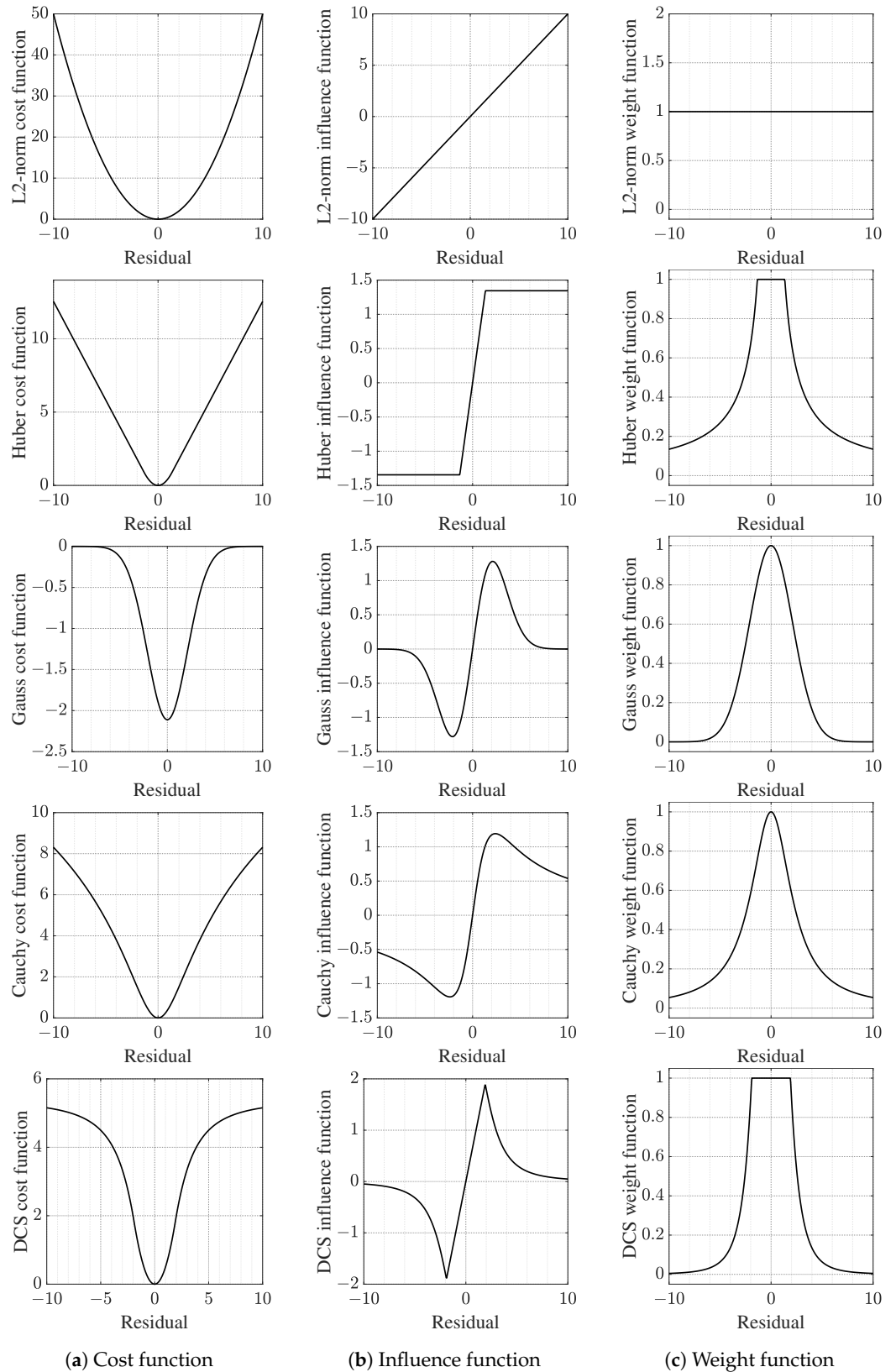

(**a**) Cost function   (**b**) Influence function   (**c**) Weight function

Figure 1. Dependence between different functions and residuals.

Table 1. Several extensively implemented cost functions.

Function Name	Cost Function	Influence Function	Weight Function
L2-norm	$\xi^2/2$	ξ	1
Huber $\begin{cases} \|\xi\| < \eta_h \\ \|\xi\| \geq \eta_h \end{cases}$	$\begin{cases} \xi^2/2 \\ \eta_h\|\xi\| - \eta_h^2/2 \end{cases}$	$\begin{cases} \xi \\ \mathrm{sgn}(\xi)\eta_h \end{cases}$	$\begin{cases} 1 \\ \eta_h/\|\xi\| \end{cases}$
Gaussian	$-\eta_g^2 \exp\left(-\dfrac{\xi^2}{2\eta_g^2}\right)$	$\xi \exp\left(-\dfrac{\xi^2}{2\eta_g^2}\right)$	$\exp\left(-\dfrac{\xi^2}{2\eta_g^2}\right)$
Cauchy	$\dfrac{\eta_c^2 \ln\left(1 + (\xi/\eta_c)^2\right)}{2}$	$\dfrac{\xi}{\left(1 + (\xi/\eta_c)^2\right)}$	$\dfrac{1}{\left(1 + (\xi/\eta_c)^2\right)}$
DCS $\begin{cases} \xi^2 < \eta_d \\ \xi^2 \geq \eta_d \end{cases}$	$\begin{cases} \xi^2/2 \\ \dfrac{\eta_d(3\xi^2 - \eta_d)}{2(\xi^2 + \eta_d)} \end{cases}$	$\begin{cases} \xi \\ \dfrac{4\eta_d^2\xi}{(\eta_d + \xi^2)^2} \end{cases}$	$\begin{cases} 1 \\ \dfrac{4\eta_d^2}{(\eta_d + \xi^2)^2} \end{cases}$

Table 2. Tuning parameters of cost functions for M-estimation.

Function	L2-Norm	Huber	Gaussian	Cauchy	DCS
Tuning parameter	none	1.345	2.1105	2.3849	3.6035

In the Gaussian noise range, the Huber and DCS functions do not impose bounds on the observations, whereas the Cauchy and Gaussian functions do. From the weight functions shown in Figure 1, even in the Gaussian noise, the Cauchy and Gaussian weight functions are less than 1, which reduces the optimality of the L2-norm-based estimation and induces the loss of high-quality observation information, whereas the Huber and DCS weight functions are still 1, which preserves the optimality of the L2-norm-based estimation. As shown in Figure 1, the L2 influence function tends to increase linearly with increasing residuals, which indicates that the L2-norm criterion is not robust to non-Gaussian noise and outliers, and this degrades the optimality of the L2-norm-based estimator, whereas the robust influence function also increases gradually with increasing residuals in the Gaussian noise range. In the non-Gaussian noise range, the influence function corresponding to greater residuals decreases gradually, which indicates that the influence of greater residuals is effectively bounded by the robust cost function, thereby achieving robustness to non-Gaussian noise and outliers.

The robustness of the robust cost function also varies over the outlier range. As shown in Figure 1, the Huber influence function does not reach 0, even when the residuals are much greater, which indicates that the Huber function fails to suppress the effect of outliers on estimates. Conversely, the Cauchy influence function shows suppression of outliers, i.e., as the residuals increase, the influence function gradually converges to 0, and the Cauchy weight function exhibits a similar tendency. The Gaussian and DCS functions show stronger robustness to outliers, i.e., the Gaussian and DCS influence functions converge fast to 0 within the outlier range, which indicates that they can effectively eliminate the effect of outliers on estimates.

3.2. Stability of Different Cost Functions

Another property is that the cost functions with outlier rejection capability exhibit non-convexity, and this renders estimation prone to local minima. Specifically, the Cauchy function is discussed as an example. If the second-order differential of a function is constantly more than or equal to 0, this indicates that the function is convex. That is, if $\ddot{f}(x) \geq 0$, then $f(x)$ is a convex function, and, conversely, $f(x)$ is a non-convex function. The second-order differential of the Cauchy function is given as

$$\ddot{\rho}_c(\xi) = \frac{(\eta_c^4 - \eta_c^2\xi^2)}{(\eta_c^2 + \xi^2)^2} \tag{1}$$

where η is the tuning parameter. As shown in Equation (1) and Figure 2, when $-\eta_c \leq \xi \leq \eta_c$, $\ddot{\rho}_c(\xi) \geq 0$, the interval is the convex interval of the Cauchy function, and when $\xi \leq -\eta_c$ and $\xi \geq \eta_c$, $\ddot{\rho}_c(\xi) < 0$, then the interval is the non-convex interval of the Cauchy function. Similarly, the second-order differential of the Gaussian function is given as

$$\ddot{\rho}_g(\xi) = \left(1 - \frac{\xi^2}{\eta_g^2}\right) \exp\left(-\frac{\xi^2}{2\eta_g^2}\right) \tag{2}$$

As shown in Equation (2) and Figure 2, the Gaussian function is not always convex over the whole domain, and it is convex only in the interval $-\eta_g \leq \xi \leq \eta_g$. The second-order differential of the DCS function is given as

$$\ddot{\rho}_d(\xi) = \begin{cases} 1, & \xi^2 < \eta_d \\ \dfrac{4\eta_d^3 - 12\eta_d^2\xi^4 - 8\eta_d^3\xi^2}{(\eta_d + \xi^2)^4}, & \xi^2 \geq \eta_d \end{cases} \tag{3}$$

Its convex interval is $-\sqrt{\eta_d} \leq \xi \leq \sqrt{\eta_d}$, and its non-convex interval is $\xi < -\sqrt{\eta_d}$ and $\xi > \sqrt{\eta_d}$.

From the above analysis, we find that the Cauchy function, the Gaussian function, and the DCS function all risk falling into the local minimum, whereas the Huber function shows an advantage. The second-order differential of the Huber function is all positive, that is

$$\ddot{\rho}_h(\xi) = \begin{cases} 1, |\xi| < \eta_h \\ 0, |\xi| \geq \eta_h \end{cases} \tag{4}$$

As shown in Equation (4) and Figure 2, the Huber function is entirely a convex function in all domains, and it does not risk falling into the local minimum.
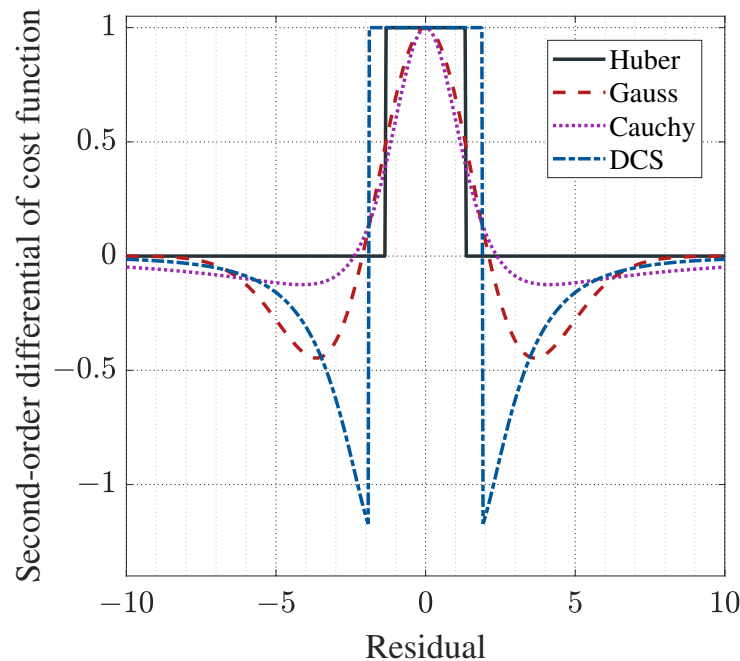

Figure 2. Dependence between second-order differentials of cost functions and residuals.

3.3. Determination of Tuning Parameters

The tuning parameter is a key parameter that affects the properties of the cost function. Despite the extensive research on M-estimation, the determination of the tuning parameter typically relies on simulation, i.e., the tuning parameter is set to different values, and then the optimal estimation accuracy is evaluated by simulation results [28]; this process allows for the determination of the optimal tuning parameter. The simulation results are

inherently random, and the given parameters are typically imprecise and fall within an approximate range, e.g., this method generally yields tuning parameters of 2 for the Gaussian function [29] and [2.3, 3.5] for the DCS function [19].

This section will present an analytical derivation of the exact tuning parameter, which is given according to the 95% asymptotic efficiency on the standard normal distribution. The tuning parameter of the Gaussian function is first given based on this criterion. As elaborated in [30], the M-estimator is asymptotically minimum variance estimation under Gaussian distribution, and its asymptotic variance is given as

$$V[\psi(\xi), p(\xi)] = \frac{\int_{-\infty}^{+\infty} \psi^2(\xi) p(\xi) \mathrm{d}\xi}{\left(\int_{-\infty}^{+\infty} \dot{\psi}(\xi) p(\xi) \mathrm{d}\xi\right)^2} \tag{5}$$

where $V[\cdot]$ is the asymptotic variance of the M-estimator; $\psi(\cdot)$ is the influence function; $p(\cdot)$ is the probability density function (PDF) of the standard normal distribution with the form of

$$p(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) \tag{6}$$

and its first-order differential is

$$\dot{p}(\xi) = -\frac{\xi}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) \tag{7}$$

As discussed in [30], when the influence function $\psi(\xi)$ is not continuous, we can replace $\int_{-\infty}^{+\infty} \dot{\psi}(\xi) p(\xi) \mathrm{d}\xi$ in Equation (5) by $-\int_{-\infty}^{+\infty} \psi(\xi) \dot{p}(\xi) \mathrm{d}\xi$ to avoid the first-order differential $\dot{\psi}(\xi)$, and then we have

$$V[\psi(\xi), p(\xi)] = \frac{\int_{-\infty}^{+\infty} \psi^2(\xi) p(\xi) \mathrm{d}\xi}{\left(\int_{-\infty}^{+\infty} \psi(\xi) \dot{p}(\xi) \mathrm{d}\xi\right)^2} \tag{8}$$

The reason for this is explained as follows: according to the rule of integration by parts, we have

$$\int_{-\infty}^{+\infty} \dot{\psi}(\xi) p(\xi) \mathrm{d}\xi = \psi(\xi) p(\xi)\big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \psi(\xi) \dot{p}(\xi) \mathrm{d}\xi \tag{9}$$

The influence function is bounded at infinity and $p(\xi)$ tends to zero, hence $\psi(\xi) p(\xi)\big|_{-\infty}^{+\infty} = 0$, which further yields

$$\int_{-\infty}^{+\infty} \dot{\psi}(\xi) p(\xi) \mathrm{d}\xi = -\int_{-\infty}^{+\infty} \psi(\xi) \dot{p}(\xi) \mathrm{d}\xi \tag{10}$$

The relative efficiency of the M-estimator is defined as

$$\mathrm{RE} = \frac{V_m[\psi_m(\xi), p(\xi)]}{V_r[\psi_r(\xi), p(\xi)]} \tag{11}$$

where $V_m[\cdot]$ is the asymptotic variance of the M-estimator; $V_r[\cdot]$ is the asymptotic variance of the robust estimator. Note that the M-estimator is applied as a reference estimator, and it is equivalent to the L2-norm-based estimator under the Gaussian distribution, whereby it can be considered as the optimal estimator.

Substituting the influence function of the M-estimator and Equations (6) and (7) into Equation (8) yields the variance of the M-estimator under standard normal distribution:

$$V_m(\psi_m(\xi), p(\xi)) = \frac{\int_{-\infty}^{+\infty} \xi^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^2\right) d\xi}{\left(\int_{-\infty}^{+\infty} -\xi \frac{\xi}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^2\right) d\xi\right)^2} = \frac{\frac{2}{\sqrt{2\pi}} \int_0^{+\infty} \xi^2 \exp\left(-\frac{1}{2}\xi^2\right) d\xi}{\left(\frac{2}{\sqrt{2\pi}} \int_0^{+\infty} \xi^2 \exp\left(-\frac{1}{2}\xi^2\right) d\xi\right)^2} \tag{12}$$

Substituting the Gamma function $\int_0^{+\infty} \xi^2 \exp\left(-\xi^2/2\right) d\xi = \sqrt{2}\Gamma(3/2) = \sqrt{2\pi}/2$ into Equation (12) yields

$$V_m[\psi_m(\xi), f(\xi)] = 1 \tag{13}$$

Therefore, Equation (11) can be reduced to

$$\text{RE} = \frac{1}{V_r[\psi_r(\xi), p(\xi)]} \tag{14}$$

Substituting the Gaussian influence function $\xi \exp\left(-\xi^2/2\eta_g^2\right)$ and Equations (6) and (7) into Equation (8) yields

$$V_g(\psi_g(\xi), p(\xi)) = \frac{\int_{-\infty}^{+\infty} \left(\xi \exp\left(-\frac{\xi^2}{2\sigma^2}\right)\right)^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^2\right) d\xi}{\left(\int_{-\infty}^{+\infty} \xi \exp\left(-\frac{\xi^2}{2\sigma^2}\right) \times \frac{\xi}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^2\right) d\xi\right)^2} = \frac{\frac{2}{\sqrt{2\pi}} \int_0^{+\infty} \xi^2 \left(\exp\left(-\frac{\xi^2}{2\sigma^2}\right)\right)^2 \exp\left(-\frac{1}{2}\xi^2\right) d\xi}{\left(\frac{2}{\sqrt{2\pi}} \int_0^{+\infty} \xi^2 \exp\left(-\frac{\xi^2}{2\sigma^2}\right) \exp\left(-\frac{1}{2}\xi^2\right) d\xi\right)^2} \tag{15}$$

Finding an analytical solution to Equation (15) is challenging, whereas numerical integration can be employed to solve the improper integral problem. The numerical solution of the anomalous integral is demonstrated by solving $\int_0^{+\infty} \xi^2 \exp\left(-\xi^2/2\eta_g^2\right) \exp(-\xi^2/2) d\xi$ as an example. Given the tuning parameter η_g, the upper bound of the integral is continuously increased until the change of the integral is less than a preset small value, i.e., the final value is the result of the improper integral. Note that the influence function is bounded at infinity, and $\exp\left(-\xi^2/2\eta_g^2\right)$ and $\exp(-\xi^2/2)$ tend to zero, thereby allowing the result of $\int_0^{+\infty} \xi^2 \exp\left(-\xi^2/2\eta_g^2\right) \exp(-\xi^2/2) d\xi$ to converge, hence the improper integral can be solved using the numerical integration.

Substituting the results of improper integral Equation (15) into Equation (14) yields the relative efficiency between the Gaussian function-based estimator and the M-estimator. By setting different values of the tuning parameter, the relative efficiency reaches 95%, and the corresponding value is the required optimal tuning parameter. The optimal tuning parameter is 2.1105, and the relationship between the tuning parameter and the relative efficiency is illustrated in Figure 3 and Table 3.

Table 3. Relative efficiency of cost functions corresponding to different tuning parameters.

Relative Efficiency	50%	90%	95%	99%
Tuning parameter of Gaussian function	0.8024	1.6852	2.1105	3.3522
Tuning parameter of DCS function	0.6476	2.6268	3.6035	6.010

The work on the DCS function [19,21] has not provided an exact value of the tuning parameter. This part will derive the optimal tuning parameter of the DCS function. Different from the Gaussian function, the DCS function is piecewise, and therefore the integration of the piecewise function is required. Substituting the DCS influence function and Equations (6) and (7) into Equation (8), the variance of the DCS-based M-estimator is

$$V_d[\psi_d(\xi), p(\xi)] = \frac{\int_{-\infty}^{+\infty} \psi_d^2(\xi) p(\xi) d\xi}{\left(\int_{-\infty}^{+\infty} \psi_d(\xi) \dot{p}(\xi) d\xi\right)^2} \tag{16}$$

where

$$\int_{-\infty}^{+\infty} \psi_d^2(\xi) p(\xi) d\xi = \frac{2}{\sqrt{2\pi}} \left(\int_0^{\sqrt{\eta_d}} \xi^2 \exp\left(-\frac{\xi^2}{2}\right) d\xi + \int_{\sqrt{\eta_d}}^{+\infty} \frac{16\eta_d^4 \xi^2}{(\eta_d + \xi^2)^4} \exp\left(-\frac{\xi^2}{2}\right) d\xi \right)$$

$$\int_{-\infty}^{+\infty} \psi_d(\xi) \dot{p}(\xi) d\xi = -\frac{2}{\sqrt{2\pi}} \left(\int_0^{\sqrt{\eta_d}} \xi^2 \exp\left(-\frac{\xi^2}{2}\right) d\xi + \int_{\sqrt{\eta_d}}^{+\infty} \frac{4\eta_d^2 \xi^2}{(\eta_d + \xi^2)^2} \exp\left(-\frac{\xi^2}{2}\right) d\xi \right)$$

(17)

The improper integrals involved in Equation (17) also can be solved using numerical integration, which is not repeated herein. Substituting the results of improper integral Equation (16) into Equation (14) yields the relative efficiency between the DCS-based estimator and the M-estimator. Following the method in the Gaussian function, the optimal tuning parameter is 3.6035, and the relationship between the tuning parameter and the relative efficiency is illustrated in Figure 3 and Table 3.
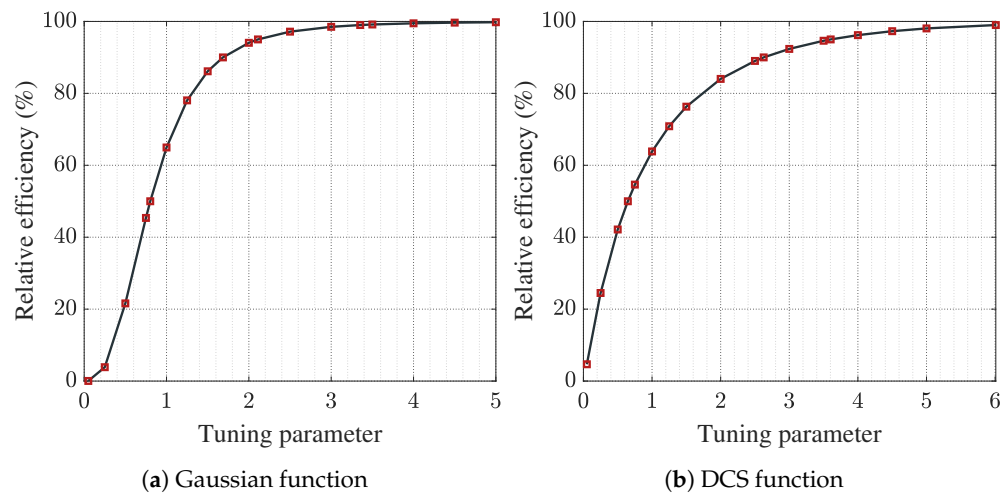

(**a**) Gaussian function   (**b**) DCS function

Figure 3. Dependence between different functions and tuning parameters.

3.4. Sequential Mixed Cost Function

As discussed in Sections 3.1 and 3.2, the Gaussian and DCS functions are robust to outliers, but their non-convexity induces the estimates to fall into the local minimum, whereas the Huber function is convex but not robust to outliers. Therefore, we synthesize the advantages of the two types of robust cost functions, thereby proposing a sequential mixed cost strategy using both the non-convex and convex cost functions. This strategy allows the estimator to combine robustness and stability. Specifically, an iterative update is required for the robust estimator, the convex robust cost function is applied until the filter converges in the first few iterations, and then the non-convex robust cost function is applied in the subsequent iterations to eliminate outliers. Generally, the Huber function is selected as the convex function, and the Cauchy, Gaussian, or DCS function is selected as the non-convex function.

This section further gives the switching strategy from the convex function to the non-convex function. Due to the stability of convex functions and the robustness of non-convex functions, iterative updates should be carried out based on convex functions as much as possible until a satisfied estimate is obtained, and fewer iterative updates should be carried out based on non-convex functions to eliminate the impact of the high level of non-Gaussian noise. Therefore, this work recommends performing only one non-convex iteration in the last iteration.

4. Robust Iterated Sigma Point Information Filter

This section incorporates the iterated sigma point Kalman filter (ISPKF) into robust estimation and further gives its information filtering form for numerical stability [31]. The derived robust ISPKF holds the same prediction step as the conventional sigma point

Kalman filter (SPKF), and the only difference lies in the update step. Therefore, this section only introduces the iterative update step of the robust filter.

4.1. ISPKF-Based Robust Iterative Update

Consider the following nonlinear model:

$$\begin{aligned} \boldsymbol{x}_k &= \boldsymbol{f}(\boldsymbol{x}_{k-1}) + \boldsymbol{\omega}_{k-1} \\ \boldsymbol{z}_k &= \boldsymbol{g}(\boldsymbol{x}_k) + \boldsymbol{v}_k \end{aligned} \tag{18}$$

where $\boldsymbol{f}(\cdot)$ and $\boldsymbol{g}(\cdot)$ are the nonlinear process model and observation model, respectively; $\boldsymbol{x}_k$ and $\boldsymbol{z}_k$ are the state vector and observation vector, respectively; $\boldsymbol{\omega}_k$ and $\boldsymbol{v}_k$ are the process noise and observation noise, respectively; and $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$ are the process noise covariance and the observation noise covariance, respectively.

Take the jth iteration as an example to illustrate the ISPKF-based robust iterative update. According to the statistical linear regression theory, the observation model is transformed into a linear one [26]:

$$\boldsymbol{z}_k = \boldsymbol{G}_{x,k}^{(j)} \boldsymbol{x}_k + \boldsymbol{G}_{c,k}^{(j)} + \boldsymbol{v}_{t,k}^{(j)} \tag{19}$$

where $\boldsymbol{G}_{x,k}^{(j)} = \left(\hat{\boldsymbol{P}}_{xz,k}^{(j)}\right)^{\mathrm{T}} \left(\hat{\boldsymbol{P}}_k^{(j-1)}\right)^{-1}$ is the linearized observation matrix; $\boldsymbol{G}_{c,k}^{(j)} = \hat{\boldsymbol{z}}_k^{(j)} - \boldsymbol{G}_{x,k}^{(j)} \hat{\boldsymbol{x}}_k^{(j-1)}$ is the constant matrix; j denotes the index of the jth iteration; $\hat{\boldsymbol{x}}_k^{(j-1)}$ is the posterior state estimate; $\hat{\boldsymbol{z}}_k^{(j)}$ is the predicted observation; $\hat{\boldsymbol{P}}_{xz,k}^{(j)}$ is the cross-covariance; $\hat{\boldsymbol{P}}_k^{(j-1)}$ is the posterior covariance; $\boldsymbol{v}_{t,k}^{(j)}$ is the total observation noise, consisting of the linearization error $\hat{\boldsymbol{v}}_k^{(j)}$ and the original observation noise $\boldsymbol{v}_k$; the total observation noise covariance is $\boldsymbol{R}_{t,k}^{(j)} = \hat{\boldsymbol{P}}_{z,k}^{(j)} - \boldsymbol{G}_{x,k}^{(j)} \hat{\boldsymbol{P}}_k^{(j-1)} \left(\boldsymbol{G}_{x,k}^{(j)}\right)^{\mathrm{T}} + \boldsymbol{R}_k$, the covariance induced by linearization errors is compensated for in the total observation covariance, which improves the estimation consistency and accuracy; $\hat{\boldsymbol{P}}_{z,k}^{(j)}$ is the covariance of predicted observation; and $p(\cdot)$ is the Gaussian probability density function. That is,

$$\begin{aligned} \hat{\boldsymbol{z}}_k^{(j)} &= \int \boldsymbol{g}(\boldsymbol{x}_k) p\left(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_k^{(j-1)}, \hat{\boldsymbol{P}}_k^{(j-1)}\right) d\boldsymbol{x}_k \\ \hat{\boldsymbol{P}}_{xz,k}^{(j)} &= \int \left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^{(j-1)}\right) \left(\boldsymbol{g}(\boldsymbol{x}_k) - \hat{\boldsymbol{z}}_k^{(j)}\right)^{\mathrm{T}} p\left(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_k^{(j-1)}, \hat{\boldsymbol{P}}_k^{(j-1)}\right) d\boldsymbol{x}_k \\ \hat{\boldsymbol{P}}_{z,k}^{(j)} &= \int \left(\hat{\boldsymbol{z}}_k^{(j)} - \boldsymbol{g}(\boldsymbol{x}_k)\right) \left(\hat{\boldsymbol{z}}_k^{(j)} - \boldsymbol{g}(\boldsymbol{x}_k)\right)^{\mathrm{T}} p\left(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_k^{(j-1)}, \hat{\boldsymbol{P}}_k^{(j-1)}\right) d\boldsymbol{x}_k \end{aligned} \tag{20}$$

The weighted least-squares cost function for the ISPKF-based update is built as

$$\Omega_k^{(j)} = \left\| \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1} \right\|_{\hat{\boldsymbol{P}}_{k|k-1}^{-1}}^2 + \left\| \boldsymbol{z}_k - \boldsymbol{G}_{x,k}^{(j)} \boldsymbol{x}_k - \boldsymbol{G}_{c,k}^{(j)} \right\|_{\left(\boldsymbol{R}_{t,k}^{(j)}\right)^{-1}}^2 \tag{21}$$

where $\hat{\boldsymbol{x}}_{k|k-1}$ is the prior state estimate; $\hat{\boldsymbol{P}}_{k|k-1}$ is the prior covariance. That is,

$$\begin{aligned} \hat{\boldsymbol{x}}_{k|k-1} &= \int \boldsymbol{f}(\boldsymbol{x}_{k-1}) p\left(\boldsymbol{x}_{k-1}; \hat{\boldsymbol{x}}_{k-1}, \hat{\boldsymbol{P}}_{k-1}\right) d\boldsymbol{x}_{k-1} \\ \hat{\boldsymbol{P}}_{k|k-1} &= \int \left(\boldsymbol{f}(\boldsymbol{x}_{k-1}) - \hat{\boldsymbol{x}}_{k|k-1}\right) \left(\boldsymbol{f}(\boldsymbol{x}_{k-1}) - \hat{\boldsymbol{x}}_{k|k-1}\right)^{T} p\left(\boldsymbol{x}_{k-1}; \hat{\boldsymbol{x}}_{k-1}, \hat{\boldsymbol{P}}_{k-1}\right) d\boldsymbol{x}_{k-1} + \boldsymbol{Q}_{k-1} \end{aligned} \tag{22}$$

To reduce the effect of non-Gaussian noise or outliers on estimation, the residuals in Equation (21) are bounded by the robust cost function. Thus, the cost function in Equation (21) is modified as

$$\Omega_{r,k}^{(j)} = \sum_{i=1}^{m} \rho\left(\bar{\xi}_{x,i}^{(j)}\right) + \sum_{i=1}^{n} \rho\left(\bar{\xi}_{z,i}^{(j)}\right) \tag{23}$$

where $\bar{\xi}_{x,i}^{(j)}$ and $\bar{\xi}_{z,i}^{(j)}$ are normalized residuals, that is,

$$\bar{\xi}_x^{(j)} = \left(\sqrt{\hat{P}_{k|k-1}}\right)^{-1}\left(x_k - \hat{x}_{k|k-1}\right)$$

$$\bar{\xi}_z^{(j)} = \left(\sqrt{R_{t,k}^{(j)}}\right)^{-1}\left(z_k - G_{x,k}^{(j)}x_k - G_{c,k}^{(j)}\right) \tag{24}$$

$\bar{\xi}_i$ denotes the ith component of $\bar{\xi}$.

As demonstrated in [27], Equation (23) can be transformed into another equivalent form as follows:

$$\Omega_{r,k}^{(j)} = \left(\xi_{x,k}^{(j)}\right)^{\mathrm{T}}\left(\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}\right)^{-1}\xi_{x,k}^{(j)} + \left(\xi_{z,k}^{(j)}\right)^{\mathrm{T}}\left(R_{t,k}^{(j),\,\mathrm{mod}}\right)^{-1}\xi_{z,k}^{(j)} \tag{25}$$

where $\xi_x^{(j)} = \left(x_k - \hat{x}_{k|k-1}\right)$ and $\xi_z^{(j)} = \left(z_k - G_{x,k}^{(j)}x_k - G_{c,k}^{(j)}\right)$ are residuals; $\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}$ is the modified prior covariance $\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}} = \left(\sqrt{\hat{P}_{k|k-1}}\right)\left(\Phi_x^{(j)}\right)^{-1}$ $\left(\sqrt{\hat{P}_{k|k-1}}\right)^{\mathrm{T}}$; $R_{t,k}^{(j),\,\mathrm{mod}}$ is the modified total observation noise covariance $R_{t,k}^{(j),\,\mathrm{mod}} = \left(\sqrt{R_{t,k}^{(j)}}\right)\left(\Phi_z^{(j)}\right)^{-1}\left(\sqrt{R_{t,k}^{(j)}}\right)^{\mathrm{T}}$; and $\Phi_x^{(j)}$ and $\Phi_z^{(j)}$ are the weight functions, which are computed as $\Phi_x^{(j)}(i,i) = \phi\left(\bar{\xi}_{x,i}^{(j)}\right)$ and $\Phi_z^{(j)}(i,i) = \phi\left(\bar{\xi}_{z,i}^{(j)}\right)$, respectively.

Based on the posterior estimate of the $(j-1)$th iteration, the posterior estimate of the jth iteration using the Gauss–Newton method can be written as follows:

$$\hat{x}_k^{(j)} = \hat{x}_k^{(j-1)} + \Delta\hat{x}_k^{(j)} \tag{26}$$

where $\Delta\hat{x}_k^{(j)}$ is the update step of the posterior state. Substituting $\hat{x}_k^{(j)}$ in Equation (26) for x_k in Equation (25) yields

$$\Omega_{r,k}^{(j)} = \left(z_k - G_{x,k}^{(j)}\left(\hat{x}_k^{(j-1)} + \Delta\hat{x}_k^{(j)}\right) - G_{c,k}^{(j)}\right)^{\mathrm{T}}\left(R_{t,k}^{(j),\,\mathrm{mod}}\right)^{-1}\left(z_k - G_{x,k}^{(j)}\left(\hat{x}_k^{(j-1)} + \Delta\hat{x}_k^{(j)}\right) - G_{c,k}^{(j)}\right) +$$
$$\left(\hat{x}_k^{(j-1)} + \Delta\hat{x}_k^{(j)} - \hat{x}_{k|k-1}\right)^{\mathrm{T}}\left(\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}\right)^{-1}\left(\hat{x}_k^{(j-1)} + \Delta\hat{x}_k^{(j)} - \hat{x}_{k|k-1}\right) \tag{27}$$

Finding the partial derivative of Equation (27) with respect to $\Delta\hat{x}_k^{(j)}$, setting it equal to zero, and considering $G_{c,k}^{(j)} = \hat{z}_k^{(j)} - G_{x,k}^{(j)}\hat{x}_k^{(j-1)}$, then we have

$$\left(\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}\right)^{-1}\left(\hat{x}_k^{(j-1)} + \Delta\hat{x}_k^{(j)} - \hat{x}_{k|k-1}\right) - \left(G_{x,k}^{(j)}\right)^{\mathrm{T}}\left(R_{t,k}^{(j),\,\mathrm{mod}}\right)^{-1}\left(z_k - \hat{z}_k^{(j)} - G_{x,k}^{(j)}\Delta\hat{x}_k^{(j)}\right) = 0 \tag{28}$$

Further simplifying Equation (28) and applying the matrix inversion lemma, we have

$$\Delta\hat{x}_k^{(j)} = \hat{x}_{k|k-1} - \hat{x}_k^{(j-1)} + K_k^{(j)}\left(z_k - \hat{z}_k^{(j)} + G_{x,k}^{(j)}\left(\hat{x}_k^{(j-1)} - \hat{x}_{k|k-1}\right)\right) \tag{29}$$

where $K_k^{(j)}$ is the Kalman gain, i.e.,

$$K_k^{(j)} = \hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}\left(G_{x,k}^{(j)}\right)^{\mathrm{T}}\left(R_{t,k}^{(j),\,\mathrm{mod}} + G_{x,k}^{(j)}\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}\left(G_{x,k}^{(j)}\right)^{\mathrm{T}}\right)^{-1} \tag{30}$$

Substituting Equation (29) into Equation (26), we get the posterior state estimate

$$\hat{x}_k^{(j)} = \hat{x}_{k|k-1} + K_k^{(j)}\left(z_k - \hat{z}_k^{(j)} - G_{x,k}^{(j)}\left(\hat{x}_{k|k-1} - \hat{x}_k^{(j-1)}\right)\right) \tag{31}$$

Finding the partial derivative of Equation (25) with respect to x_k, setting it equal to zero, and substituting $\hat{x}_k^{(j)}$ into x_k, we have

$$\left(\hat{x}_k^{(j)} - \hat{x}_{k|k-1}\right)^{\mathrm{T}}\left(\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}\right)^{-1} - \left(z_k - G_{x,k}^{(j)}\hat{x}_k^{(j)} - G_{c,k}^{(j)}\right)^{\mathrm{T}}\left(R_{t,k}^{(j),\,\mathrm{mod}}\right)^{-1}G_{x,k}^{(j)} = 0 \tag{32}$$

Substituting Equation (19) into Equation (32), we have

$$\left(\hat{x}_k^{(j)} - x_k + x_k - \hat{x}_{k|k-1}\right)^{\mathrm{T}}\left(\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}\right)^{-1} - \left(G_{x,k}^{(j)}\left(x_k - \hat{x}_k^{(j)}\right) + v_{t,k}^{(j)}\right)^{\mathrm{T}}\left(R_{t,k}^{(j),\,\mathrm{mod}}\right)^{-1}G_{x,k}^{(j)} = 0 \tag{33}$$

Simplifying Equation (33) yields

$$\left(x_k - \hat{x}_k^{(j)}\right) = -\left(\left(G_{x,k}^{(j)}\right)^{\mathrm{T}}\left(R_{t,k}^{(j),\,\mathrm{mod}}\right)^{-1}G_{x,k}^{(j)} + \left(\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}\right)^{-1}\right)^{-1}\left(\left(G_{x,k}^{(j)}\right)^{\mathrm{T}}\left(R_{t,k}^{(j),\,\mathrm{mod}}\right)^{-1}v_{t,k}^{(j)} + \left(\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}\right)^{-1}\left(x_k - \hat{x}_{k|k-1}\right)\right) \tag{34}$$

The posterior covariance is estimated as

$$\hat{P}_k^{(j)} = E\left[\left(x_k - \hat{x}_k^{(j)}\right)\left(x_k - \hat{x}_k^{(j)}\right)^{\mathrm{T}}\right] = \left(\left(G_{x,k}^{(j)}\right)^{\mathrm{T}}\left(R_{t,k}^{(j),\,\mathrm{mod}}\right)^{-1}G_{x,k}^{(j)} + \left(\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}\right)^{-1}\right)^{-1} \tag{35}$$

Applying the matrix inversion lemma and then substituting Equation (30) into Equation (35) results in the simplification of Equation (35) to

$$\hat{P}_k^{(j)} = \left(I - K_k^{(j)}G_{x,k}^{(j)}\right)\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}} \tag{36}$$

where I is the identity matrix.

Remark 1. *Compared to the IUKF-based robust filters given in [19,27], the ISPKF-based robust filter proposed in this work is more robust and consistent. In IUKF, the innovations are used to modify the posterior estimates, which have already incorporated the observation information, thereby rendering the state correlated with the observation, but IUKF ignores the correlation between the two mentioned above, which results in a degraded consistency of IUKF. On the contrary, in ISPKF, the innovations are used to modify the prior estimates, where the state is not correlated with the observation, hence the weak consistency problem of IUKF is addressed.*

4.2. Information Filtering Form

As shown in Equation (30), the Kalman gain is calculated using the modified form of total observation noise covariance $R_{t,k}^{(j),\,\mathrm{mod}}$. When the residuals are relatively great, the corresponding weight tends to be 0, so inverting the weight matrix may potentially induce matrix singularity, and this further results in the Kalman filter failing to work. Therefore, we give the information filtering form to address the problem.

Applying the matrix inversion lemma to Equation (35) yields

$$\hat{Y}_k^{(j)} = \hat{Y}_{k|k-1}^{(j),\mathrm{mod}} + \left(G_{x,k}^{(j)}\right)^{\mathrm{T}}\left(R_{t,k}^{(j),\mathrm{mod}}\right)^{-1}G_{x,k}^{(j)} \tag{37}$$

where $\hat{Y}_k^{(j)}$ and $\hat{Y}_{k|k-1}^{(j),\,\mathrm{mod}}$ are the information matrices defined as $\hat{Y}_k^{(j)} = \left(\hat{P}_k^{(j)}\right)^{-1}$ and $\hat{Y}_{k|k-1}^{(j),\,\mathrm{mod}} = \left(\hat{P}_{k|k-1}^{(j),\,\mathrm{mod}}\right)^{-1}$, respectively. As shown in Equation (37), the update of the information matrix involves the inverse matrices of the prior covariance and the total observed noise covariance

$$\hat{Y}_{k|k-1}^{(j),\mathrm{mod}} = \left(\hat{P}_{k|k-1}^{(j),\mathrm{mod}}\right)^{-1} = \left(\sqrt{\hat{P}_{k|k-1}}\right)^{-\mathrm{T}}\Phi_x^{(j)}\left(\sqrt{\hat{P}_{k|k-1}}\right)^{-1}$$
$$\left(R_{t,k}^{(j),\mathrm{mod}}\right)^{-1} = \left(\sqrt{R_{t,k}^{(j)}}\right)^{-\mathrm{T}}\Phi_z^{(j)}\left(\sqrt{R_{t,k}^{(j)}}\right)^{-1} \tag{38}$$

The inverse of the weight matrices is not involved in Equation (38), thereby effectively avoiding the matrix singularity.

Applying the matrix inversion lemma to Equation (30) and considering the relationship between the posterior and prior covariances in Equation (36) yields

$$K_k^{(j)} = \hat{P}_k^{(j)} \left(G_{x,k}^{(j)} \right)^{\mathrm{T}} \left(R_{t,k}^{(j),\mathrm{mod}} \right)^{-1} \tag{39}$$

Multiplying $\hat{Y}_k^{(j)}$ by the left-hand side of Equation (31) yields

$$\hat{Y}_k^{(j)} \hat{x}_k^{(j)} = \hat{Y}_k^{(j)} \hat{x}_{k|k-1} + \hat{Y}_k^{(j)} K_k^{(j)} \left(z_k - \hat{z}_k^{(j)} - G_{x,k}^{(j)} \left(\hat{x}_{k|k-1} - \hat{x}_k^{(j-1)} \right) \right) \tag{40}$$

and then substituting Equation (39) into Equation (40) yields

$$\begin{aligned} \hat{Y}_k^{(j)} \hat{x}_k^{(j)} = & \left(\hat{Y}_k^{(j)} - \left(G_{x,k}^{(j)} \right)^{\mathrm{T}} \left(R_{t,k}^{(j),\mathrm{mod}} \right)^{-1} G_{x,k}^{(j)} \right) \hat{x}_{k|k-1} + \\ & \left(G_{x,k}^{(j)} \right)^{\mathrm{T}} \left(R_{t,k}^{(j),\mathrm{mod}} \right)^{-1} \left(z_k - \hat{z}_k^{(j)} + G_{x,k}^{(j)} \hat{x}_k^{(j-1)} \right) \end{aligned} \tag{41}$$

further substituting Equation (37) into Equation (41), Equation (41) is reduced to

$$\hat{y}_k^{(j)} = \hat{y}_{k|k-1}^{(j)} + \left(G_{x,k}^{(j)} \right)^{\mathrm{T}} \left(R_{t,k}^{(j),\mathrm{mod}} \right)^{-1} \left(z_k - \hat{z}_k^{(j)} + G_{x,k}^{(j)} \hat{x}_k^{(j-1)} \right) \tag{42}$$

where $\hat{y}_k^{(j)}$ and $\hat{y}_{k|k-1}^{(j)}$ are information state defined as $\hat{y}_k^{(j)} = \left(\hat{P}_k^{(j)} \right)^{-1} \hat{x}_k^{(j)}$ and $\hat{y}_{k|k-1}^{(j)} = \left(\hat{P}_{k|k-1}^{(j),\mathrm{mod}} \right)^{-1} \hat{x}_{k|k-1}$, respectively. Similarly, the inverse of the weight matrices is not involved in Equation (42), thereby also avoiding the matrix singularity.

Remark 2. *Note that the observation noise covariance involved in Equations (37) and (42) is the total observation noise covariance* $R_{t,k}^{(j)}$, *which consists of two components: the linearized error covariance* $\tilde{R}_k^{(j)}$ *and the original observation noise covariance* R_k. *The covariance induced by the linearization errors is incorporated into the observation noise covariance of the information filter.*

To avoid confusion, the algorithm is summarized in Algorithm 1.

4.3. Property of Robust and Iterative Strategy

How does the non-convex function induce estimates to fall into the local minimum? As shown in Figure 1, the weight function corresponding to the much greater residual rapidly approaches 0 for the Gaussian, Cauchy, and DCS functions. This property facilitates the estimator to eliminate the effect of the greater residuals on estimation. However, when the initial error of the estimator is great, even if the observation information is not contaminated by non-Gaussian noise or outliers, the residual between the observation and its predicted value is still great, i.e., the residual of the innovation $\xi_z = z_k - \hat{z}_k$ is great. The corresponding weight approaches 0, i.e., $\Phi_z \approx 0$, hence the inverse of the modified total observation covariance approaches 0, as shown in Equation (38), i.e., $\left(R_{t,k}^{(j),\mathrm{mod}} \right)^{-1} = \left(\sqrt{R_{t,k}^{(j)}} \right)^{-\mathrm{T}} \Phi_z^{(j)} \left(\sqrt{R_{t,k}^{(j)}} \right)^{-1} \approx 0$. Further, the corresponding state innovation also approaches 0 as shown in Equation (42), i.e., $\left(G_{x,k}^{(j)} \right)^{\mathrm{T}} \left(R_{t,k}^{(j),\mathrm{mod}} \right)^{-1} \left(z_k - \hat{z}_k^{(j)} + G_{x,k}^{(j)} \hat{x}_k^{(j-1)} \right) \approx 0$. Therefore, the modification of the innovation to the prior estimation is significantly reduced, or even approaching ineffectiveness, and the state is not effectively updated. The estimator only can rely on the process model to predict the estimate. However, the estimate remains within the neighborhood of the prior estimate, and the residual between the nominal state and the prior estimate is still great. Consequently, the subsequent posterior updates remain ineffective, eventually leading to filtering divergence.

Algorithm 1: Robust iterated sigma point information filter using sequential mixed cost

---

    Data: $\hat{x}_{k-1}, \hat{P}_{k-1}, z_k$

    Result: $\hat{x}_k, \hat{P}_k$

1 Compute prior state and covariance using Equation (22);

2 Initialize iteration $\hat{x}_k^{(0)} = \hat{x}_{k|k-1}$, $\hat{P}_k^{(0)} = \hat{P}_{k|k-1}$;

3 **for** $j \in [1, 2, \cdots]$ **do**

4   Compute $\hat{z}_k^{(j)}, \hat{P}_{z,k}^{(j)}$, and $\hat{P}_{xz,k}^{(j)}$ using Equation (20);

5   Compute $\xi_{x,i}^{(j)}, \xi_{z,i}^{(j)}$ using Equation (24);

6   **if** *iterate using convex function* **then**

7           $\Phi_x^{(j)}(i,i) = \phi_{Huber}\left(\xi_{x,i}^{(j)}\right), \Phi_z^{(j)}(i,i) = \phi_{Huber}\left(\xi_{z,i}^{(j)}\right)$;

8   **else**

9           $\Phi_x^{(j)}(i,i) = \phi_{nonconvex}\left(\xi_{x,i}^{(j)}\right), \Phi_z^{(j)}(i,i) = \phi_{nonconvex}\left(\xi_{z,i}^{(j)}\right)$;

10   **end**

11   Compute $\hat{Y}_{k|k-1}^{(j),\text{mod}}, \left(R_{t,k}^{(j),\text{mod}}\right)^{-1}$ using Equation (38) and $\hat{y}_{k|k-1}^{(j)} = \hat{Y}_{k|k-1}^{(j),\text{mod}}\hat{x}_{k|k-1}$;

12   Compute $\hat{Y}_k^{(j)}, \hat{y}_k^{(j)}$ using Equations (37) and (42);

13   Compute $\hat{P}_k^{(j)} = \left(\hat{Y}_k^{(j)}\right)^{-1}, \hat{x}_k^{(j)} = \hat{P}_k^{(j)}\hat{y}_k^{(j)}$;

14 **end**

---

5. Simulation and Results

One hundred Monte Carlo simulation runs were performed under different levels of non-Gaussian noise. The cost functions involved in the comparison are the Huber function, the Gaussian function, the Cauchy function, the DCS function, and the mixed Huber and non-convex function (henceforth referred to as Huber/Gauss, Huber/Cauchy, and Huber/DCS), respectively. The number of iterations were all set to four. For the mixed cost strategy, the first three iterations are performed based on the Huber function, and the subsequent iteration is based on the non-convex functions (Gauss, Cauchy, and DCS, respectively).

5.1. Process and Observation Models

The simulation scenario is spacecraft relative navigation in an elliptical orbit, hence the process model is the Tschauner–Hempel (T-H) equations, which are expressed as follows:

$$
\begin{aligned}
\ddot{x}_k &= \omega_c^2 x_k + 2\omega_c \dot{y}_k + \dot{\omega}_c y_k + \frac{\mu}{r_c^2} - \frac{\mu(r_c + x_k)}{r_d^3} + f_{ux,k} + f_{wx,k} \\
\ddot{y}_k &= \omega_c^2 y_k - 2\omega_c \dot{x}_k - \dot{\omega}_c x_k - \frac{\mu y_k}{r_d^3} + f_{uy,k} + f_{wy,k} \\
\ddot{z}_k &= -\frac{\mu z_k}{r_d^3} + f_{uz,k} + f_{wz,k}
\end{aligned}
\tag{43}
$$

where the state vector consists of the relative positions and their first-order differentials, i.e., $x_k = [x_k, y_k, z_k, \dot{x}_k, \dot{y}_k, \dot{z}_k]^\mathrm{T}$; f_w is the perturbing acceleration; f_u is the control specific force; μ is the gravitational constant; ω_c and r_c are the orbital angular velocity and orbital radius of the chief spacecraft, respectively; r_d is the orbital radius of the deputy spacecraft; and ω_c, r_c, and r_d are related as follows:

$$\ddot{r}_c = r_c \omega_c^2 - \frac{\mu}{r_c^2}$$

$$\dot{\omega}_c = -\frac{2\dot{r}_c \omega_c}{r_c} \tag{44}$$

$$r_d = \sqrt{(r_c + x)^2 + y^2 + z^2}$$

The relative observation z_k is provided by the radar of the chief spacecraft and the observation model is given as follows:

$$z_k = \begin{bmatrix} \sqrt{x_k^2 + y_k^2 + z_k^2} \\ \arctan\left(\dfrac{y_k}{x_k}\right) \\ \arctan\left(\dfrac{z_k}{\sqrt{x_k^2 + y_k^2}}\right) \end{bmatrix} + v_k \tag{45}$$

5.2. Simulation Condition Settings

The orbital elements of the chief spacecraft are given in Table 4. The simulation parameters are given in Table 5.

Table 4. Orbital elements of chief spacecraft.

Orbital Element	Corresponding Value
Semimajor axis	8200 km
Orbital eccentricity	0.15
Orbital inclination	50°
Right ascension of ascending node	105°
Argument of perigee	10°
True anomaly	30°

Table 5. Simulation parameters.

Simulation Parameter	Corresponding Value
Initial nominal vector	$x_0 = [(10, 15, 10)\text{km}, (0.3, 0.5, 0.3)\text{km/s}]^T$
Control specific force	$f_u = \left[\underbrace{1 \times 10^{-5}, \cdots, 1 \times 10^{-5}}_{3}\right] \text{km/s}^2$
Process noise covariance	$Q_k = \text{diag}\left[\underbrace{1 \times 10^{-11}, \cdots, 1 \times 10^{-11}}_{3}\right] (\text{km/s})^2$
Observation noise covariance	$R_1 = \text{diag}\left[\left(1.1 \times 10^{-2}\text{km}\right)^2, (0.12°)^2, (0.12°)^2\right]$
Dynamics discrete interval, observation update interval	$0.1\text{s}, 0.5\text{s}$
Filters' initial covariance	$P_0 = \text{diag}\left[\underbrace{(0.3)^2, \cdots, (0.3)^2}_{3}\text{km}, \underbrace{(0.1)^2, \cdots, (0.1)^2}_{3}\text{km/s}\right]$
Filters' initial state	$\hat{x}_0 \sim \mathcal{N}(x_0, \hat{P}_0)$
Filters' initial covariance	$R_2 = \alpha^2 R_1$

The non-Gaussian observation noise is modeled as $p(v) = (1 - \varepsilon)p_1 + \varepsilon p_2$, where ε is the perturbing parameter; p_1 and p_2 are the PDFs of the nominal and contaminated Gaussian distributions, respectively; their covariance matrices are denoted as R_1 and R_2, respectively; and R_2 is α^2 times as large as R_1. As elaborated in [27], we use the root mean square error (RMSE) to evaluate estimation accuracy and the averaged normalized estima-

tion error squared (ANEES) to evaluate estimation consistency [32], and the equations of RMSE and ANEES are given as follows:

$$
\begin{aligned}
\text{RMSE}_{k,i} &= \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left(\hat{x}_{k,i}^{m} - x_{k,i}^{m} \right)^2} \\
\text{ANEES}_{k} &= \frac{1}{Mn} \sum_{m=1}^{M} \left(x_{k}^{m} - \hat{x}_{k}^{m} \right)^T \left(\hat{P}_{k}^{m} \right)^{-1} \left(x_{k}^{m} - \hat{x}_{k}^{m} \right)
\end{aligned}
\tag{46}
$$

where M denotes the number of Monte Carlo simulation runs; m denotes the mth Monte Carlo simulation run; n denotes the dimension of the state vector; the subscript k denotes the kth instant; the subscript i denotes the ith component of the state vector; x denotes the nominal state vector; $\hat{x}$ denotes the posterior state estimate; and $\hat{P}$ denotes the posterior covariance estimate. A detailed explanation of the ANEES matrix is demonstrated in [27], and the theoretical value of ANEES is 1. In this work, given the degrees of freedom (DOF) $Mn = 100 \times 6 = 600$ and the significance level of $\alpha = 5\%$, the critical value of the chi-squared test can be approximated as $\chi_{\alpha}^{2}(Mn) \approx \left(z_{\alpha} + \sqrt{2Mn - 1} \right)^2 / 2$, where z_{α} denotes the value to the left of a critical value z, and α denotes the probability mass of the standard normal distribution. In this work, the upper and lower bounds of the 95% confidence interval are computed as $\chi_{\alpha/2}^{2}(Mn)/(Mn) = 0.8893$ and $\chi_{1-\alpha/2}^{2}(Mn)/(Mn) = 1.1155$, respectively, where $z_{0.025} = -1.96$ and $z_{0.975} = 1.96$.

5.3. Comparison Under Gaussian Noise

Figure 4 depicts the RMSEs of the filters based on different functions under Gaussian observation noise. We find that the filters using non-convex functions (Gaussian and DCS functions) diverge, i.e., these filters converge to local minima, whereas the filter using the non-convex function (Cauchy function) converges, due to its weaker robustness compared to the other two non-convex functions, i.e., the Cauchy influence function or weight function does not converge to 0 rapidly in the outlier range, as shown in Figure 1. The convex cost function-based (Huber function) filter converges well, which demonstrates that the convex function is more stable than the non-convex functions. Furthermore, by using the Huber function, the mixed strategies (Huber/MCC, Huber/Cauchy, and Huber/DCS) can avoid the flaws of non-convex functions, improve the filtering accuracy, and prevent non-convex function-based filters from diverging.
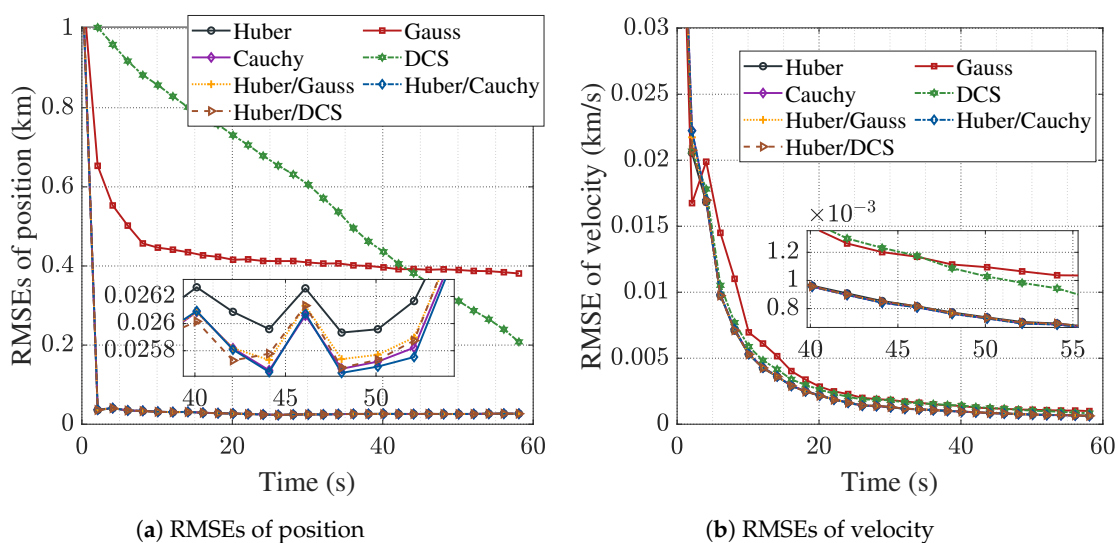

(**a**) RMSEs of position

(**b**) RMSEs of velocity

Figure 4. RMSEs of filters under Gaussian noise ($\varepsilon = 0$).

5.4. Comparison Under Different Levels of Non-Gaussian Noise

Figures 5 and 6 depict the RMSEs of filters based on different functions under different levels of non-Gaussian observation noise. Under the general level of non-Gaussian noise, the convergence of the filters based on different functions is essentially the same as that under Gaussian noise, whereas for the mixed strategies, the estimation accuracy of several filters (Huber, Huber/MCC, Huber/Cauchy, and Huber/DCS) appears to be different. Specifically, the filters using the mixed strategies obtain better estimation accuracies than the Huber-based filter, which fully reveals the strong robustness of the non-convex functions over the convex one. As the level of non-Gaussian noise increases, the flaws of the non-convex functions become more obvious, and, especially, even with the help of the Huber function, the DCS-based filter still falls into the local minimum. Furthermore, with the help of the Huber function, the Huber/Cauchy-based filter achieves better estimation than the Cauchy-based one; with the help of the Cauchy function, the Huber/Cauchy-based filter achieves better estimation than the Huber-based one. The results re-emphasize that the mixed strategy allows the filters to combine both estimation stability and robustness. Figure 7 depicts the ANEESs of filters based on different functions; the mixed strategies allow the filters to obtain better estimation consistency.
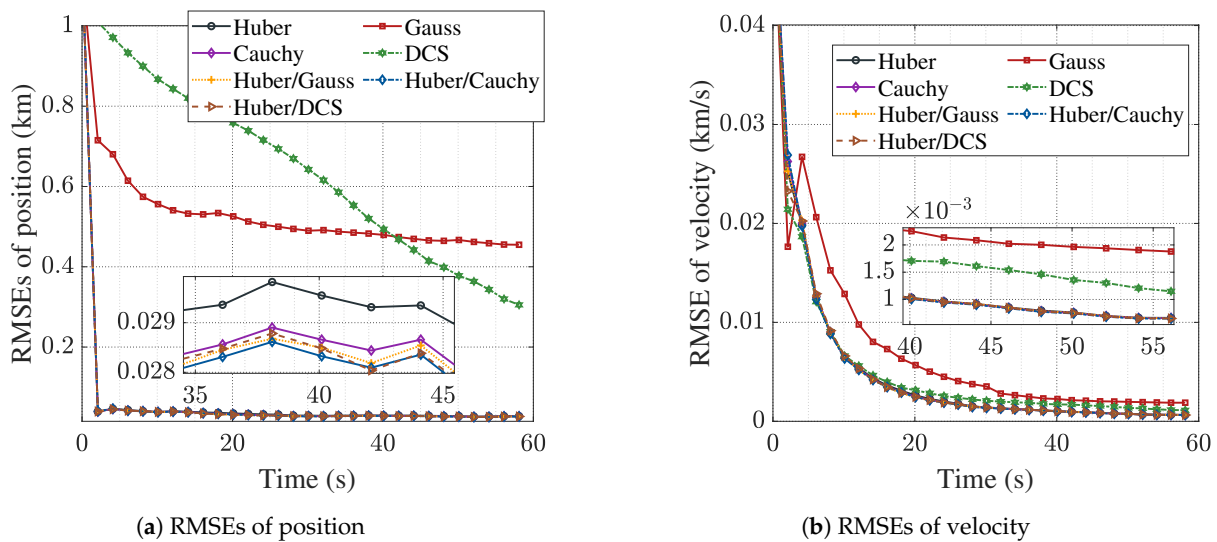

(**a**) RMSEs of position

(**b**) RMSEs of velocity

Figure 5. RMSEs of filters under a general level of non-Gaussian noise ($\varepsilon = 0.1, \alpha = 5$).

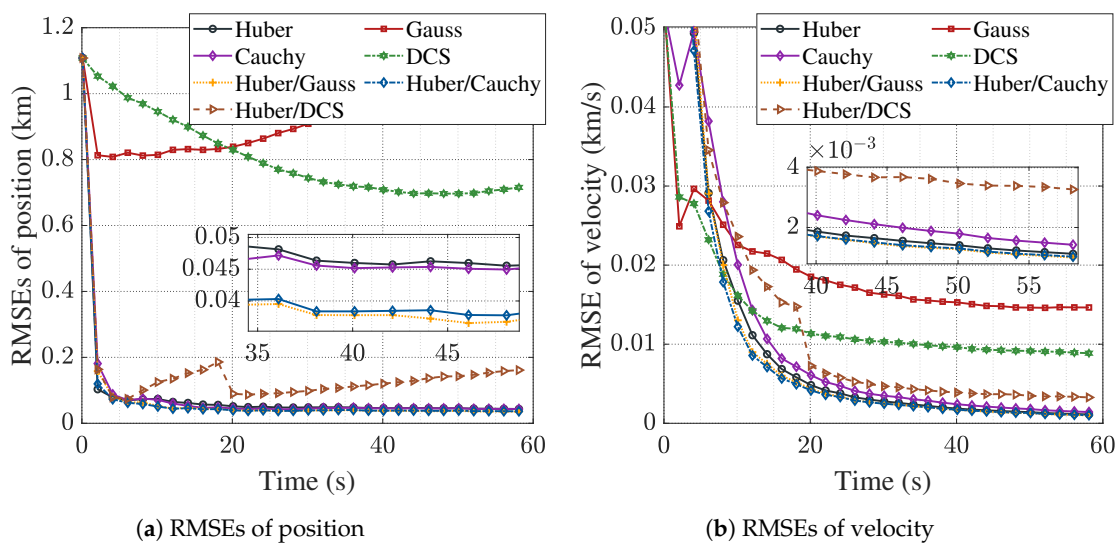(**a**) RMSEs of position

(**b**) RMSEs of velocity

Figure 6. RMSEs of filters under a high level of non-Gaussian noise ($\varepsilon = 0.2, \alpha = 10$).

5.5. Analysis on the Number of Iterations in Sequential Mixed Cost

To elaborate the switching strategy from the convex function to the non-convex function in iterations, 100 Monte Carlo simulation runs are performed under the high level of non-Gaussian noise ($\varepsilon = 0.2, \alpha = 10$). The mixed cost strategies involved in the comparison are Huber/Gauss, Huber/Cauchy, and Huber/DCS, respectively. The total number of iterations are all set to four. For the mixed cost strategies, the number of Huber-based iterations are set to one, two, and three, respectively, and then the corresponding number of non-convex-based iterations are set as three, two, and one, respectively. The above algorithms are abbreviated as Huber1/Gauss3, Huber1/Cauchy3, Huber1/DCS3, Huber2/Gauss2, Huber2/Cauchy2, Huber2/DCS2, Huber3/Gauss1, Huber3/Cauchy1, and Huber3/DCS1, respectively.

(**a**) General level ($\varepsilon = 0.1, \alpha = 5$)　　(**b**) High level ($\varepsilon = 0.2, \alpha = 10$)
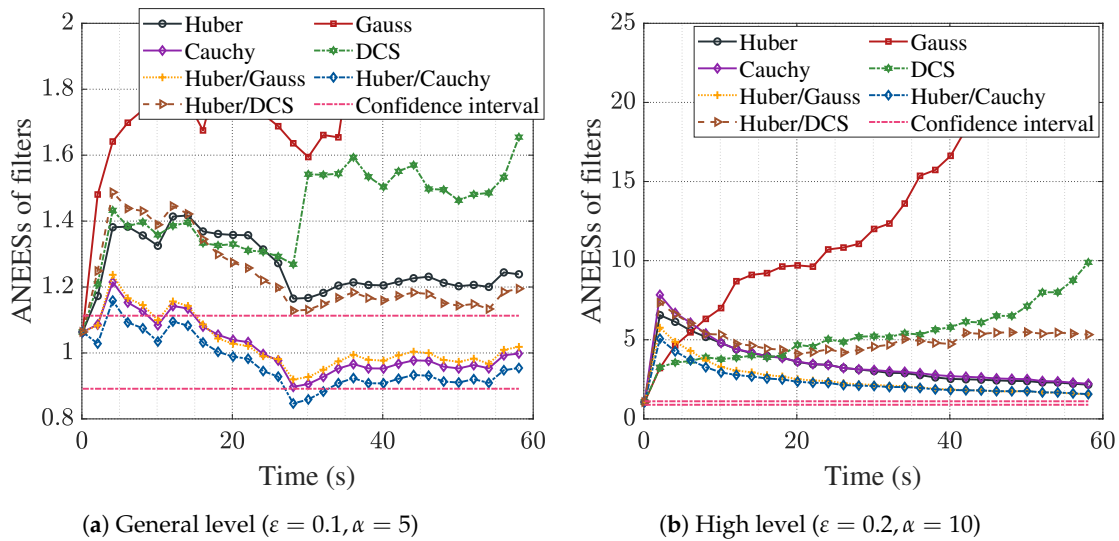
Figure 7. ANEESs of filters under different levels of non-Gaussian noise.

Figures 8 and 9 depict the RMSEs and ANEESs of filters based on the mixed cost functions using different numbers of iterations, respectively. As shown in Figure 8, we find that all filters using the DCS function diverge (Huber1/DCS3, Huber2/DCS2, and Huber3/DCS1), and for the Gaussian function, only the filter using one Gaussian iteration converges (Huber3/Gauss1). On the contrary, all filters using the Cauchy function converge (Huber1/Cauchy3, Huber2/Cauchy2, and Huber3/Cauchy1), and the estimation accuracy and consistency of the filters gradually are improved as the number of Cauchy-based iterations decreases and the number of Huber-based iterations increases. The same results can be seen from the consistency metrics of the filters in Figure 9. Therefore, we can conclude that in the sequential mixed cost strategy, as many Huber-based iterations as possible should be performed to obtain relatively accurate estimates, and a non-convex function should be used only in the last iteration to obtain stronger robustness. Furthermore, the Cauchy-based filters achieve better robustness and stability compared with the other two non-convex functions; the reason is that the Cauchy influence function or weight function does not converge to 0 rapidly in the outlier range compared with the Gaussian and DCS functions. Therefore, we can conclude that the Cauchy function is the better choice for the non-convex iteration.
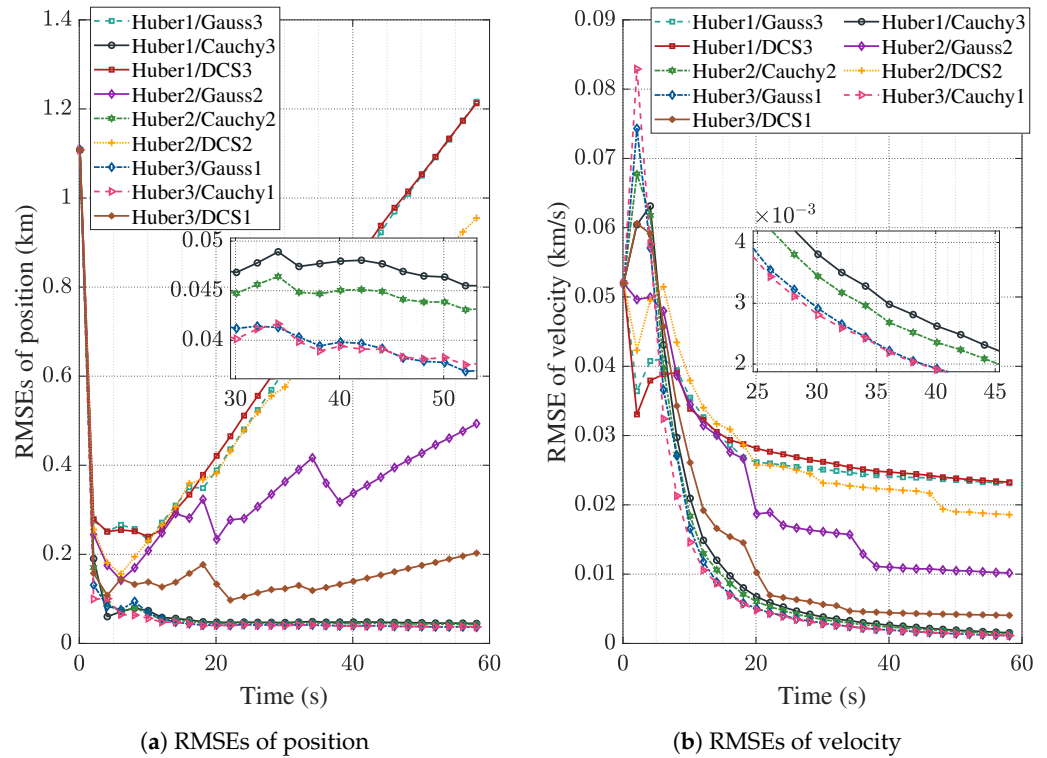
(**a**) RMSEs of position

(**b**) RMSEs of velocity

Figure 8. RMSEs of filters using different numbers of iterations ($\varepsilon = 0.2, \alpha = 10$).
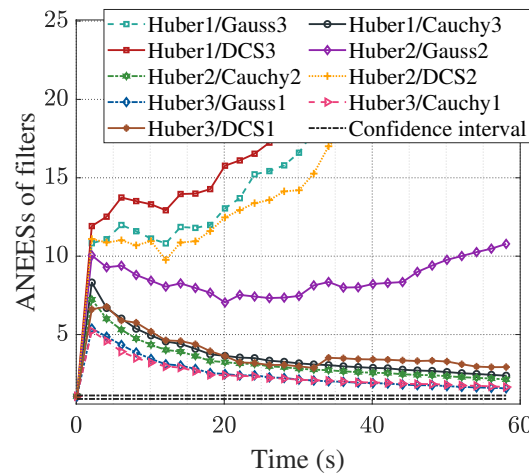

Figure 9. ANEESs of filters using different numbers of iterations ($\varepsilon = 0.2, \alpha = 10$).

6. Discussion

To address the problem induced by non-Gaussian observation noise, the update step of the Gaussian filters is modified based on M-estimation. The robustness of such robust filters depends on the cost criterion adopted, where the more extensively used cost function is the Huber function. As demonstrated in Section 3.1, the Huber function is non-redescending; this property results in relatively weak robustness to cope with the high level of non-Gaussian noise. To improve the robustness of the M-estimation type robust Kalman filter, the redescending cost functions (e.g., the Gaussian function, the DCS function, and the Cauchy function) emerge. However, the flaws of the redescending cost function are also obvious. As discussed in Section 3.2, the redescending functions are non-convex and tend to induce the estimation to fall into local minima; this flaw is also proved by the numerical simulation in Section 5. Regarding the perspective of the stability property, the convex functions (e.g., the Huber function) have an advantage over the non-convex functions (e.g., the Gaussian function, the DCS function, and the Cauchy function).

To synthesize the advantages of the two types of robust cost functions, this work proposes a sequential mixed cost strategy using both the non-convex and convex cost functions. Specifically, an iterative update is required for the robust estimator, where the convex cost function is applied until the filter converges in the first few iterations, and then the non-convex cost function is applied in the subsequent iterations to eliminate outliers. The iterative strategy is given based on the iterated sigma point Kalman filtering framework.

The simulation results in Section 5.4 demonstrate that the convex cost function facilitates the non-convex function to achieve more stability, and the non-convex function facilitates the convex function to achieve stronger robustness. The best combination is the Huber and Cauchy mixed cost strategy. The reason lies in that the Cauchy influence function or weight function does not converge to zero rapidly in the outlier range compared with the Gaussian and DCS functions. Therefore, we can conclude that the sequential mixed cost strategy facilitates the M-estimation type robust Kalman filters to cope with different levels of non-Gaussian noise. Furthermore, the simulation results in Section 5.5 demonstrate that as many Huber-based iterations as possible should be performed to obtain relatively accurate estimates, and a non-convex-based iteration should be used only in the last iteration to obtain stronger robustness. For future work, we will perform research on more types of convex function and non-convex function mixed cost strategies, and on the effect of iterative strategies on estimation robustness and stability.

7. Conclusions

This work investigates the robustness and stability of different cost functions in M-estimation. Due to non-redescent, the convex cost function exhibits strong stability and weak robustness, and due to redescent, the non-convex cost function exhibits weak stability and strong robustness. To combine the properties of different cost functions, we propose the sequential mixed cost strategy and provide the robust iterated sigma point information filtering framework. The simulation results show that, under a general level of non-Gaussian noise, the mixed strategy avoids the non-convex function-based estimation falling into the local minimum; under a high level of non-Gaussian noise, the mixed strategy provides more accurate estimates; in the sequential mixed cost strategy, as many Huber-based iterations as possible should be performed to obtain relatively accurate estimates, and a non-convex function should be used only in the last iteration to obtain stronger robustness. Therefore, the mixed strategy can comprehensively improve the efficiency of the M-estimation type iterated filter.

References

1. Liao, T.; Hirota, K.; Wu, X.; Shao, S.; Dai, Y. A dynamic self-tuning maximum correntropy Kalman filter for wireless sensors networks positioning systems. *Remote Sens.* **2022**, *14*, 4345. [CrossRef]
2. Cui, B.; Chen, W.; Weng, D.; Wei, X.; Sun, Z.; Zhao, Y.; Liu, Y. Observability-constrained resampling-rree cubature Kalman filter for GNSS/INS with measurement outliers. *Remote Sens.* **2023**, *15*, 4591. [CrossRef]
3. Liu, D.; Chen, X.; Xu, Y.; Liu, X.; Shi, C. Maximum correntropy generalized high-degree cubature Kalman filter with application to the attitude determination system of missile. *Aerosp. Sci. Technol.* **2019**, *95*, 105441. [CrossRef]

4. Huber, P.J. Robust estimation of a location parameter. In *Breakthroughs in Statistics: Methodology and Distribution*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 492–518.

5. Karlgaard, C.D.; Schaub, H. Huber-based divided difference filtering. *J. Guid. Control. Dyn.* **2007**, *30*, 885–891. [CrossRef]

6. Chang, L.; Hu, B.; Chang, G.; Li, A. Huber-based novel robust unscented Kalman filter. *IET Sci. Meas. Technol.* **2012**, *6*, 502–509. [CrossRef]

7. Huang, Y.; Zhang, Y.; Li, N.; Wu, Z.; Chambers, J.A. A novel robust Student's t-based Kalman filter. *IEEE Trans. Aerosp. Electron. Syst.* **2017**, *53*, 1545–1554. [CrossRef]

8. Chen, B.; Liu, X.; Zhao, H.; Principe, J.C. Maximum correntropy Kalman filter. *Automatica* **2017**, *76*, 70–77. [CrossRef]

9. Wang, G.; Zhang, Y.; Wang, X. Iterated maximum correntropy unscented Kalman filters for non-Gaussian systems. *Signal Process.* **2019**, *163*, 87–94. [CrossRef]

10. Huai, L.; Li, B.; Yun, P.; Song, C.; Wang, J. Weighted Maximum Correntropy Criterion-Based Interacting Multiple-Model Filter for Maneuvering Target Tracking. *Remote Sens.* **2023**, *15*, 4513. [CrossRef]

11. Wang, D.; Zhang, H.; Huang, H.; Ge, B. A redundant measurement-based maximum correntropy extended Kalman filter for the noise covariance estimation in INS/GNSS integration. *Remote Sens.* **2023**, *15*, 2430. [CrossRef]

12. MacTavish, K.; Barfoot, T.D. At all costs: A comparison of robust cost functions for camera correspondence outliers. In Proceedings of the 2015 12th Conference on Computer and Robot Vision, Halifax, NS, Canada, 3–5 June 2015; pp. 62–69.

13. Chen, B.; Wang, X.; Lu, N.; Wang, S.; Cao, J.; Qin, J. Mixture correntropy for robust learning. *Pattern Recognit.* **2018**, *79*, 318–327. [CrossRef]

14. Dang, L.; Huang, Y.; Zhang, Y.; Chen, B. Multi-kernel correntropy based extended Kalman filtering for state-of-charge estimation. *ISA Trans.* **2022**, *129*, 271–283. [CrossRef] [PubMed]

15. Chen, B.; Xie, Y.; Li, Z.; Li, Y.; Ren, P. Asymmetric correntropy for robust adaptive filtering. *IEEE Trans. Circuits Syst. II Express Briefs* **2021**, *69*, 1922–1926. [CrossRef]

16. Qu, H.; Wang, M.; Zhao, J.; Zhao, S.; Li, T.; Yue, P. Generalized asymmetric correntropy for robust adaptive filtering: A theoretical and simulation study. *Remote Sens.* **2022**, *14*, 3677. [CrossRef]

17. Qi, L.; Shen, M.; Wang, D.; Wang, S. Robust Cauchy kernel conjugate gradient algorithm for non-Gaussian noises. *IEEE Signal Process. Lett.* **2021**, *28*, 1011–1015. [CrossRef]

18. Huang, H.; Zhang, H. Student's t-Kernel-Based Maximum Correntropy Kalman Filter. *Sensors* **2022**, *22*, 1683. [CrossRef]

19. Li, S.; Cui, N.; Mu, R. Dynamic-covariance-scaling-based robust sigma-point information filtering. *J. Guid. Control. Dyn.* **2021**, *44*, 1677–1684. [CrossRef]

20. Agarwal, P. Robust graph-based localization and mapping. Ph.D. Thesis, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany, 2015.

21. Agarwal, P.; Tipaldi, G.D.; Spinello, L.; Stachniss, C.; Burgard, W. Robust map optimization using dynamic covariance scaling. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; pp. 62–69.

22. Zhang, Z. Parameter estimation techniques: A tutorial with application to conic fitting. *Image Vis. Comput.* **1997**, *15*, 59–76. [CrossRef]

23. Wang, X.; Cui, N.; Guo, J. Huber-based unscented filtering and its application to vision-based relative navigation. *IET Radar Sonar Navig.* **2010**, *4*, 134–141. [CrossRef]

24. Zhan, R.; Wan, J. Iterated unscented Kalman filter for passive target tracking. *IEEE Trans. Aerosp. Electron. Syst.* **2007**, *43*, 1155–1163. [CrossRef]

25. Karlgaard, C.D. Nonlinear regression Huber-Kalman filtering and fixed-interval smoothing. *J. Guid. Control. Dyn.* **2015**, *38*, 322–330. [CrossRef]

26. Wang, G.; Li, N.; Zhang, Y. Distributed maximum correntropy linear and nonlinear filters for systems with non-Gaussian noises. *Signal Process.* **2021**, *182*, 107937. [CrossRef]

27. Li, S.; Zhang, X.; Liu, W.; Cui, N. Optimization-based iterative and robust strategy for spacecraft relative navigation in elliptical orbit. *Aerosp. Sci. Technol.* **2023**, *133*, 108138. [CrossRef]

28. Wang, Y.; Zheng, W.; Sun, S.; Li, L. Robust information filter based on maximum correntropy criterion. *J. Guid. Control. Dyn.* **2016**, *39*, 1126–1131. [CrossRef]

29. Liu, X.; Chen, B.; Xu, B.; Wu, Z.; Honeine, P. Maximum correntropy unscented filter. *Int. J. Syst. Sci.* **2017**, *48*, 1607–1615. [CrossRef]

30. De Menezes, D.; Prata, D.M.; Secchi, A.R.; Pinto, J.C. A review on robust M-estimators for regression analysis. *Comput. Chem. Eng.* **2021**, *147*, 107254. [CrossRef]

31. Sibley, G.; Sukhatme, G.S.; Matthies, L.H. The iterated sigma point Kalman filter with applications to long range stereo. *Robot. Sci. Syst.* **2006**, *8*, 235–244.

32. Crassidis, J.L.; Junkins, J.L., Optimal estimation of dynamic systems; CRC Press, Taylor and Francis Group: Boca Raton, FL, USA, 2004; pp. 228–231.