Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

A novel multiple-outlier-robust Kalman filter*

Yulong HUANG, Mingming BAI, Yonggang ZHANG‡

College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China

E-mail: heuedu@163.com; mingming.bai@hrbeu.edu.cn; zhangyg@hrbeu.edu.cn

Received Nov. 17, 2020; Revision accepted Mar. 1, 2021; Crosschecked Oct. 9, 2021; Published online Jan. 24, 2022

Abstract: This paper presents a novel multiple-outlier-robust Kalman filter (MORKF) for linear stochastic discrete-time systems. A new multiple statistical similarity measure is first proposed to evaluate the similarity between two random vectors from dimension to dimension. Then, the proposed MORKF is derived via maximizing a multiple statistical similarity measure based cost function. The MORKF guarantees the convergence of iterations in mild conditions, and the boundedness of the approximation errors is analyzed theoretically. The selection strategy for the similarity function and comparisons with existing robust methods are presented. Simulation results show the advantages of the proposed filter.

Key words: Kalman filtering; Multiple statistical similarity measure; Multiple outliers; Fixed-point iteration; State estimate

https://doi.org/10.1631/FITEE.2000642 **CLC number:** TP273

1 Introduction

The Kalman filter (KF) has played an important role in many engineering fields such as navigation, positioning, target tracking, control, and communications (Simon, 2006). The outlier interference problem often occurs in these applications because of unreliable sensor measurements, external disturbances, and unknown modeling errors. In general, a linear state-space model for such an outlier-corrupted state estimate problem can be formulated as follows:

$$\begin{cases} \boldsymbol{x}_k = \boldsymbol{F}_k \boldsymbol{x}_{k-1} + \boldsymbol{w}_k, & \text{(state equation)} \\ \boldsymbol{z}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{v}_k, & \text{(measurement equation)} \end{cases} \tag{1}$$

where k denotes the discrete time index, $\boldsymbol{x}_k \in \mathbb{R}^n$ and $\boldsymbol{z}_k \in \mathbb{R}^m$ denote the state and measurement vectors, respectively, $\boldsymbol{F}_k \in \mathbb{R}^{n\times n}$ and $\boldsymbol{H}_k \in \mathbb{R}^{m\times n}$ represent the state transition and measurement matrices, respectively, and $\boldsymbol{w}_k \in \mathbb{R}^n$ and $\boldsymbol{v}_k \in \mathbb{R}^m$ denote the outlier-contaminated state and measurement noise vectors, respectively, both of which have non-Gaussian heavy-tailed distributions. Unfortunately, for an outlier-contaminated linear system, the optimality of the classical KF is violated, and its filtering performance degrades remarkably.

To solve this problem, many efforts have been made to improve the robustness of the classical KF. By using the influence function approach in the KF, a series of M-estimators has been constructed by minimizing the well-chosen robust cost function, in which the Huber KF (HKF) serves as the best known M-estimator (Huber, 2011). An alternative robust M-estimate method, named the maximum correntropy KF (MCKF), has also been proposed by maximizing the correntropy of the predictive error and the residual error (Chen et al., 2017). To further use the heavy-tailed features inherent in the outlier-contaminated noise, many robust filters have been proposed based on non-Gaussian distribution modeling (Ting et al., 2007; Huang et al., 2016, 2019c; Roth et al., 2017), in which the robust Student's

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61903097 and 61773133)

ORCID: Yulong HUANG, https://orcid.org/0000-0001-9303-9083; Mingming BAI, https://orcid.org/0000-0002-4790-8791; Yonggang ZHANG, https://orcid.org/0000-0003-4548-1111

t based KF (RSTKF) (Piché et al., 2012; Huang et al., 2017, 2019a, 2019b) acts as a typical example. Recently, a novel statistical similarity measure based Kalman filtering (SSMKF) framework has been proposed (Huang et al., 2020) which maximizes a statistical similarity measure based cost function. The SSMKF provides a general solution for outlier-contaminated linear systems, and it includes the popular RSTKF as its special case when a logarithmic similarity function is selected (Huang et al., 2020).

In a state-space model, the state and measurement variables often vary from dimension to dimension. For example, in a target tracking problem, the state variables of position and velocity have different magnitudes and propagation features, and the measurement variables suffer from different external disturbances. Consequently, the outliers of different state and measurement variables should indeed be different in intensity and occurrence probability. As a result, the outliers occurring in different state and measurement dimensions may possess different statistical properties in practical applications, and are therefore named multiple outliers in this study. The newly emerging SSMKF is incapable of addressing multiple outliers because it was designed based on an assumption that the outliers occurring in different state and measurement dimensions have the same statistical properties (Huang et al., 2020). To some extent, the existing M-estimator can reduce the effects of multiple outliers (Huber, 2011), but the randomness inherent in the state vector is neglected, which limits its estimation accuracy. The main aspects of the methods mentioned above and the proposed filter are summarized in Table 1.

In this paper, we present a novel multiple-outlier-robust KF (MORKF) for linear stochastic discrete-time systems. A new multiple statistical similarity measure (MSSM) is first proposed to evaluate the similarity between two random vectors from dimension to dimension. The MORKF is developed by maximizing an MSSM-based cost function. Convergence is guaranteed under mild conditions and the rationality of assumptions is discussed. The similarity function selections and comparisons with existing robust KFs are also presented. Simulation results illustrate that the developed MORKF has improved estimation accuracy but with heavier computational burden than the existing HKF, MCKF, and SSMKF. Table 2 presents the acronyms and nomenclature that are used in this paper.

2 Proposed multiple-outlier-robust Kalman filter

2.1 Proposed multiple statistical similarity measure

Different from Huang et al. (2020), in this study we focus on evaluating the similarity between two random vectors, denoted as $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, from dimension to dimension. Hence, a novel MSSM is proposed and formulated as follows:

$$s(\boldsymbol{\alpha},\boldsymbol{\beta})=\sum_{i=1}^{p}\int\int f((\alpha_i-\beta_i)^2)p(\boldsymbol{\alpha},\boldsymbol{\beta})\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{\beta}, \quad (2)$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ denote two p-dimensional random vectors, and α_i and β_i are the i^{th} elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively. $f(\cdot)$ denotes the similarity function, which satisfies the following conditions: (1) $f(\cdot)$ is continuous and differentiable on $[0,+\infty)$; (2) $\dot{f}(l)<0$ on $[0,+\infty)$; (3) $\ddot{f}(l)\geq 0$ on $[0,+\infty)$.

Table 1 Summary of the main points of the existing methods and the proposed filter

Filter	Core technique	Optimization method	Deficiency
HKF (Huber, 2011)	Huber cost function	Gauss–Newton minimization	Robust, but with poor accuracy
MCKF (Chen et al., 2017)	Correntropy cost function	Gauss–Newton maximization	Robust, but with poor accuracy
RSTKF (Huang et al., 2017)	Student's t distribution modeling	Variational Bayesian approach	Incapable of addressing the multiple outliers
SSMKF (Huang et al., 2020)	Statistical similarity measure based cost function	Lower bound maximization	Incapable of addressing the multiple outliers
Proposed MORKF	MSSM-based cost function	Lower bound maximization	Minor increase of the computational burden

MSSM: multiple statistical similarity measure

Table 2 Acronyms and nomenclature

Notation	Definition
$s(\cdot,\cdot)$	Proposed MSSM
$p(\boldsymbol{x},\boldsymbol{y})$	Joint PDF of $\boldsymbol{x}$ and $\boldsymbol{y}$
$f(\cdot)$	Similarity function
$\dot{f}(\cdot)$	First-order derivative of $f(\cdot)$
$\ddot{f}(\cdot)$	Second-order derivative of $f(\cdot)$
$E[\cdot]$	Expectation operation
Var$[\cdot]$	Variance operation
$\boldsymbol{I}_m$	Identity matrix of dimension $m\times m$
$N(\cdot;\bar{\boldsymbol{x}},\boldsymbol{C})$	Multivariate Gaussian PDF with mean vector $\bar{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{C}$
$\|\cdot\|_{\mathcal{F}}$	Frobenius norm of a matrix
M^*	Optimal solution of variable M

MSSM: multiple statistical similarity measure; PDF: probability density function

It is obvious from Eq. (2) that the proposed MSSM satisfies $s(\boldsymbol{\alpha},\boldsymbol{\beta})=s(\boldsymbol{\beta},\boldsymbol{\alpha})$, which means that the proposed MSSM is a symmetric measure for the two evaluated random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The second condition indicates that the MSSM $s(\boldsymbol{\alpha},\boldsymbol{\beta})$ increases monotonously as the difference between the two evaluated random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ decreases, and vice versa.

Proposition 1 The proposed MSSM achieves the unique maximum point if and only if $\boldsymbol{\alpha}=\boldsymbol{\beta}$.

The proof of Proposition 1 is given in Appendix A.

The proposed MSSM is a generalized form for the existing similarity measures. For instance, the proposed MSSM $s(\boldsymbol{\alpha},\boldsymbol{\beta})$ becomes the negative form of the well-known mean squared error (MSE) measure when $f(\cdot)$ is selected as $f(l)=-l$. In addition, the proposed MSSM $s(\boldsymbol{\alpha},\boldsymbol{\beta})$ is identical to the existing correntropy measure when $f(\cdot)$ is selected as $f(l)=\sigma^2\exp(\frac{1-l}{2\sigma^2})$ (Chen et al., 2017). The MSSM can be diverse when various similarity functions are selected, and thus different MORKFs can be constructed by maximizing the corresponding MSSM-based cost function.

Remark 1 The statistical similarity measure proposed in Huang et al. (2020) is used to evaluate the overall similarity between two random vectors, which makes it suitable for detecting the outliers with the same statistical properties in different dimensions. In contrast, the proposed MSSM in this study can be employed to evaluate the separate similarity between two random vectors from dimension to dimension, and therefore it is more suitable for detecting multiple outliers as compared with the previous statistical similarity measure in Huang et al. (2020).

2.2 Design of the proposed multiple-outlier-robust Kalman filter

The core design of the proposed MORKF is to look for an optimal posterior PDF $q^*(\boldsymbol{x}_k)$ via maximizing the MSSM-based cost function as follows:

$$q^*(\boldsymbol{x}_k)=\arg\max_{q(\boldsymbol{x}_k)}\Big\{s\left(\boldsymbol{S}_{k|k-1}^{-1}\boldsymbol{x}_k,\boldsymbol{S}_{k|k-1}^{-1}\hat{\boldsymbol{x}}_{k|k-1}\right)+s\left(\boldsymbol{S}_{\boldsymbol{R}_k}^{-1}\boldsymbol{z}_k,\boldsymbol{S}_{\boldsymbol{R}_k}^{-1}\boldsymbol{H}_k\boldsymbol{x}_k\right)\Big\},\tag{3}$$

where $\boldsymbol{S}_{k|k-1}$ and $\boldsymbol{S}_{\boldsymbol{R}_k}$ are the square root matrices of the nominal predictive error covariance matrix (PECM) $\boldsymbol{P}_{k|k-1}$ and the nominal measurement noise covariance matrix (MNCM) $\boldsymbol{R}_k$, respectively, i.e.,

$$\boldsymbol{P}_{k|k-1}=\boldsymbol{S}_{k|k-1}\boldsymbol{S}_{k|k-1}^{\mathrm{T}},\ \boldsymbol{R}_k=\boldsymbol{S}_{\boldsymbol{R}_k}\boldsymbol{S}_{\boldsymbol{R}_k}^{\mathrm{T}}.\tag{4}$$

The predictive mean vector $\hat{\boldsymbol{x}}_{k|k-1}$ and the nominal PECM are respectively given by

$$\hat{\boldsymbol{x}}_{k|k-1}=\boldsymbol{F}_k\hat{\boldsymbol{x}}_{k-1|k-1},\tag{5}$$

$$\boldsymbol{P}_{k|k-1}=\boldsymbol{F}_k\boldsymbol{P}_{k-1|k-1}\boldsymbol{F}_k^{\mathrm{T}}+\boldsymbol{Q}_k,\tag{6}$$

where $\hat{\boldsymbol{x}}_{k-1|k-1}$ and $\boldsymbol{P}_{k-1|k-1}$ are the posterior mean vector and posterior covariance matrix at the previous time respectively, and $\boldsymbol{Q}_k$ is the nominal state noise covariance matrix (SNCM).

Because the predictive mean vector $\hat{\boldsymbol{x}}_{k|k-1}$ can be calculated using Eq. (5) and the measurement vector $\boldsymbol{z}_k$ can be obtained from external sensors, these two quantities $\hat{\boldsymbol{x}}_{k|k-1}$ and $\boldsymbol{z}_k$ are totally known in the design of the proposed MORKF. As a result, the joint PDFs in Eq. (3) are marginalized, and the MSSM-based cost function can be rewritten as follows:

$$q^*(\boldsymbol{x}_k)=\arg\max_{q(\boldsymbol{x}_k)}\left\{\sum_{i=1}^{n}\int f_x([\boldsymbol{T}_{ki}(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k-1})]^2)\cdot q(\boldsymbol{x}_k)\mathrm{d}\boldsymbol{x}_k+\sum_{j=1}^{m}\int f_z([\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{x}_k)]^2)q(\boldsymbol{x}_k)\mathrm{d}\boldsymbol{x}_k\right\},\tag{7}$$

where $f_x(\cdot)$ and $f_z(\cdot)$ are the similarity functions with respect to state and measurement vectors respectively, and $\boldsymbol{T}_{ki}$ and $\boldsymbol{U}_{kj}$ are the column vectors of the inverse square root matrices $\boldsymbol{S}_{k|k-1}^{-1}$ and $\boldsymbol{S}_{\boldsymbol{R}_k}^{-1}$ respectively, which are formulated as follows:

$$\boldsymbol{S}_{k|k-1}^{-1}=\left[\boldsymbol{T}_{k1}^{\mathrm{T}},\boldsymbol{T}_{k2}^{\mathrm{T}},\ldots,\boldsymbol{T}_{kn}^{\mathrm{T}}\right]^{\mathrm{T}},\tag{8}$$

$$\boldsymbol{S}_{\boldsymbol{R}_k}^{-1}=\left[\boldsymbol{U}_{k1}^{\mathrm{T}},\boldsymbol{U}_{k2}^{\mathrm{T}},\ldots,\boldsymbol{U}_{km}^{\mathrm{T}}\right]^{\mathrm{T}}.\tag{9}$$

Due to the non-Gaussianity of the posterior PDF, it is not feasible to solve the maximization problem (7) analytically. To overcome this difficulty and obtain an approximate solution for Eq. (7), a heuristic idea is to assume the posterior PDF as Gaussian, namely, $q(\boldsymbol{x}_k) \approx N(\boldsymbol{x}_k;\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$. Based on this, the original MSSM-based cost function in Eq. (7) can be solved analytically by maximizing the lower bound. Employing Jensen's inequality, the right-hand terms of Eq. (7) have the following lower bounds:

$$\begin{aligned}&\int f_x([\boldsymbol{T}_{ki}(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k-1})]^2)q(\boldsymbol{x}_k)\mathrm{d}\boldsymbol{x}_k\\ \geq& f_x\left(\int[\boldsymbol{T}_{ki}(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k-1})]^2q(\boldsymbol{x}_k)\mathrm{d}\boldsymbol{x}_k\right),\end{aligned}\tag{10}$$

$$\begin{aligned}&\int f_z([\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{x}_k)]^2)q(\boldsymbol{x}_k)\mathrm{d}\boldsymbol{x}_k\\ \geq& f_z\left(\int[\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{x}_k)]^2q(\boldsymbol{x}_k)\mathrm{d}\boldsymbol{x}_k\right).\end{aligned}\tag{11}$$

Using inequalities (10) and (11) in Eq. (7) and the Gaussian assumption for the posterior PDF, the maximization problem is further converted to

$$\{\boldsymbol{\mu}_k^*,\boldsymbol{\Sigma}_k^*\}\approx\arg\max_{\{\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k\}}J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\quad\text{s.t.}\quad\boldsymbol{\Sigma}_k>\boldsymbol{0},\tag{12}$$

where $\boldsymbol{\mu}_k^*$ denotes the optimal posterior mean vector and $\boldsymbol{\Sigma}_k^*$ denotes the optimal posterior covariance matrix. Rearranging the lower bounds of Eq. (7), the approximated cost function $J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ is given by

$$J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)=\sum_{i=1}^{n}f_x\left(\boldsymbol{T}_{ki}\boldsymbol{A}_k\boldsymbol{T}_{ki}^{\mathrm{T}}\right)+\sum_{j=1}^{m}f_z\left(\boldsymbol{U}_{kj}\boldsymbol{B}_k\boldsymbol{U}_{kj}^{\mathrm{T}}\right),\tag{13}$$

and the auxiliary matrices $\boldsymbol{A}_k$ and $\boldsymbol{B}_k$ are calculated as

$$\boldsymbol{A}_k=\boldsymbol{\Sigma}_k+(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}},\tag{14}$$

$$\boldsymbol{B}_k=(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k)(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k)^{\mathrm{T}}+\boldsymbol{H}_k\boldsymbol{\Sigma}_k\boldsymbol{H}_k^{\mathrm{T}}.\tag{15}$$

Before presenting the main results, to aid in the development, a cluster of intermediate variables and matrices are defined as follows:

$$\begin{cases}\alpha_{ki}^*=-2\dot{f}_x(\boldsymbol{T}_{ki}\boldsymbol{A}_k^*\boldsymbol{T}_{ki}^{\mathrm{T}}),\\ \beta_{kj}^*=-2\dot{f}_z(\boldsymbol{U}_{kj}\boldsymbol{B}_k^*\boldsymbol{U}_{kj}^{\mathrm{T}}),\\ \tilde{\alpha}_{ki}^*=2\ddot{f}_x(\boldsymbol{T}_{ki}\boldsymbol{A}_k^*\boldsymbol{T}_{ki}^{\mathrm{T}}),\\ \tilde{\beta}_{kj}^*=2\ddot{f}_z(\boldsymbol{U}_{kj}\boldsymbol{B}_k^*\boldsymbol{U}_{kj}^{\mathrm{T}}),\end{cases}\tag{16}$$

$$\begin{cases}\boldsymbol{\Lambda}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)=\frac{\partial J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\partial\boldsymbol{\mu}_k},\\ \boldsymbol{\Lambda}_{\boldsymbol{\Sigma}_k}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)=\frac{\partial J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\partial\boldsymbol{\Sigma}_k},\\ \boldsymbol{\Pi}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)=\frac{\partial^2 J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\partial\boldsymbol{\mu}_k\partial\boldsymbol{\mu}_k^{\mathrm{T}}},\end{cases}\tag{17}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ and $\boldsymbol{\Lambda}_{\boldsymbol{\Sigma}_k}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ denote the Jacobian matrices of the cost function $J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ with respect to the posterior mean vector $\boldsymbol{\mu}_k$ and the posterior covariance matrix $\boldsymbol{\Sigma}_k$ respectively, and $\boldsymbol{\Pi}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ denotes the Hessian matrix of the cost function $J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ with respect to $\boldsymbol{\mu}_k$.

Theorem 1 By solving the maximization problem (12), the optimal posterior mean vector $\boldsymbol{\mu}_k^*$ can be formulated as follows:

$$\tilde{\boldsymbol{K}}_k^*=\tilde{\boldsymbol{P}}_{k|k-1}^*\boldsymbol{H}_k^{\mathrm{T}}\left(\boldsymbol{H}_k\tilde{\boldsymbol{P}}_{k|k-1}^*\boldsymbol{H}_k^{\mathrm{T}}+\tilde{\boldsymbol{R}}_k^*\right)^{-1},\tag{18}$$

$$\boldsymbol{\mu}_k^*=\hat{\boldsymbol{x}}_{k|k-1}+\tilde{\boldsymbol{K}}_k^*(\boldsymbol{z}_k-\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1}),\tag{19}$$

where the modified PECM $\tilde{\boldsymbol{P}}_{k|k-1}^*$ and the modified MNCM $\tilde{\boldsymbol{R}}_k^*$ are respectively given by

$$\tilde{\boldsymbol{P}}_{k|k-1}^*=\boldsymbol{S}_{k|k-1}\left(\boldsymbol{\Psi}_{\boldsymbol{x}_k}^*\right)^{-1}\boldsymbol{S}_{k|k-1}^{\mathrm{T}},\tag{20}$$

$$\tilde{\boldsymbol{R}}_k^*=\boldsymbol{S}_{\boldsymbol{R}_k}\left(\boldsymbol{\Psi}_{\boldsymbol{z}_k}^*\right)^{-1}\boldsymbol{S}_{\boldsymbol{R}_k}^{\mathrm{T}}.\tag{21}$$

The weighted matrices $\boldsymbol{\Psi}_{\boldsymbol{x}_k}^*$ and $\boldsymbol{\Psi}_{\boldsymbol{z}_k}^*$ are formulated as

$$\begin{aligned}\boldsymbol{\Psi}_{\boldsymbol{x}_k}^*=&\mathrm{diag}\left(-2\dot{f}_x(\boldsymbol{T}_{k1}\boldsymbol{A}_k^*\boldsymbol{T}_{k1}^{\mathrm{T}}),-2\dot{f}_x(\boldsymbol{T}_{k2}\boldsymbol{A}_k^*\boldsymbol{T}_{k2}^{\mathrm{T}}),\right.\\ &\left.\ldots,-2\dot{f}_x(\boldsymbol{T}_{kn}\boldsymbol{A}_k^*\boldsymbol{T}_{kn}^{\mathrm{T}})\right),\end{aligned}\tag{22}$$

$$\begin{aligned}\boldsymbol{\Psi}_{\boldsymbol{z}_k}^*=&\mathrm{diag}\left(-2\dot{f}_z(\boldsymbol{U}_{k1}\boldsymbol{B}_k^*\boldsymbol{U}_{k1}^{\mathrm{T}}),-2\dot{f}_z(\boldsymbol{U}_{k2}\boldsymbol{B}_k^*\boldsymbol{U}_{k2}^{\mathrm{T}}),\right.\\ &\left.\ldots,-2\dot{f}_z(\boldsymbol{U}_{km}\boldsymbol{B}_k^*\boldsymbol{U}_{km}^{\mathrm{T}})\right).\end{aligned}\tag{23}$$

The auxiliary matrices are given by

$$\boldsymbol{A}_k^*=\boldsymbol{\Sigma}_k^*+(\boldsymbol{\mu}_k^*-\hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{\mu}_k^*-\hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}},\tag{24}$$

$$\boldsymbol{B}_k^*=(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k^*)(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k^*)^{\mathrm{T}}+\boldsymbol{H}_k\boldsymbol{\Sigma}_k^*\boldsymbol{H}_k^{\mathrm{T}}.\tag{25}$$

The proof of Theorem 1 is given in Appendix B.

Theorem 1 indicates that the optimal posterior mean vector $\boldsymbol{\mu}_k^*$ is indeed a modified Kalman filtering estimate, achieved by using the modified PECM $\tilde{\boldsymbol{P}}_{k|k-1}^*$ and the modified MNCM $\tilde{\boldsymbol{R}}_k^*$ in Eqs. (20) and (21). Also, it is observed from Theorem 1 that the optimal posterior mean vector $\boldsymbol{\mu}_k^*$ depends on the optimal posterior covariance matrix $\boldsymbol{\Sigma}_k^*$. Similar to our previous work (Huang et al., 2020), we can obtain the following two propositions:

Proposition 2 The result of $\boldsymbol{\mu}_k^*$ given in Eq. (19) is a global optimal solution when and only when the following inequalities hold:

$$\begin{cases} (\alpha_{ki}^*)^2 - 2\tilde{\alpha}_{ki}^* > 0, \\ (\beta_{kj}^*)^2 - 2\tilde{\beta}_{kj}^* > 0. \end{cases} \tag{26}$$

The proof of Proposition 2 is given in Appendix C.

Proposition 3 The approximated cost function $J(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is monotonically decreasing with respect to $\boldsymbol{\Sigma}_k$.

The proof of Proposition 3 is given in Appendix D.

Proposition 2 provides a sufficient condition to guarantee that the result $\boldsymbol{\mu}_k^*$ in Eq. (19) is a global optimal solution. Proposition 3 implies that the optimal solution of the posterior covariance matrix $\boldsymbol{\Sigma}_k^*$ should be selected as the lower bound of $\boldsymbol{\Sigma}_k$. Therefore, to guarantee the filtering consistency, the optimal posterior covariance matrix $\boldsymbol{\Sigma}_k^*$ is given by

$$\boldsymbol{\Sigma}_k^* = \left(\boldsymbol{S}_{k|k-1}^{-\mathrm{T}} \boldsymbol{\Psi}_{\boldsymbol{x}_k}^* \boldsymbol{S}_{k|k-1}^{-1} + \boldsymbol{H}_k^{\mathrm{T}} \boldsymbol{S}_{\boldsymbol{R}_k}^{-\mathrm{T}} \boldsymbol{\Psi}_{\boldsymbol{z}_k}^* \boldsymbol{S}_{\boldsymbol{R}_k}^{-1} \boldsymbol{H}_k\right)^{-1}. \tag{27}$$

Substituting Eqs. (20) and (21) into Eq. (27) and using the well-known matrix inversion lemma yield

$$\boldsymbol{\Sigma}_k^* = \tilde{\boldsymbol{P}}_{k|k-1}^* - \tilde{\boldsymbol{K}}_k^* \boldsymbol{H}_k \tilde{\boldsymbol{P}}_{k|k-1}^*. \tag{28}$$

It can be observed from Theorem 1 and Eq. (28) that the optimal posterior mean vector $\boldsymbol{\mu}_k^*$ and the optimal posterior covariance matrix $\boldsymbol{\Sigma}_k^*$ are intercoupled. Then, the fixed-point iteration method is employed to solve $\boldsymbol{\mu}_k^*$ and $\boldsymbol{\Sigma}_k^*$ approximately, from which the proposed MORKF can be implemented, as summarized in Algorithm 1, where ϵ denotes the iteration threshold and N_{v} denotes the maximum number of iterations.

Next, the convergence conditions of the fixed-point iteration will be provided.

Proposition 4 If the iterative initial value $\boldsymbol{\mu}_k^{(0)}$ and the optimal solution $\boldsymbol{\mu}_k^*$ are close enough and the following inequalities hold $\forall l \geq 0$, the fixed-point iteration approach will converge locally:

$$\begin{cases} 0 \leq \ddot{f}_x(l^2) l \leq \theta_1, \\ 0 \leq \ddot{f}_z(l^2) l \leq \theta_2, \end{cases} \tag{29}$$

where θ_1 and θ_2 are arbitrary positive finite real numbers.

The proof of Proposition 4 is given in Appendix E.

Algorithm 1 Implementation of the multiple-outlier-robust Kalman filter (MORKF)

Input: $\hat{\boldsymbol{x}}_{k-1|k-1}$, $\boldsymbol{P}_{k-1|k-1}$, $\boldsymbol{F}_k$, $\boldsymbol{H}_k$, $\boldsymbol{z}_k$, $\boldsymbol{Q}_k$, $\boldsymbol{R}_k$, $f_x(\cdot)$, $f_z(\cdot)$, m, n, ϵ, N_{v}

Output: $\hat{\boldsymbol{x}}_{k|k}$ and $\boldsymbol{P}_{k|k}$

// Time-update

$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{F}_k \hat{\boldsymbol{x}}_{k-1|k-1}$

$\boldsymbol{P}_{k|k-1} = \boldsymbol{F}_k \boldsymbol{P}_{k-1|k-1} \boldsymbol{F}_k^{\mathrm{T}} + \boldsymbol{Q}_k$

// Measurement-update

Initialization: $\boldsymbol{\Psi}_{\boldsymbol{x}_k}^{(0)} = \boldsymbol{I}_n$, $\boldsymbol{\Psi}_{\boldsymbol{z}_k}^{(0)} = \boldsymbol{I}_m$, $\left\{\boldsymbol{T}_{ki} = \boldsymbol{S}_{k|k-1}^{-1}(i,:)\right\}_{i=1}^{n}$, $\left\{\boldsymbol{U}_{kj} = \boldsymbol{S}_{\boldsymbol{R}_k}^{-1}(j,:)\right\}_{j=1}^{m}$

for $i = 0 : N_{\mathrm{v}} - 1$ **do**

// Calculate $\tilde{\boldsymbol{P}}_{k|k-1}^{(i+1)}$ and $\tilde{\boldsymbol{R}}_k^{(i+1)}$

$\tilde{\boldsymbol{P}}_{k|k-1}^{(i+1)} = \boldsymbol{S}_{k|k-1} \left(\boldsymbol{\Psi}_{\boldsymbol{x}_k}^{(i)}\right)^{-1} \boldsymbol{S}_{k|k-1}^{\mathrm{T}}$

$\tilde{\boldsymbol{R}}_k^{(i+1)} = \boldsymbol{S}_{\boldsymbol{R}_k} \left(\boldsymbol{\Psi}_{\boldsymbol{z}_k}^{(i)}\right)^{-1} \boldsymbol{S}_{\boldsymbol{R}_k}^{\mathrm{T}}$

// Calculate $\boldsymbol{\mu}_k^{(i+1)}$ and $\boldsymbol{\Sigma}_k^{(i+1)}$

$\tilde{\boldsymbol{K}}_k^{(i+1)} = \tilde{\boldsymbol{P}}_{k|k-1}^{(i+1)} \boldsymbol{H}_k^{\mathrm{T}} \left(\boldsymbol{H}_k \tilde{\boldsymbol{P}}_{k|k-1}^{(i+1)} \boldsymbol{H}_k^{\mathrm{T}} + \tilde{\boldsymbol{R}}_k^{(i+1)}\right)^{-1}$

$\boldsymbol{\mu}_k^{(i+1)} = \hat{\boldsymbol{x}}_{k|k-1} + \tilde{\boldsymbol{K}}_k^{(i+1)} (\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k-1})$

$\boldsymbol{\Sigma}_k^{(i+1)} = \tilde{\boldsymbol{P}}_{k|k-1}^{(i+1)} - \tilde{\boldsymbol{K}}_k^{(i+1)} \boldsymbol{H}_k \tilde{\boldsymbol{P}}_{k|k-1}^{(i+1)}$

// Compare $\boldsymbol{\mu}_k^{(i+1)}$ and $\boldsymbol{\mu}_k^{(i)}$

if $\left\|\boldsymbol{\mu}_k^{(i+1)} - \boldsymbol{\mu}_k^{(i)}\right\| \Big/ \left\|\boldsymbol{\mu}_k^{(i)}\right\| \leq \epsilon$ **then**

Terminate the iteration

end if

// Calculate $\boldsymbol{A}_k^{(i+1)}$ and $\boldsymbol{B}_k^{(i+1)}$

$\boldsymbol{A}_k^{(i+1)} = \boldsymbol{\Sigma}_k^{(i+1)} + \left(\boldsymbol{\mu}_k^{(i+1)} - \hat{\boldsymbol{x}}_{k|k-1}\right) \cdot \left(\boldsymbol{\mu}_k^{(i+1)} - \hat{\boldsymbol{x}}_{k|k-1}\right)^{\mathrm{T}}$

$\boldsymbol{B}_k^{(i+1)} = \left(\boldsymbol{z}_k - \boldsymbol{H}_k \boldsymbol{\mu}_k^{(i+1)}\right) \left(\boldsymbol{z}_k - \boldsymbol{H}_k \boldsymbol{\mu}_k^{(i+1)}\right)^{\mathrm{T}} + \boldsymbol{H}_k \boldsymbol{\Sigma}_k^{(i+1)} \boldsymbol{H}_k^{\mathrm{T}}$

// Calculate $\boldsymbol{\Psi}_{\boldsymbol{x}_k}^{(i+1)}$ and $\boldsymbol{\Psi}_{\boldsymbol{z}_k}^{(i+1)}$

$\boldsymbol{\Psi}_{\boldsymbol{x}_k}^{(i+1)} = \mathrm{diag}\left(-2\dot{f}_x(\boldsymbol{T}_{k1} \boldsymbol{A}_k^{(i+1)} \boldsymbol{T}_{k1}^{\mathrm{T}}), -2\dot{f}_x(\boldsymbol{T}_{k2} \boldsymbol{A}_k^{(i+1)} \boldsymbol{T}_{k2}^{\mathrm{T}}), \ldots, -2\dot{f}_x(\boldsymbol{T}_{kn} \boldsymbol{A}_k^{(i+1)} \boldsymbol{T}_{kn}^{\mathrm{T}})\right)$

$\boldsymbol{\Psi}_{\boldsymbol{z}_k}^{(i+1)} = \mathrm{diag}\left(-2\dot{f}_z(\boldsymbol{U}_{k1} \boldsymbol{B}_k^{(i+1)} \boldsymbol{U}_{k1}^{\mathrm{T}}), -2\dot{f}_z(\boldsymbol{U}_{k2} \boldsymbol{B}_k^{(i+1)} \boldsymbol{U}_{k2}^{\mathrm{T}}), \ldots, -2\dot{f}_z(\boldsymbol{U}_{km} \boldsymbol{B}_k^{(i+1)} \boldsymbol{U}_{km}^{\mathrm{T}})\right)$

end for

$\hat{\boldsymbol{x}}_{k|k} = \boldsymbol{\mu}_k^{(i+1)}$, $\boldsymbol{P}_{k|k} = \boldsymbol{\Sigma}_k^{(i+1)}$

2.3 Analysis and discussion

2.3.1 Computational complexity analysis

Next, we analyze and compare the computational complexities of the proposed MORKF and the existing SSMKF by calculating the number of floating point operations (NoFPO). Taking the square-root similarity functions as an example, the NoFPOs of some main equations are listed in Table 3.

The fixed-point iteration method is employed

Table 3 NoFPOs of some main equations

Equation	NoFPO_A	NoFPO_B
(5)	$2n^2-n$	0
(6)	$4n^3-n^2$	0
(18)	$4n^2m+4nm^2-3nm$	$O(m^3)$
(19)	$4nm$	0
(20)	$2n^3$	$n+O(n^3)$
(21)	$2m^3$	$m+O(m^3)$
(22)	n	$2n$
(23)	m	$2m$
(24)	$2n^2+2n$	0
(25)	$2n^2m+2m^2n+3mn$	0
(28)	$2n^3+2n^2m-n^2$	0

NoFPO_A denotes the NoFPO of addition/subtraction and multiplication operations; NoFPO_B denotes the NoFPO of division, matrix inversion, Cholesky decomposition, and square-root operations

to implement the proposed MORKF. Assuming that the average fixed-point iteration number is N_{m}, the NoFPO of the proposed MORKF can be calculated according to Table 3:

$$\begin{aligned}&\text{NoFPO}_{\text{MORKF}}=\\&(4N_{\mathrm{m}}+4)n^3+2N_{\mathrm{m}}m^3+8N_{\mathrm{m}}n^2m+6N_{\mathrm{m}}nm^2\\&+(N_{\mathrm{m}}+1)n^2+4N_{\mathrm{m}}nm+(6N_{\mathrm{m}}-1)n+4N_{\mathrm{m}}m\\&+N_{\mathrm{m}}O(n^3)+2N_{\mathrm{m}}O(m^3).\end{aligned}\tag{30}$$

Similarly, the NoFPO of the existing SSMKF (Huang et al., 2020) can be calculated as follows:

$$\begin{aligned}&\text{NoFPO}_{\text{SSMKF}}=\\&(4N_{\mathrm{m}}+4)n^3+2N_{\mathrm{m}}m^3+8N_{\mathrm{m}}n^2m+6N_{\mathrm{m}}nm^2\\&+(N_{\mathrm{m}}+1)n^2+4N_{\mathrm{m}}nm+(3N_{\mathrm{m}}-1)n+N_{\mathrm{m}}m\\&+2N_{\mathrm{m}}+N_{\mathrm{m}}O(n^3)+2N_{\mathrm{m}}O(m^3).\end{aligned}\tag{31}$$

Comparing Eqs. (30) and (31), we find that the computational complexity of the proposed MORKF is moderately greater than that of the existing SSMKF because of the different adjustment styles when facing the outliers. However, note that the weighted matrices $\boldsymbol{\Psi}^*_{\boldsymbol{x}_k}$ and $\boldsymbol{\Psi}^*_{\boldsymbol{z}_k}$ in Eqs. (20) and (21) are diagonal matrices, so it is easy to obtain their inverse matrices. Considering the improved performance in addressing the multiple outliers, the increased computational complexity of the proposed MORKF is acceptable.

2.3.2 Approximation error analysis

Three assumptions are presented to facilitate the derivation of the proposed MORKF:

Assumption 1 The optimal posterior PDF $q^*(\boldsymbol{x}_k)$ is assumed to be Gaussian.

Assumption 2 The lower bound of the original MSSM-based cost function is maximized.

Assumption 3 The lower bound of the posterior covariance matrix $\boldsymbol{\Sigma}_k$ is assumed as the estimation error covariance matrix of the modified Kalman filter with modified PECM $\tilde{\boldsymbol{P}}^*_{k|k-1}$ and modified MNCM $\tilde{\boldsymbol{R}}^*_k$.

As for Assumption 1, in Bayesian filtering, it is always difficult to analytically formulate the non-Gaussian posterior PDF that is caused by the state and measurement outliers (Roth et al., 2017). In this study, to overcome this difficulty, the widely accepted Gaussian assumption is employed to look for an analytical presentation of the posterior PDF, thereby reaching a compromise between filtering accuracy and computational burden. Although such Gaussian assumption may introduce approximation errors into the posterior PDF to some extent, it exhibits satisfactory filtering accuracy with tolerable computational burden in engineering practice, as shown in later simulation study. Therefore, the Gaussian assumption for the non-Gaussian posterior PDF is reasonable.

Next, we analyze the influence of Assumption 2. For the sake of descriptions, a cluster of variables are defined as

$$\begin{cases}L^i_{1k}=\left[\boldsymbol{T}_{ki}(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k-1})\right]^2, & L^{i*}_{1k}=\boldsymbol{T}_{ki}\boldsymbol{A}^*_k\boldsymbol{T}^{\mathrm{T}}_{ki},\\ L^j_{2k}=\left[\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{x}_k)\right]^2, & L^{j*}_{2k}=\boldsymbol{U}_{kj}\boldsymbol{B}^*_k\boldsymbol{U}^{\mathrm{T}}_{kj},\end{cases}\tag{32}$$

where $i=1,2,\ldots,n$, $j=1,2,\ldots,m$, and L^{i*}_{1k} and L^{j*}_{2k} denote the expectations of L^i_{1k} and L^j_{2k} with respect to the optimal posterior PDF $q^*(\boldsymbol{x}_k)$. The first-order Taylor series expansions are performed on $f_x(l)$ and $f_z(l)$ at $l=L^{i*}_{1k}$ and $l=L^{i*}_{2k}$, respectively, which yields

$$\begin{cases}f_x(l)=f_x(L^{i*}_{1k})+\dot{f}_x(L^{i*}_{1k})(l-L^{i*}_{1k})+o(l-L^{i*}_{1k}),\\ f_z(l)=f_z(L^{j*}_{2k})+\dot{f}_z(L^{j*}_{2k})(l-L^{j*}_{2k})+o(l-L^{j*}_{2k}),\end{cases}\tag{33}$$

where $o(\cdot)$ denotes the high-order terms of the Taylor series expansions. Dropping the high-order terms yields the following first-order linearization approximations:

$$\begin{cases}f_x(L^i_{1k})\approx f_x(L^{i*}_{1k})+\dot{f}_x(L^{i*}_{1k})(L^i_{1k}-L^{i*}_{1k}),\\ f_z(L^j_{2k})\approx f_z(L^{j*}_{2k})+\dot{f}_z(L^{j*}_{2k})(L^j_{2k}-L^{j*}_{2k}).\end{cases}\tag{34}$$

By employing methods similar to that in Huang et al. (2020), we obtain the following propositions:

Proposition 5 Using the first-order linearization approximations (Eq. (33)) and the Gaussian assumption to the posterior PDF in Eq. (7), the original maximization problem (7) becomes the approximate maximization problem (12).

The proof of Proposition 5 is given in Appendix F.

Proposition 6 The variances of auxiliary variables L_{1k}^i and L_{2k}^j are upper-bounded, and can be formulated as follows:

$$\begin{cases} \mathrm{Var}\left[\sum\limits_{i=1}^{n} L_{1k}^i\right] \le n\left(n^2+2n\right)\sum\limits_{i=1}^{n}\left(\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k\boldsymbol{T}_{ki}^{\mathrm{T}}\right)^2 \\ \quad +n\sum\limits_{i=1}^{n}\left[\boldsymbol{T}_{ki}(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})\right]^2\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k\boldsymbol{T}_{ki}^{\mathrm{T}}, \\ \mathrm{Var}\left[\sum\limits_{j=1}^{m} L_{2k}^j\right] \\ \quad \le m\left(n^2+2n-1\right)\sum\limits_{j=1}^{m}\left(\boldsymbol{U}_{kj}\boldsymbol{H}_k\boldsymbol{\Sigma}_k\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{U}_{kj}^{\mathrm{T}}\right)^2 \\ \quad +4m\sum\limits_{j=1}^{m}\left[\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k)\right]^2\boldsymbol{U}_{kj}\boldsymbol{H}_k\boldsymbol{\Sigma}_k\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{U}_{kj}^{\mathrm{T}}. \end{cases} \tag{35}$$

The proof of Proposition 6 is given in Appendix G.

Proposition 5 means that the approximation errors of Assumption 2 are dominated mainly by the second-order moments of L_{1k}^i and L_{2k}^j. Meanwhile, the boundedness of the second-order moments of L_{1k}^i and L_{2k}^j, which is given in Proposition 6, guarantees that the approximation errors of Assumption 2 are bounded. As shown in inequality (35), these two upper bounds depend critically on the posterior covariance matrix $\boldsymbol{\Sigma}_k$ and the state dimension. The larger the posterior covariance matrix $\boldsymbol{\Sigma}_k$ or the state dimension n, the larger approximation errors will be induced. Because the posterior covariance matrix will decrease with the convergence of the filter, the approximation errors induced by Assumption 2 can be limited. Moreover, the exemplary similarity functions provided in this study possess much smaller high-order derivatives than the first-order derivatives, which helps Assumption 2 be more reasonable.

As for Assumption 3, the outlier-robust Kalman filters tend to generate a lager posterior covariance matrix $\boldsymbol{\Sigma}_k$ than the classical Kalman filter with the modified PECM $\tilde{\boldsymbol{P}}_{k|k-1}^*$ and the modified MNCM $\tilde{\boldsymbol{R}}_k^*$. Such a constraint is often beneficial in guaranteeing filtering consistency and stability.

3 Selection strategy for similarity functions and comparisons with existing RKFs

3.1 Selection strategy for similarity functions

The selection strategy for similarity functions $f_x(\cdot)$ and $f_z(\cdot)$ will be discussed in this subsection to facilitate the implementation of the proposed MORKF. First, the optimality of the proposed MORKF should be guaranteed when the state and measurement noises are Gaussian-distributed. According to Huang et al. (2020), the auxiliary matrices $\boldsymbol{A}_k^*$ and $\boldsymbol{B}_k^*$ can be approximated as the nominal one-step PECM $\boldsymbol{P}_{k|k-1}$ and MNCM $\boldsymbol{R}_k$ respectively when the state and measurement noises are Gaussian-distributed, i.e.,

$$\boldsymbol{A}_k^* \approx \boldsymbol{P}_{k|k-1},\ \boldsymbol{B}_k^* \approx \boldsymbol{R}_k. \tag{36}$$

According to Eqs. (8), (9), and (36), we have

$$\begin{aligned} &\begin{bmatrix} \boldsymbol{T}_{k1} \\ \boldsymbol{T}_{k2} \\ \vdots \\ \boldsymbol{T}_{kn} \end{bmatrix} \boldsymbol{A}_k^* \left[\boldsymbol{T}_{k1}^{\mathrm{T}}, \boldsymbol{T}_{k2}^{\mathrm{T}}, \ldots, \boldsymbol{T}_{kn}^{\mathrm{T}}\right] \\ \approx &\boldsymbol{S}_{k|k-1}^{-1}\boldsymbol{P}_{k|k-1}\boldsymbol{S}_{k|k-1}^{-\mathrm{T}} = \boldsymbol{I}_n, \end{aligned} \tag{37}$$

$$\begin{aligned} &\begin{bmatrix} \boldsymbol{U}_{k1} \\ \boldsymbol{U}_{k2} \\ \vdots \\ \boldsymbol{U}_{km} \end{bmatrix} \boldsymbol{B}_k^* \left[\boldsymbol{U}_{k1}^{\mathrm{T}}, \boldsymbol{U}_{k2}^{\mathrm{T}}, \ldots, \boldsymbol{U}_{km}^{\mathrm{T}}\right] \\ \approx &\boldsymbol{S}_{\boldsymbol{R}_k}^{-1}\boldsymbol{R}_k\boldsymbol{S}_{\boldsymbol{R}_k}^{-\mathrm{T}} = \boldsymbol{I}_m. \end{aligned} \tag{38}$$

Exploiting Eqs. (37) and (38) yields

$$\begin{cases} \boldsymbol{T}_{ki}\boldsymbol{A}_k^*\boldsymbol{T}_{ki}^{\mathrm{T}} = 1, & i = 1, 2, \ldots, n, \\ \boldsymbol{U}_{kj}\boldsymbol{B}_k^*\boldsymbol{U}_{kj}^{\mathrm{T}} = 1, & j = 1, 2, \ldots, m. \end{cases} \tag{39}$$

Substituting Eq. (39) into Eqs. (22) and (23), the diagonal weighted matrices can be rewritten as follows:

$$\boldsymbol{\Psi}_{\boldsymbol{x}_k}^* = \mathrm{diag}\left(-2\dot{f}_x(1), -2\dot{f}_x(1), \ldots, -2\dot{f}_x(1)\right), \tag{40}$$

$$\boldsymbol{\Psi}_{\boldsymbol{z}_k}^* = \mathrm{diag}\left(-2\dot{f}_z(1), -2\dot{f}_z(1), \ldots, -2\dot{f}_z(1)\right). \tag{41}$$

Algorithm 1 indicates that the proposed MORKF degrades into the classical KF when the weighted matrices satisfy $\boldsymbol{\Psi}_{\boldsymbol{x}_k}^{(i)}=\boldsymbol{I}_n$ and $\boldsymbol{\Psi}_{\boldsymbol{z}_k}^{(i)}=\boldsymbol{I}_m$. To meet such conditions, according to Eqs. (40) and (41), the similarity functions $f_x(\cdot)$ and $f_z(\cdot)$ should satisfy

$$\dot{f}_x(1)=-0.5,\ \dot{f}_z(1)=-0.5. \tag{42}$$

Next, we need to guarantee the robustness of the proposed MORKF. If the state and measurement noises are contaminated by outliers, the auxiliary matrices $\boldsymbol{A}_k^*$ and $\boldsymbol{B}_k^*$ satisfy the following inequalities (Huang et al., 2020):

$$\boldsymbol{A}_k^*\geq\boldsymbol{P}_{k|k-1},\ \boldsymbol{B}_k^*\geq\boldsymbol{R}_k. \tag{43}$$

Using the above inequalities, we can obtain the following theorem:

Theorem 2 For a linear system, the proposed MORKF exhibits robustness if the similarity functions $f_x(\cdot)$ and $f_z(\cdot)$ are chosen such that

$$\begin{cases}\dot{f}_x(l)<0,\ \ddot{f}_x(l)\geq 0,\ \dot{f}_x(1)=-0.5,\ l\in[0,+\infty),\\ \dot{f}_z(l)<0,\ \ddot{f}_z(l)\geq 0,\ \dot{f}_z(1)=-0.5,\ l\in[0,+\infty),\end{cases} \tag{44}$$

and then the diagonal weighted matrices satisfy

$$\begin{cases}\boldsymbol{0}<\boldsymbol{\Psi}_{\boldsymbol{x}_k}^*\leq\boldsymbol{I}_n,\\ \boldsymbol{0}<\boldsymbol{\Psi}_{\boldsymbol{z}_k}^*\leq\boldsymbol{I}_m.\end{cases} \tag{45}$$

The proof of Theorem 2 is given in Appendix H.

Employing inequality (45) in Eqs. (20) and (21) yields

$$\tilde{\boldsymbol{P}}_{k|k-1}^*-\boldsymbol{P}_{k|k-1}\geq\boldsymbol{0},\ \tilde{\boldsymbol{R}}_k^*-\boldsymbol{R}_k\geq\boldsymbol{0}. \tag{46}$$

Inequality (46) indicates that the modified PECM and the modified MNCM are not less than the nominal PECM and the nominal MNCM, respectively. Furthermore, according to the second and third conditions of similarity function $f(\cdot)$ and Eqs. (20) and (23), violent outliers may result in small diagonal elements of $\boldsymbol{\Psi}_{\boldsymbol{x}_k}^*$ and $\boldsymbol{\Psi}_{\boldsymbol{z}_k}^*$, and then a significantly modified PECM and MNCM will be obtained. Consequently, the modified PECM and MNCM can be adaptively adjusted along with the intensity and occurrence probability of outliers.

Several similarity functions are recommended and listed in Table 4, from which several exemplary MORKFs can be obtained, where σ denotes the kernel size, and ν and ω denote the degree-of-freedom (DOF) parameters. It is easy to demonstrate that all the similarity functions listed in Table 4 satisfy the conditions of Proposition 2. To meet the constraints of Propositions 2 and 4, the recommended similarity functions need to satisfy the following corollaries:

Table 4 Recommended similarity functions $f(\cdot)$ and their first- and second-order derivatives

$f(l)$	$\dot{f}(l)$	$\ddot{f}(l)$
$-0.5l$	-0.5	0
$\sigma^2\exp\left(\frac{1-l}{2\sigma^2}\right)$	$-0.5\exp\left(\frac{1-l}{2\sigma^2}\right)$	$\frac{1}{4\sigma^2}\exp\left(\frac{1-l}{2\sigma^2}\right)$
$-0.5(\nu+1)\log(1+\frac{l}{\nu})$	$-0.5\frac{\nu+1}{\nu+l}$	$0.5\frac{\nu+1}{(\nu+l)^2}$
$-\sqrt{(\omega+1)(\omega+l)}$	$-0.5\sqrt{\frac{\omega+1}{\omega+l}}$	$0.25\frac{\sqrt{\omega+1}}{\sqrt[3]{\omega+l}}$

σ: kernel size; ν and ω: degree-of-freedom (DOF) parameters

Corollary 1 For the recommended similarity functions listed in Table 4, inequality (26) in Proposition 2 holds only when the following conditions are satisfied:

$$\begin{cases}\sigma^2\exp\left(\dfrac{1-L_{1k}^{i*}}{2\sigma^2}\right)>1,\\ \sigma^2\exp\left(\dfrac{1-L_{2k}^{j*}}{2\sigma^2}\right)>1,\\ \nu>1,\end{cases} \tag{47}$$

where L_{1k}^{i*} and L_{2k}^{j*} are as given in Eq. (32).

The proof of Corollary 1 is given in Appendix I.

Corollary 2 The positive finite real numbers θ_1 and θ_2 in Proposition 4 exist only when $\sigma\nrightarrow 0$, $\nu\nrightarrow 0$, and $\omega\nrightarrow 0$.

The proof of Corollary 2 is given in Appendix J.

Remark 2 It is observed that all the first-order derivatives of the recommended similarity functions in Table 4 are -0.5 when σ, ν, and ω tend to infinity, i.e., $\{\sigma,\nu,\omega\}\to+\infty$. Therefore, the resultant exemplary MORKFs will degrade into the classical KF when $\{\sigma,\nu,\omega\}\to+\infty$.

3.2 Comparisons with existing RKFs

The M-estimator is a generalized maximum likelihood estimator that provides a robust state estimate by solving the following minimization problem (Huber, 2011):

$$\hat{\boldsymbol{x}}_{k|k}=\arg\min_{\boldsymbol{x}_k}\rho(\boldsymbol{x}_k), \tag{48}$$

where the cost function $\rho(\boldsymbol{x}_k)$ can be formulated as

$$\begin{aligned}\rho(\boldsymbol{x}_k)=&\sum_{i=1}^{n}\rho_x(\boldsymbol{T}_{ki}(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k-1}))\\&+\sum_{j=1}^{m}\rho_z(\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{x}_k)),\end{aligned} \tag{49}$$

where $\rho_x(\cdot)$ and $\rho_z(\cdot)$ are the robust cost functions that are applied to the predictive errors and residual errors, respectively.

It is observed from Eqs. (12), (13), (48), and (49) that the cost function of the proposed MORKF has a form similar to that of the M-estimator. The M-estimator optimizes only the state vector, whereas the proposed MORKF optimizes the state vector and the covariance matrix simultaneously. The M-estimator takes the stochastic state as a deterministic quantity and updates the covariance matrix independent of the optimization of the cost function $\rho(\boldsymbol{x}_k)$. The proposed MORKF exploits the randomness of the stochastic state by updating the state and the covariance matrix alternately, which benefits the performance of the proposed MORKF.

Remark 3 The existing SSMKF (Huang et al., 2020) was derived based on an assumption that the outliers occurring in different state and measurement dimensions have the same statistical properties, which may not be suitable in the scenarios with multiple outliers. By constructing a new MSSM-based cost function which imposes separate constraints on each state and measurement dimension, the proposed MORKF produces a couple of diagonal weighted matrices $\boldsymbol{\Psi}^*_{\boldsymbol{x}_k}$ and $\boldsymbol{\Psi}^*_{\boldsymbol{z}_k}$ to adaptively adjust the nominal PECM and MNCM, respectively, rather than two scalar-scale factors in the SSMKF.

4 Simulation study

The performance of the proposed MORKF is validated in a two-dimensional (2D) maneuvering target tracking example, where the positions are observed in a noisy scenario, and the positions and velocities are estimated simultaneously. The state transition matrix and measurement matrix are $\boldsymbol{F}_k = \begin{bmatrix} \boldsymbol{I}_2 & T\boldsymbol{I}_2 \\ \boldsymbol{0} & \boldsymbol{I}_2 \end{bmatrix}$ and $\boldsymbol{H}_k = \begin{bmatrix} \boldsymbol{I}_2 & \boldsymbol{0} \end{bmatrix}$ respectively, where $T = 1$ s. The state and measurement noises can be formulated as $\boldsymbol{w}_k = [w_{1,k},\ w_{2,k},\ w_{3,k},\ w_{4,k}]^{\mathrm{T}}$ and $\boldsymbol{v}_k = [v_{1,k},\ v_{2,k}]^{\mathrm{T}}$ respectively, whose nominal noise covariance matrices are given by $\boldsymbol{Q} = \begin{bmatrix} \frac{T^3}{3}\boldsymbol{I}_2 & \frac{T^2}{2}\boldsymbol{I}_2 \\ \frac{T^2}{2}\boldsymbol{I}_2 & T\boldsymbol{I}_2 \end{bmatrix}$ and $\boldsymbol{R} = r\boldsymbol{I}_2$ (r denotes a scale factor), respectively.

The superiority of the proposed MORKF is evaluated through comparisons with the classical KF with nominal noise covariance matrices (KFNNCM), the existing HKF (Huber, 2011), MCKF (Chen et al., 2017), and SSMKF (Huang et al., 2020). As discussed in our previous work (Huang et al., 2020), SSMKF can achieve the best estimation accuracy when the similarity function is selected as the square-root function, and the resultant SSMKF has better estimation performance than the existing RSTKF (Huang et al., 2017). To better show the advantages of the proposed method, the square-root similarity functions are used to implement the existing SSMKF and the proposed MORKF; then two algorithms that are abbreviated as SSMKF-sqrt and MORKF-sqrt respectively can be obtained. The parameter settings for all algorithms are listed in Table 5. The iteration threshold is set as $\epsilon = 10^{-16}$ and the maximum number of iterations is set as $N_{\mathrm{m}} = 50$. The simulation time is 1000 s, and 1000 Monte-Carlo runs are executed. All the algorithms are coded with MATLAB and executed on a computer with Intel Core i3-3110M CPU @2.40 GHz.

The initial state estimate $\hat{\boldsymbol{x}}_{0|0}$ is randomly extracted from a Gaussian PDF $N(\hat{\boldsymbol{x}}_{0|0}; \boldsymbol{x}_0, \boldsymbol{P}_0)$, where the true initial state vector is set as $\boldsymbol{x}_0 = [0,\ 0,\ 10,\ 10]^{\mathrm{T}}$, and the initial estimation error variance is set as $\boldsymbol{P}_0 = \mathrm{diag}(10\,000,\ 10\,000,\ 100,\ 100)$. The whole simulation is divided into two stages: a scenario with outliers having identical properties is simulated in the first stage (1–500 s), and a scenario with multiple outliers with different properties is simulated in the later stage (501–1000 s). The root mean square errors (RMSEs) and averaged RMSEs (ARMSEs) of position and velocity are taken as the performance metrics, both of which were defined in Huang et al. (2017). For a better presentation, the RMSEs are all smoothed by a moving window with a span of 50 s.

Case 1: We consider the case where both state and measurement noises are Gaussian-mixture. Particularly, in the first stage, the identical outlier-contaminated state and measurement noises are produced as follows:

$$\begin{cases} \boldsymbol{w}_k \sim \begin{cases} N(\boldsymbol{0}, \boldsymbol{Q}), & \text{w.p. } 0.97, \\ N(\boldsymbol{0}, U_1\boldsymbol{Q}), & \text{w.p. } 0.03, \end{cases} \\ \boldsymbol{v}_k \sim \begin{cases} N(0, \boldsymbol{R}), & \text{w.p. } 0.97, \\ N(0, U_3\boldsymbol{R}), & \text{w.p. } 0.03, \end{cases} \end{cases} \tag{50}$$

where "w.p." is short for "with probability."

In the second stage, the noise covariance matrices for $[w_{1,k},\ w_{3,k}]^{\mathrm{T}}$ and $[w_{2,k},\ w_{4,k}]^{\mathrm{T}}$ can be

Table 5 Parameter settings for compared algorithms

Filter	Parameter setting
HKF	Tuning parameter $\gamma = 1.345$
MCKF	Kernel size $\sigma = 15$
SSMKF-sqrt	DOF parameter $\nu = 5$
MORKF-sqrt	DOF parameter $\nu = 5$

DOF: degree-of-freedom

defined as $\boldsymbol{Q}_1 = \boldsymbol{Q}_2 = \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix}$, respectively. The multiple outlier-contaminated state and measurement noises are produced as follows:

$$\begin{cases} [w_{1,k},\ w_{3,k}]^{\mathrm{T}} \sim \begin{cases} N(\boldsymbol{0}, \boldsymbol{Q}_1), & \text{w.p. } 0.97, \\ N(\boldsymbol{0}, U_1\boldsymbol{Q}_1), & \text{w.p. } 0.03, \end{cases} \\ [w_{2,k},\ w_{4,k}]^{\mathrm{T}} \sim \begin{cases} N(\boldsymbol{0}, \boldsymbol{Q}_2), & \text{w.p. } 0.90, \\ N(\boldsymbol{0}, U_2\boldsymbol{Q}_2), & \text{w.p. } 0.10, \end{cases} \\ v_{1,k} \sim \begin{cases} N(0, r), & \text{w.p. } 0.97, \\ N(0, U_3 r), & \text{w.p. } 0.03, \end{cases} \\ v_{2,k} \sim \begin{cases} N(0, r), & \text{w.p. } 0.90, \\ N(0, U_4 r), & \text{w.p. } 0.10, \end{cases} \end{cases} \tag{51}$$

where the coefficients are $U_1 = U_4 = 300$, $U_2 = 400$, $U_3 = 500$, and $r = 100$.

Figs. 1 and 2 illustrate the RMSEs of position and velocity from all compared filters. We can observe that the proposed MORKF-sqrt has similar performance to the existing SSMKF-sqrt in the first stage but exhibits the smallest RMSEs in the second stage. In the first stage, both the proposed MORKF and the existing SSMKF outperform the HKF and MCKF because the randomness inherent in the stochastic state vector has been extensively exploited by using the posterior covariance matrix during the fixed-point iteration. However, in the second stage, SSMKF is inferior to the proposed MORKF because SSMKF is constructed based on an assumption that the outliers occurring in different dimensions possess the same statistical properties. The steady-state ARMSEs during the second stage (600–1000 s) and the runtime in a single step are summarized in Table 6. It can be seen from Table 6 that in the scenario with multiple outlier corrupted state and measurement noises, the proposed MORKF-sqrt has smaller steady-state ARMSEs than existing filters in position and velocity but greater runtime is required. As compared with the $\mathrm{ARMSE}_{\mathrm{pos}}$ and $\mathrm{ARMSE}_{\mathrm{vel}}$ from MCKF, the steady-state ARMSEs of the proposed MORKF-sqrt have been reduced by 25.55% and 4.60% in position and velocity, respectively.

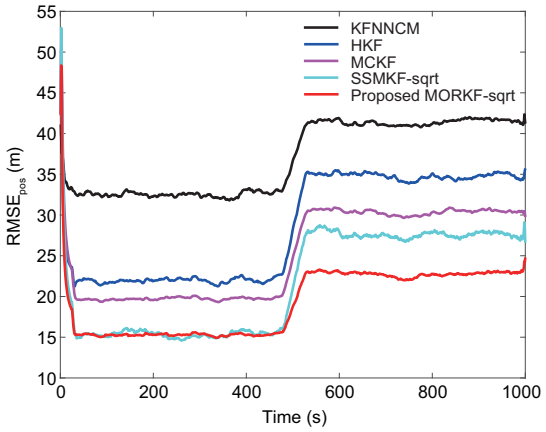


Fig. 1 Root mean square errors (RMSEs) of position from all filters in case 1 (References to color refer to the online version of this figure)

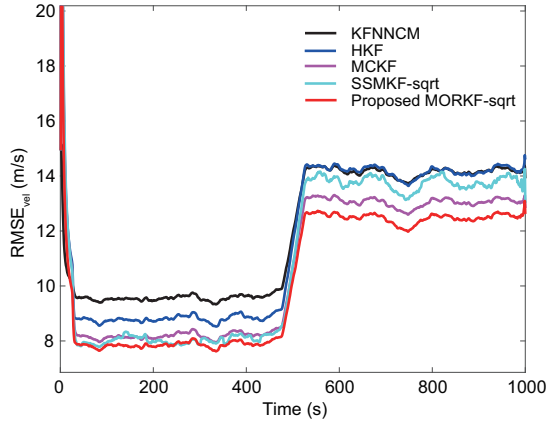


Fig. 2 Root mean square errors (RMSEs) of velocity from all filters in case 1 (References to color refer to the online version of this figure)

Table 6 Steady-state ARMSEs during 600–1000 s and runtime in a single step in case 1

Filter	$\mathrm{ARMSE}_{\mathrm{pos}}$ (m)	$\mathrm{ARMSE}_{\mathrm{vel}}$ (m/s)	Time (ms)
KFNNCM	41.61	14.13	0.049
HKF	34.80	14.15	1.397
MCKF	30.52	13.03	1.185
SSMKF-sqrt	28.21	13.94	1.819
MORKF-sqrt	22.72	12.43	2.023

ARMSE: average root mean square error

Next, we describe why the proposed MORKF outperforms the existing SSMKF for multiple outliers. The diagonal elements of the weighted diagonal matrices $\boldsymbol{\Psi}_{\boldsymbol{x}_k}$ and $\boldsymbol{\Psi}_{\boldsymbol{z}_k}$ from MORKF as well

as the scalar-scale factors ξ_k and λ_k from SSMKF are collected and averaged during 1000 Monte-Carlo runs, and are depicted in Figs. 3 and 4, respectively. In the first stage, the diagonal elements of the weighted matrix $\boldsymbol{\Psi}_{\boldsymbol{x}_k}$ and the scalar-scale factor ξ_k are of similar magnitude, so are matrix $\boldsymbol{\Psi}_{\boldsymbol{z}_k}$ and scalar λ_k, which results in similar performance for MORKF and SSMKF in the scenario with the same form of outliers. However, in the second stage, the scalar factors ξ_k and λ_k and the diagonal elements $\Psi_{x_k}(2,2)$, $\Psi_{x_k}(4,4)$, and $\Psi_{z_k}(2,2)$ are reduced to accommodate the suddenly increased state outliers in the second and fourth dimensions and the suddenly increased measurement outliers in the second dimension as described in Eq. (51), which results in the various enlargements of PECM and MNCM from dimension to dimension in the proposed MORKF, rather than the scalar enlargements of PECM and MNCM in SSMKF. Therefore, the proposed MORKF produces a couple of diagonal weighted matrices $\boldsymbol{\Psi}_{\boldsymbol{x}_k}$ and $\boldsymbol{\Psi}_{\boldsymbol{z}_k}$ to adaptively adjust the nominal PECM and MNCM, respectively, rather than two scalar-scale factors ξ_k and λ_k as in the SSMKF, which leads to better estimation accuracy compared with the existing cutting-edge SSMKF for addressing multiple outliers.

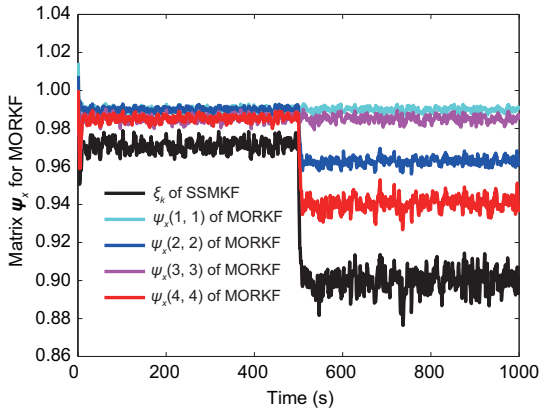


Fig. 3 Comparisons of scalar-scale factor ξ_k from SSMKF and diagonal elements of $\Psi_{\boldsymbol{x}_k}$ from MORKF in case 1 (References to color refer to the online version of this figure)

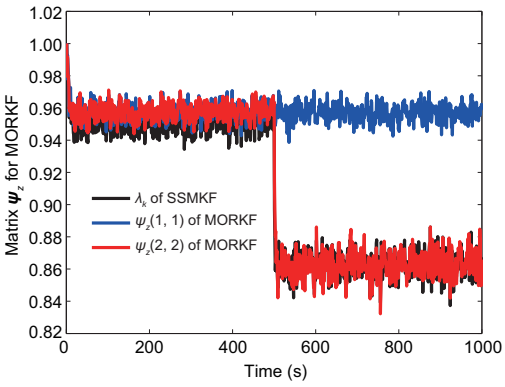


Fig. 4 Comparisons of scalar-scale factor λ_k from SSMKF and diagonal elements of $\Psi_{\boldsymbol{z}_k}$ from MORKF in case 1 (References to color refer to the online version of this figure)

Case 2: In this case, the identical outlier-contaminated state and measurement noises in the first stage are generated in the same way as in case 1. However, in the second stage, the state and measurement noises are produced by mixing the Gaussian distribution and uniform distribution in some dimensions. The specific formulations of noise generation are given as follows:

$$\begin{cases} [w_{1,k},\ w_{3,k}]^{\mathrm{T}} \sim \begin{cases} N(\boldsymbol{0},\boldsymbol{Q}_1), & \text{w.p. } 0.97, \\ N(\boldsymbol{0},U_1\boldsymbol{Q}_1), & \text{w.p. } 0.03, \end{cases} \\ [w_{2,k},\ w_{4,k}]^{\mathrm{T}} \sim \begin{cases} N(\boldsymbol{0},\boldsymbol{Q}_2), & \text{w.p. } 0.90, \\ U(w_{2,k};-100,100), & \text{w.p. } 0.10, \\ U(w_{4,k};-200,200), & \text{w.p. } 0.10, \end{cases} \\ v_{1,k} \sim \begin{cases} N(0,r), & \text{w.p. } 0.97, \\ N(0,U_3r), & \text{w.p. } 0.03, \end{cases} \\ v_{2,k} \sim \begin{cases} N(0,r), & \text{w.p. } 0.90, \\ U(v_{2,k};-600,600), & \text{w.p. } 0.10, \end{cases} \end{cases} \tag{52}$$

where $U(\sigma;a,b)$ denotes that variable σ is randomly extracted from a uniform distribution upon $[a,b]$.

The simulation results are given in Figs. 5 and

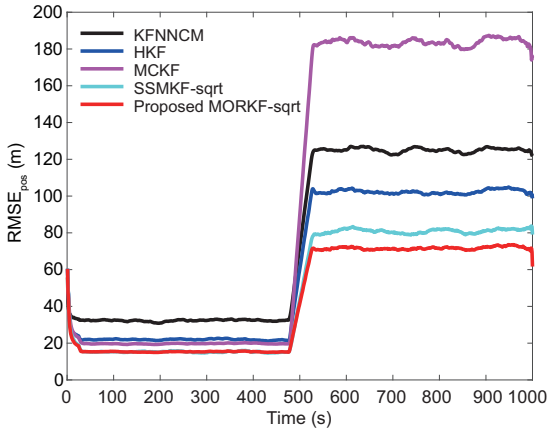


Fig. 5 Root mean square errors (RMSEs) of position from all filters in case 2 (References to color refer to the online version of this figure)

6, and the ARMSEs are summarized in Table 7. Similar to case 1, the proposed MORKF-sqrt has similar performance to the existing SSMKF-sqrt in the first stage. However, the proposed MORKF-sqrt presents the best performance in the second stage, because the multiple outliers can be addressed from dimension to dimension by optimizing the proposed MSSM-based cost function.

5 Conclusions

In this paper, we have presented a novel MORKF for linear stochastic discrete-time systems. To evaluate the similarity between two random vectors from dimension to dimension, a new MSSM was first introduced. The MORKF was derived by maximizing an MSSM-based cost function. To illustrate the effectiveness and superiority of the proposed MORKF, theoretical analysis and discussion have been provided, and the similarity function selections and comparisons with existing robust methods have also been presented. Simulation results demonstrated that the developed MORKF outperforms the existing cutting-edge robust KFs in terms of estimation accuracy for linear systems when the state and measurement noises are corrupted by multiple outliers.

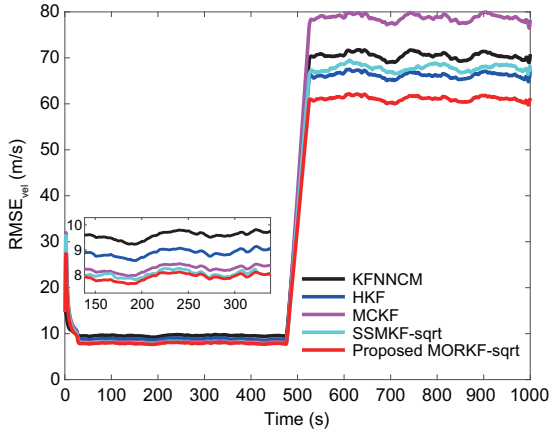


Fig. 6 Root mean square errors (RMSEs) of velocity from all filters in case 2 (References to color refer to the online version of this figure)

Table 7 Steady-state ARMSEs during 600–1000 s in case 2

Filter	ARMSE$_{pos}$ (m)	ARMSE$_{vel}$ (m/s)
KFNNCM	125.12	70.50
HKF	102.36	66.27
MCKF	183.59	78.74
SSMKF-sqrt	81.11	67.90
MORKF-sqrt	71.73	61.14

ARMSE: average root mean square error

Contributors

Yulong HUANG designed the algorithm. Yulong HUANG and Mingming BAI coded the simulation. Mingming BAI drafted the paper. Yonggang ZHANG helped organize the paper. Mingming BAI and Yonggang ZHANG revised and finalized the paper.

Compliance with ethics guidelines

Yulong HUANG, Mingming BAI, and Yonggang ZHANG declare that they have no conflict of interest.

References

Chen BD, Liu X, Zhao HQ, et al., 2017. Maximum correntropy Kalman filter. *Automatica*, 76:70-77. https://doi.org/10.1016/j.automatica.2016.10.004

Huang YL, Zhang YG, Li N, et al., 2016. Robust Student's t based nonlinear filter and smoother. *IEEE Trans Aerosp Electron Syst*, 52(5):2586-2596. https://doi.org/10.1109/TAES.2016.150722

Huang YL, Zhang YG, Li N, et al., 2017. A novel robust Student's t-based Kalman filter. *IEEE Trans Aerosp Electron Syst*, 53(3):1545-1554. https://doi.org/10.1109/TAES.2017.2651684

Huang YL, Zhang YG, Zhao YX, et al., 2019a. A novel robust Gaussian-Student's t mixture distribution based Kalman filter. *IEEE Trans Signal Process*, 67(13):3606-3620. https://doi.org/10.1109/TSP.2019.2916755

Huang YL, Zhang YG, Chambers JA, 2019b. A novel Kullback–Leibler divergence minimization-based adaptive Student's t-filter. *IEEE Trans Signal Process*, 67(20):5417-5432. https://doi.org/10.1109/TSP.2019.2939079

Huang YL, Zhang YG, Shi P, et al., 2019c. Robust Kalman filters based on Gaussian scale mixture distributions with application to target tracking. *IEEE Trans Syst Man Cybern Syst*, 49(10):2082-2096. https://doi.org/10.1109/TSMC.2017.2778269

Huang YL, Zhang YG, Zhao YX, et al., 2020. A novel outlier-robust Kalman filtering framework based on statistical similarity measure. *IEEE Trans Autom Contr*, 66(6):2677-2692. https://doi.org/10.1109/TAC.2020.3011443

Huber PJ, 2011. Robust statistics. In: Lovric M (Ed.), International Encyclopedia of Statistical Science. Springer, Berlin, Germany.

Piché R, Särkkä S, Hartikainen J, 2012. Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution. Proc IEEE Int Workshop on Machine Learning for Signal Processing, p.1-6. https://doi.org/10.1109/MLSP.2012.6349794

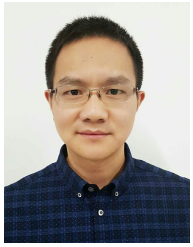
Roth M, Ardeshiri T, Özkan E, et al., 2017. Robust Bayesian filtering and smoothing using Student's t distribution. https://arxiv.org/abs/1703.02428

Simon D, 2006. Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches. John Wiley, Hoboken, USA.

Ting JA, Theodorou E, Schaal S, 2007. Learning an outlier-robust Kalman filter. Proc 18th European Conf on Machine Learning, p.748-756. https://doi.org/10.1007/978-3-540-74958-5_76

Yulong HUANG, first author of this invited paper, received his BS degree in automation and PhD degree in control science and engineering from the College of Automation, Harbin Engineering University (HEU), Harbin, China, in 2012 and 2018, respectively. From Nov. 2016 to Nov. 2017, he was a visiting researcher at the Electrical Engineering Department of Columbia University, New York, USA. Currently, he is an associate professor of navigation, guidance, and control in HEU. From Dec. 2019 to Dec. 2021, he was associated with the Department of Mechanical Engineering, City University of Hong Kong, as a Hong Kong Scholar. His current research interests include state estimation, intelligent information fusion, and their applications in navigation technology, such as inertial navigation, integrated navigation, intelligent navigation, and cooperative navigation.

Mingming BAI received the BS degree in measurement and control technology and instrumentation from the College of Automation, China University of Geosciences, Wuhan, China, in 2016. He is currently pursuing the PhD degree in control science and engineering with the HEU. His current research interests include target tracking, data fusion, multiagent systems, and state estimation theory.



Yonggang ZHANG, corresponding author of this invited paper, received his BS and MS degrees from the College of Automation, HEU, Harbin, China, in 2002 and 2004, respectively. He received his PhD degree in electronic engineering from Cardiff University, UK in 2007 and worked as a post-doctoral fellow at Loughborough University, UK from 2007 to 2008 in the area of adaptive signal processing. Currently, he is a professor of navigation, guidance, and control in HEU. His current research interests include signal processing, information fusion, and their applications in navigation technology, such as fiber optical gyroscope, inertial navigation, and integrated navigation. Prof. ZHANG currently serves as a corresponding expert for *Front Inform Technol Electron Eng.*

Appendix A: Proof of Proposition 1

Since $\dot{f}(l)<0$ and $(\alpha_i-\beta_i)^2\geq 0$, we have

$$f((\alpha_i-\beta_i)^2)\leq f(0). \tag{A1}$$

Using inequality (A1) in Eq. (2), we obtain

$$\begin{aligned} s(\boldsymbol{\alpha},\boldsymbol{\beta}) &= \sum_{i=1}^{p}\int\int f((\alpha_i-\beta_i)^2)p(\boldsymbol{\alpha},\boldsymbol{\beta})\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{\beta} \\ &\leq \sum_{i=1}^{p}\int\int f(0)p(\boldsymbol{\alpha},\boldsymbol{\beta})\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{\beta} \\ &= pf(0). \end{aligned} \tag{A2}$$

Hence, the maximum point $pf(0)$ is reached only when $\boldsymbol{\alpha}=\boldsymbol{\beta}$.

Appendix B: Proof of Theorem 1

The Jacobian matrix of $J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ with respect to $\boldsymbol{\mu}_k$ can be formulated as follows:

$$\begin{aligned} &\boldsymbol{\Lambda}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k) \\ =&2\sum_{i=1}^{n}\dot{f}_x(\boldsymbol{T}_{ki}\boldsymbol{A}_k\boldsymbol{T}_{ki}^{\mathrm{T}})\boldsymbol{T}_{ki}^{\mathrm{T}}\boldsymbol{T}_{ki}(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1}) \\ &-2\sum_{j=1}^{m}\dot{f}_z(\boldsymbol{U}_{kj}\boldsymbol{B}_k\boldsymbol{U}_{kj}^{\mathrm{T}})\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{U}_{kj}^{\mathrm{T}}\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k). \end{aligned} \tag{B1}$$

Define a couple of diagonal weighted matrices:

$$\begin{cases} \boldsymbol{\Psi}_{\boldsymbol{x}_k}=\mathrm{diag}\left(-2\dot{f}_x(\boldsymbol{T}_{k1}\boldsymbol{A}_k\boldsymbol{T}_{k1}^{\mathrm{T}}),\right. \\ \quad \left.-2\dot{f}_x(\boldsymbol{T}_{k2}\boldsymbol{A}_k\boldsymbol{T}_{k2}^{\mathrm{T}}),\ldots,-2\dot{f}_x(\boldsymbol{T}_{kn}\boldsymbol{A}_k\boldsymbol{T}_{kn}^{\mathrm{T}})\right), \\ \boldsymbol{\Psi}_{\boldsymbol{z}_k}=\mathrm{diag}\left(-2\dot{f}_z(\boldsymbol{U}_{k1}\boldsymbol{B}_k\boldsymbol{U}_{k1}^{\mathrm{T}}),\right. \\ \quad \left.-2\dot{f}_z(\boldsymbol{U}_{k2}\boldsymbol{B}_k\boldsymbol{U}_{k2}^{\mathrm{T}}),\ldots,-2\dot{f}_z(\boldsymbol{U}_{km}\boldsymbol{B}_k\boldsymbol{U}_{km}^{\mathrm{T}})\right), \end{cases} \tag{B2}$$

and then the Jacobian matrix in Eq. (B1) can be rewritten as follows:

$$\begin{aligned} \boldsymbol{\Lambda}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k) =& -\boldsymbol{S}_{k|k-1}^{-\mathrm{T}}\boldsymbol{\Psi}_{\boldsymbol{x}_k}\boldsymbol{S}_{k|k-1}^{-1}(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1}) \\ &+\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{S}_{\boldsymbol{R}_k}^{-\mathrm{T}}\boldsymbol{\Psi}_{\boldsymbol{z}_k}\boldsymbol{S}_{\boldsymbol{R}_k}^{-1}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k). \end{aligned} \tag{B3}$$

According to the maximum point criterion, exploiting $\boldsymbol{\Lambda}_{\boldsymbol{\mu}_k}^{*}(\boldsymbol{\mu}_k^{*},\boldsymbol{\Sigma}_k^{*})=\boldsymbol{0}$ yields

$$\begin{aligned} &-\boldsymbol{S}_{k|k-1}^{-\mathrm{T}}\boldsymbol{\Psi}_{\boldsymbol{x}_k}^{*}\boldsymbol{S}_{k|k-1}^{-1}(\boldsymbol{\mu}_k^{*}-\hat{\boldsymbol{x}}_{k|k-1}) \\ &+\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{S}_{\boldsymbol{R}_k}^{-\mathrm{T}}\boldsymbol{\Psi}_{\boldsymbol{z}_k}^{*}\boldsymbol{S}_{\boldsymbol{R}_k}^{-1}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k^{*})=\boldsymbol{0}. \end{aligned} \tag{B4}$$

Substituting Eqs. (20) and (21) into Eq. (B3) and using the matrix inversion lemma in Simon (2006) yield the results in Theorem 1.

Appendix C: Proof of Proposition 2

Using Eqs. (13) and (16), the Hessian matrix of $J(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ with respect to $\boldsymbol{\mu}_k$ can be given by

$$\begin{aligned}&\boldsymbol{\Pi}^*_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}^*_k, \boldsymbol{\Sigma}^*_k)\\ =&-\sum_{i=1}^{n}\alpha^*_{ki}\boldsymbol{T}^{\mathrm{T}}_{ki}\boldsymbol{T}_{ki}-\sum_{j=1}^{m}\beta^*_{kj}\boldsymbol{H}^{\mathrm{T}}_k\boldsymbol{U}^{\mathrm{T}}_{kj}\boldsymbol{U}_{kj}\boldsymbol{H}_k+\boldsymbol{O}^*_{\boldsymbol{\mu}_k},\end{aligned} \tag{C1}$$

where the quadratic term $\boldsymbol{O}^*_{\boldsymbol{\mu}_k}$ is given by

$$\begin{aligned}\boldsymbol{O}^*_{\boldsymbol{\mu}_k}=&\sum_{j=1}^{m}2\tilde{\beta}^*_{kj}\boldsymbol{H}^{\mathrm{T}}_k\boldsymbol{U}^{\mathrm{T}}_{kj}\boldsymbol{F}_{kj}\boldsymbol{U}_{kj}\boldsymbol{H}_k\\ &+\sum_{i=1}^{n}2\tilde{\alpha}^*_{ki}\boldsymbol{T}^{\mathrm{T}}_{ki}\boldsymbol{D}_{ki}\boldsymbol{T}_{ki},\end{aligned} \tag{C2}$$

and the auxiliary matrices $\boldsymbol{D}_{ki}$ and $\boldsymbol{F}_{kj}$ satisfy (Huang et al., 2020)

$$\begin{cases}\boldsymbol{D}_{ki}=\boldsymbol{T}_{ki}(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\boldsymbol{T}^{\mathrm{T}}_{ki}\\ \quad<\boldsymbol{T}_{ki}\tilde{\boldsymbol{P}}^*_{k|k-1}\boldsymbol{T}^{\mathrm{T}}_{ki},\\ \boldsymbol{F}_{kj}=\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k)(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k)^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}_{kj}\\ \quad<\boldsymbol{U}_{kj}\tilde{\boldsymbol{R}}^*_k\boldsymbol{U}^{\mathrm{T}}_{kj}.\end{cases} \tag{C3}$$

Exploiting Eqs. (16) and (20)–(23) in inequality (C3) yields

$$\boldsymbol{D}_{ki}<1/\alpha^*_{ki},\ \boldsymbol{F}_{kj}<1/\beta^*_{kj}. \tag{C4}$$

Substituting inequality (C4) into Eq. (C2), the Hessian matrix satisfies

$$\begin{aligned}\boldsymbol{\Pi}^*_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}^*_k, \boldsymbol{\Sigma}^*_k)<&\sum_{i=1}^{n}(-\alpha^*_{ki}+2\tilde{\alpha}^*_{ki}/\alpha^*_{ki})\boldsymbol{T}^{\mathrm{T}}_{ki}\boldsymbol{T}_{ki}\\ &+\sum_{j=1}^{m}(-\beta^*_{kj}+2\tilde{\beta}^*_{kj}\beta^*_{kj})\boldsymbol{H}^{\mathrm{T}}_k\boldsymbol{U}^{\mathrm{T}}_{kj}\boldsymbol{U}_{kj}\boldsymbol{H}_k.\end{aligned} \tag{C5}$$

The Hessian matrix is negative definite if the conditions in Proposition 2 hold.

Appendix D: Proof of Proposition 3

Using Eqs. (14) and (15) in Eq. (13) and calculating the derivative of the resultant cost function $J(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ with respect to $\boldsymbol{\Sigma}_k$, the Jacobian matrix $\boldsymbol{\Lambda}_{\boldsymbol{\Sigma}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ can be formulated as follows:

$$\begin{aligned}\boldsymbol{\Lambda}_{\boldsymbol{\Sigma}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)=&-0.5\boldsymbol{S}^{-\mathrm{T}}_{k|k-1}\boldsymbol{\Psi}_{\boldsymbol{x}_k}\boldsymbol{S}^{-1}_{k|k-1}\\ &-0.5\boldsymbol{H}^{\mathrm{T}}_k\boldsymbol{S}^{-\mathrm{T}}_{\boldsymbol{R}_k}\boldsymbol{\Psi}_{\boldsymbol{z}_k}\boldsymbol{S}^{-1}_{\boldsymbol{R}_k}\boldsymbol{H}_k.\end{aligned} \tag{D1}$$

According to the second condition of $f(\cdot)$, we know that both $\boldsymbol{\Psi}_{\boldsymbol{x}_k}$ and $\boldsymbol{\Psi}_{\boldsymbol{z}_k}$ are positive definite. Therefore, the Jacobian matrix $\boldsymbol{\Lambda}_{\boldsymbol{\Sigma}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ in Eq. (D1) is negative definite.

Appendix E: Proof of Proposition 4

Dropping the quadratic term in Eq. (C1) and taking the Frobenius norm of the derivative of the modified Hessian matrix $\tilde{\boldsymbol{\Pi}}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ yield

$$\begin{aligned}\left\|\frac{\partial\tilde{\boldsymbol{\Pi}}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\partial\boldsymbol{\mu}_k}\right\|_{\mathcal{F}}=&\sum_{i=1}^{n}2\tilde{\alpha}_{ki}\eta_1\sqrt{\boldsymbol{T}_{ki}(\boldsymbol{\mu}-\hat{\boldsymbol{x}}_{k|k-1})}\\ &+\sum_{j=1}^{m}2\tilde{\beta}_{kj}\eta_2\sqrt{\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu})},\end{aligned} \tag{E1}$$

where

$$\begin{cases}\eta_1=\|\boldsymbol{T}_{ki}\|_{\mathcal{F}}\|\boldsymbol{T}^{\mathrm{T}}_{ki}\boldsymbol{T}_{ki}\|_{\mathcal{F}},\\ \eta_2=\|\boldsymbol{H}^{\mathrm{T}}_k\boldsymbol{U}^{\mathrm{T}}_{kj}\|_{\mathcal{F}}\|\boldsymbol{H}^{\mathrm{T}}_k\boldsymbol{U}^{\mathrm{T}}_{kj}\boldsymbol{H}_k\boldsymbol{U}_{kj}\|_{\mathcal{F}}.\end{cases} \tag{E2}$$

According to Eqs. (14) and (15), we obtain

$$\begin{cases}\sqrt{\boldsymbol{T}_{ki}(\boldsymbol{\mu}-\hat{\boldsymbol{x}}_{k|k-1})}\le\sqrt{\boldsymbol{T}_{ki}\boldsymbol{A}_k\boldsymbol{T}^{\mathrm{T}}_{ki}},\\ \sqrt{\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu})}\le\sqrt{\boldsymbol{U}_{kj}\boldsymbol{B}_k\boldsymbol{U}^{\mathrm{T}}_{kj}}.\end{cases} \tag{E3}$$

Using inequality (29) and Eq. (C2) in Eq. (B4) yields

$$\left\|\frac{\partial\tilde{\boldsymbol{\Pi}}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\partial\boldsymbol{\mu}_k}\right\|_{\mathcal{F}}\le 4n\theta_1\eta_1+4m\theta_2\eta_2. \tag{E4}$$

Similar to our previous work (Huang et al., 2020), using inequality (C3), we can prove that the modified Hessian matrix $\tilde{\boldsymbol{\Pi}}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ satisfies the Lipschitz condition, from which the results in Proposition 4 hold.

Appendix F: Proof of Proposition 5

Substituting Eq. (34) into Eq. (7) yields

$$\begin{aligned}\tilde{J}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\approx&\sum_{i=1}^{n}\int \dot{f}_x(L_{1k}^{i*})L_{1k}^{i}q(\boldsymbol{x}_k)\mathrm{d}\boldsymbol{x}_k\\&+\sum_{j=1}^{m}\int \dot{f}_z(L_{2k}^{j*})L_{2k}^{j}q(\boldsymbol{x}_k)\mathrm{d}\boldsymbol{x}_k+c_{\{\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k\}}\\=&\sum_{i=1}^{n}\dot{f}_x(L_{1k}^{i*})\boldsymbol{T}_{ki}\boldsymbol{A}_k\boldsymbol{T}_{ki}^{\mathrm{T}}\\&+\sum_{j=1}^{m}\dot{f}_z(L_{2k}^{j*})\boldsymbol{U}_{kj}\boldsymbol{B}_k\boldsymbol{T}_{kj}^{\mathrm{T}}+c_{\{\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k\}},\end{aligned}\tag{F1}$$

where $c_{\{\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k\}}$ denotes a constant with no respect to $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$.

Making derivative operations on $\tilde{J}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ yields

$$\frac{\partial\tilde{J}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\partial\boldsymbol{\mu}_k}=\boldsymbol{\Lambda}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k),\tag{F2}$$

$$\frac{\partial\tilde{J}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\partial\boldsymbol{\Sigma}_k}=\boldsymbol{\Lambda}_{\boldsymbol{\Sigma}_k}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k),\tag{F3}$$

$$\frac{\partial^2\tilde{J}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\partial\boldsymbol{\mu}_k\partial\boldsymbol{\mu}_k^{\mathrm{T}}}=\boldsymbol{\Pi}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)-\boldsymbol{O}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k),\tag{F4}$$

where $\boldsymbol{O}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ represents the quadratic term of the Hessian matrix of the original cost function $J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ with respect to $\boldsymbol{\mu}_k$, and is given by

$$\begin{aligned}\boldsymbol{O}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)=&\sum_{i=1}^{n}4\ddot{f}_x(\boldsymbol{T}_{ki}\boldsymbol{A}_k\boldsymbol{T}_{ki}^{\mathrm{T}})\boldsymbol{T}_{ki}^{\mathrm{T}}\boldsymbol{T}_{ki}(\boldsymbol{\mu}_k\\&-\hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\boldsymbol{T}_{ki}^{\mathrm{T}}\boldsymbol{T}_{ki}+\sum_{j=1}^{m}4\ddot{f}_z(\boldsymbol{U}_{kj}\boldsymbol{B}_k\boldsymbol{U}_{kj}^{\mathrm{T}})\\&\cdot\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{U}_{kj}^{\mathrm{T}}\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k)(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k)^{\mathrm{T}}\boldsymbol{U}_{kj}^{\mathrm{T}}\boldsymbol{U}_{kj}\boldsymbol{H}_k.\end{aligned}\tag{F5}$$

The cost functions $\tilde{J}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ and $J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ have the same Jacobin matrices, and the Hessian matrix $\frac{\partial^2\tilde{J}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\partial\boldsymbol{\mu}_k\partial\boldsymbol{\mu}_k^{\mathrm{T}}}$ is negative definite. Therefore, the cost functions $\tilde{J}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ and $J(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ have the same optimal solutions.

Appendix G: Proof of Proposition 6

The variable in Eq. (32) can be rewritten as follows:

$$\begin{aligned}L_{1k}^{i}=&\left[\boldsymbol{T}_{ki}(\boldsymbol{x}_k-\boldsymbol{\mu}_k)+\boldsymbol{T}_{ki}(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})\right]^2\\=&\left[\boldsymbol{T}_{ki}(\boldsymbol{x}_k-\boldsymbol{\mu}_k)\right]^2+2\boldsymbol{T}_{ki}(\boldsymbol{x}_k-\boldsymbol{\mu}_k)(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\\&\cdot\boldsymbol{T}_{ki}^{\mathrm{T}}+\left[\boldsymbol{T}_{ki}(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})\right]^2.\end{aligned}\tag{G1}$$

By introducing the posterior covariance matrix $\boldsymbol{\Sigma}_k$, the formulation in Eq. (G1) can be further rewritten as

$$\begin{aligned}&\left[\boldsymbol{T}_{ki}(\boldsymbol{x}_k-\boldsymbol{\mu}_k)\right]^2+2\boldsymbol{T}_{ki}(\boldsymbol{x}_k-\boldsymbol{\mu}_k)(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\\&\cdot\boldsymbol{T}_{ki}^{\mathrm{T}}+\left[\boldsymbol{T}_{ki}(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})\right]^2\\=&\left[\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{\Sigma}_k^{-\frac{1}{2}}(\boldsymbol{x}_k-\boldsymbol{\mu}_k)\right]^2+2\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{\Sigma}_k^{-\frac{1}{2}}\\&\cdot(\boldsymbol{x}_k-\boldsymbol{\mu}_k)(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\boldsymbol{T}_{ki}^{\mathrm{T}}+\left[\boldsymbol{T}_{ki}(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})\right]^2\\=&\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{e}_x\boldsymbol{e}_x^{\mathrm{T}}\boldsymbol{\Sigma}_k^{\frac{\mathrm{T}}{2}}\boldsymbol{T}_{ki}^{\mathrm{T}}+2\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{e}_x(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\\&\cdot\boldsymbol{T}_{ki}^{\mathrm{T}}+\left[\boldsymbol{T}_{ki}(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})\right]^2,\end{aligned}\tag{G2}$$

where $\boldsymbol{e}_x=\boldsymbol{\Sigma}_k^{-\frac{1}{2}}(\boldsymbol{x}_k-\boldsymbol{\mu}_k)$. The random vector $\boldsymbol{e}_x$ satisfies a standard normal distribution, namely, $\boldsymbol{e}_x\sim N(\boldsymbol{e}_x;\boldsymbol{0},\boldsymbol{I}_m)$.

The variance of L_{1k}^{i} can be formulated as

$$\begin{aligned}\mathrm{Var}\left[L_{1k}^{i}\right]=&\mathrm{Var}\left[\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{e}_x\boldsymbol{e}_x^{\mathrm{T}}\boldsymbol{\Sigma}_k^{\frac{\mathrm{T}}{2}}\boldsymbol{T}_{ki}^{\mathrm{T}}\right]\\&+4\mathrm{Var}\left[\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{e}_x(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\boldsymbol{T}_{ki}^{\mathrm{T}}\right],\end{aligned}\tag{G3}$$

where the terms in Eq. (G3) can be given by

$$\begin{cases}\mathrm{Var}\left[\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{e}_x\boldsymbol{e}_x^{\mathrm{T}}\boldsymbol{\Sigma}_k^{\frac{\mathrm{T}}{2}}\boldsymbol{T}_{ki}^{\mathrm{T}}\right]\\=E\left[\left(\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{e}_x\right)^4\right]-\left(\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k\boldsymbol{T}_{ki}^{\mathrm{T}}\right)^2,\\\mathrm{Var}\left[\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{e}_x(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\boldsymbol{T}_{ki}^{\mathrm{T}}\right]\\=\left[\boldsymbol{T}_{ki}(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})\right]^2\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k\boldsymbol{T}_{ki}^{\mathrm{T}}.\end{cases}\tag{G4}$$

Due to $\left(\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{e}_x\right)^2\le\left\|\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\right\|_{\mathcal{F}}^2\gamma_x$, we can easily obtain

$$\left(\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{e}_x\right)^4\le\left\|\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\right\|_{\mathcal{F}}^4\gamma_x^2,\tag{G5}$$

where $\gamma_x=\|\boldsymbol{e}_x\|^2$ satisfies a chi-square distribution with DOF parameter n, namely, $\gamma_x\sim\chi^2(n)$. According to the property of the chi-square distribution, the mean of the random variable γ_x is n and the variance is $2n$.

Making an expectation operation on both sides of inequality (G5), we obtain

$$E\left[\left(\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\boldsymbol{e}_x\right)^4\right]\leq\left(n^2+2n\right)\left(\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k\boldsymbol{T}_{ki}^{\mathrm{T}}\right)^2. \quad \text{(G6)}$$

Then we have

$$\begin{aligned}\mathrm{Var}\left[L_{1k}^i\right]\leq&\left(n^2+2n\right)\left(\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k\boldsymbol{T}_{ki}^{\mathrm{T}}\right)^2\\&+\left[\boldsymbol{T}_{ki}(\boldsymbol{\mu}_k-\hat{\boldsymbol{x}}_{k|k-1})\right]^2\boldsymbol{T}_{ki}\boldsymbol{\Sigma}_k\boldsymbol{T}_{ki}^{\mathrm{T}}.\end{aligned} \quad \text{(G7)}$$

Using similar means, we can obtain the following result:

$$\begin{aligned}&\mathrm{Var}\left[L_{2k}^j\right]\leq\left(n^2+2n-1\right)\left(\boldsymbol{U}_{kj}\boldsymbol{H}_k\boldsymbol{\Sigma}_k\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{U}_{kj}^{\mathrm{T}}\right)^2\\&+4\left[\boldsymbol{U}_{kj}(\boldsymbol{z}_k-\boldsymbol{H}_k\boldsymbol{\mu}_k)\right]^2\boldsymbol{U}_{kj}\boldsymbol{H}_k\boldsymbol{\Sigma}_k\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{U}_{kj}^{\mathrm{T}}.\end{aligned} \quad \text{(G8)}$$

Assuming that L_{1k}^i and L_{2k}^j are independent of each other, we obtain inequality (35).

Appendix H: Proof of Theorem 2

Using Eq. (4) and inequality (43) yields

$$\begin{cases}\boldsymbol{T}_{ki}\boldsymbol{A}_k^*\boldsymbol{T}_{ki}^{\mathrm{T}}\geq\boldsymbol{T}_{ki}\boldsymbol{P}_{k|k-1}\boldsymbol{T}_{ki}^{\mathrm{T}}=1,\\\boldsymbol{U}_{kj}\boldsymbol{B}_k^*\boldsymbol{U}_{kj}^{\mathrm{T}}\geq\boldsymbol{U}_{kj}\boldsymbol{R}_k\boldsymbol{U}_{kj}^{\mathrm{T}}=1.\end{cases} \quad \text{(H1)}$$

According to condition (44), we have

$$\begin{cases}-0.5\leq\dot{f}_x(\boldsymbol{T}_{ki}\boldsymbol{A}_k^*\boldsymbol{T}_{ki}^{\mathrm{T}})<0,\\-0.5\leq\dot{f}_z(\boldsymbol{U}_{kj}\boldsymbol{B}_k^*\boldsymbol{U}_{kj}^{\mathrm{T}})<0.\end{cases} \quad \text{(H2)}$$

Substituting inequality (C5) into Eqs. (22) and (23) yields the results given in Theorem 2.

Appendix I: Proof of Corollary 1

Substituting Eq. (32) into Eq. (16), inequality (26) can be reformulated as follows:

$$\begin{cases}\left[\dot{f}_x(L_{1k}^{i*})\right]^2>\ddot{f}_x(L_{1k}^{i*}),\\\left[\dot{f}_z(L_{2k}^{j*})\right]^2>\ddot{f}_z(L_{2k}^{j*}).\end{cases} \quad \text{(I1)}$$

Using the exemplary similarity functions given in Table 4 seriatim in inequality (I1), we obtain the results in inequality (47).

Appendix J: Proof of Corollary 2

For the case of exponential similarity function, constructing an auxiliary function $h(l)=\ddot{f}(l^2)l$ and making a derivative operation on it yield

$$\dot{h}(l)=\frac{1}{4\sigma^2}\exp\left(\frac{1-l^2}{2\sigma^2}\right)\left(1-\frac{l^2}{\sigma^2}\right). \quad \text{(J1)}$$

It is obvious that $h(l)$ has a unique maximum value $h(\sigma)=\dfrac{1}{4\sigma}\exp\left(\dfrac{1}{2\sigma^2}-\dfrac{1}{2}\right)$, which is positively bounded when $\sigma\nrightarrow 0$.

Using similar means, the logarithmic and square-root similarity functions can be verified to satisfy the inequalities given in inequality (29) when $\nu\nrightarrow 0$ and $\omega\nrightarrow 0$.