Map-Based Localization Under Adversarial Attacks

Yulin Yang and Guoquan Huang

1 Introduction and Related Work

It is conceivable that thousands of autonomous vehicles will be operated in a wide range of civilian and military application domains, such as self-driving cars, unmanned aerial vehicles (UAVs), and autonomous underwater vehicles (AUVs). However, current onboard navigation systems for these vehicles are often vulnerable to malicious attacks—that is, terrorists and criminals may easily hijack vehicles to attack the public. While the study of secure control has made important advances over the past few years, the vast majority of this literature focuses on cyber attacks. However, sensor attacks—manipulating physical fields such as electromagnetic and pressure which are measured by sensors and/or directly compromising measurements even if communication is secure (e.g. see [1, 2])—pose a more menacing threat to autonomous navigation systems.

In particular, secure state estimation and control in cyber-physical systems has gained significant attention (e.g., [3–8]), because it was realized that adversarial attacks on sensors truly occur in real life. For example, the first-time-ever attack (Stuxnet) on the Supervisory Control And Data Acquisition (SCADA) system was found in 2010 [9], where sensor measurements were replaced by previously recorded data and fed to the controller, thus leading to possible catastrophic damages; false data can be injected into smart power grids [10]; and an attacker can spoof the GPS to misguide an $80 million yacht off route [11].

To secure state estimation in *linear* dynamical systems, one can formulate a non-convex ℓ_0-minimization problem when sensor measurements are either noise-free [3, 4] or being corrupted by noise [6], which is then relaxed into a convex ℓ_r/ℓ_1 (sum

Y. Yang (✉) · G. Huang
Department of Mechanical Engineering, University of Delaware, Newark, DE 19716, USA
e-mail: yuyang@udel.edu

G. Huang
e-mail: ghuang@udel.edu

N. M. Amato et al. (eds.), *Robotics Research*, Springer Proceedings
in Advanced Robotics 10, https://doi.org/10.1007/978-3-030-28619-4_54

of ℓ_r norms) problem. In particular, Fawzi et al. [4] studied the secure estimation problem for a noiseless linear time invariant (LTI) system with a fixed set of attacked sensors which are less than one half of the total number of sensors, but the attack signals can be arbitrary. Pajic et al. [12, 13] extended [4] to noisy system with bounded noise assumption, and proved that the worst-case estimation error of their algorithms is linear with the bound of the noise. If there is no (processing) resource constraint, a minimax optimization can be formulated to construct an optimal estimator by minimizing the worst-case mean square error against *all possible* attacked sensors and *all possible* sensor noise [5, 8]. Moreover, in [3, 14] a *complete* set of fault-monitor filters are generated to detect the existence of an attack. However, if only an upper bound on the number of the attacked sensors is available, this method is *not* scalable since the number of monitors is combinatorial in the size of the attacked sensors. In [14] observability analysis was also performed for a linear system under attacks, showing that the system is observable if and only if less than a half of the sensors are attacked. In robotics, Bezzo et al. [15] introduced a secure Kalman filter (KF) for the LTI system by inflating the covariance of attacked sensors' measurements. Recently, Hu et al. [16] addressed secure localization for UAVs by using error correction techniques [17] to identify the attack signals based on the sparse attack assumption but relaxing the assumption of a fixed set of attack sensors and allowing different sets of sensors to be attacked each time. Additionally, in *noise-free* cases, Satisfiability Modulo Theory (SMT)-based algorithms can also be employed to detect and isolate the compromised sensors for both linear dynamical systems [7] and nonlinear differentially flat systems [18].

In this paper, we seek to secure state estimation for *stochastic nonlinear* systems with the particular application to map-based localization. In particular, based on the MCC-KF [19], we first perform in-depth analysis of the maximum correntropy criterion (MCC)-based **E**KF. Then, we analytically derive the weighted MCC-EKF (WMCC-EKF) that shows to improve accuracy and robustness to unbounded attacks as compared to the state-of-the-art methods. Different with [20], the proposed WMCC-EKF is derived for nonlinear measurement model and the weights are determined partially according to the known noise level. Furthermore, as a conservative solution, we generalize the secure estimation algorithm [16] to nonlinear systems and develop the Secure Estimation (SE)-EKF that integrates the attack detection within a sliding-window filtering framework. The proposed secure EKFs are validated through both Monte-Carlo simulations and experiments on real datasets.

2 Problem Statement

Consider a nonlinear system with measurements possibly attacked by adversaries:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{w}_k) \tag{1}$$

$$\mathbf{y}_{k+1} = \mathbf{h}(\mathbf{x}_{k+1}) + \mathbf{n}_{k+1} + \mathbf{a}_{k+1} \tag{2}$$

$$\mathbf{z}_{k+1} = \mathbf{y}_{k+1} - \mathbf{a}_{k+1} = \mathbf{h}(\mathbf{x}_{k+1}) + \mathbf{n}_{k+1} \tag{3}$$

where $\mathbf{x}_k \in \mathbb{R}^{m\times 1}$ represents the system states at the time step k, $\mathbf{f}$ represents the system dynamic model and $\mathbf{w}$ is the input white Gaussian noise with covariance $\mathbf{Q}$. $\mathbf{y} \in \mathbb{R}^{p\times 1}$ denotes the measurements from p sensors, $\mathbf{h}$ represents the nonlinear measurement model function. $\mathbf{a} \in \mathbb{R}^{p\times 1}$ denotes the attack signals and is assumed to be sparse vector that at least one sensor cannot be attacked. We also define $\mathbf{z} \in \mathbb{R}^{p\times 1}$ as the un-attacked output. $\mathbf{n} \in \mathbb{R}^{p\times 1}$ represents zero-mean Gaussian white noises with covariance $\mathbf{R} = \mathbf{diag}\{\sigma_1^2 \dots \sigma_i^2 \dots \sigma_p^2\}$, where $\sigma_i, i = 1 \dots p$ represents the i-th sensor's noise variance and $\mathbf{diag}\{\cdot\}$ is the diagonal matrix form. If the $\mathbf{R}$ is a full (not diagonal or block diagonal) matrix, a noise pre-whitening operation (see [21]) can be performed to transform $\mathbf{R}$ into diagonal form. The corresponding linearized system can be computed as follows:

$$\tilde{\mathbf{x}}_{k+1} \simeq \mathbf{F}_k \tilde{\mathbf{x}}_k + \mathbf{G}_k \mathbf{w}_k \tag{4}$$

$$\tilde{\mathbf{y}}_{k+1} \simeq \mathbf{H}_{k+1} \tilde{\mathbf{x}}_{k+1} + \mathbf{n}_{k+1} + \mathbf{a}_{k+1} \tag{5}$$

$$\tilde{\mathbf{z}}_{k+1} \simeq \mathbf{H}_{k+1} \tilde{\mathbf{x}}_{k+1} + \mathbf{n}_{k+1} \tag{6}$$

where $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ denotes the error states, the $\mathbf{F}_k$ and $\mathbf{G}_k$ represent the Jacobians regarding to the state $\mathbf{x}_k$ and the noise $\mathbf{w}_k$ respectively. $\tilde{\mathbf{y}}$ denotes the measurement residual, while $\tilde{\mathbf{z}}$ describes the un-attacked measurement residual. $\mathbf{H}_{k+1}$ represents the measurement Jacobian with respect to the state $\mathbf{x}_{k+1}$.

2.1 Map-Based Localization with Malicious Attacks

While this paper particularly focuses on 2D map-based localization as an example to illustrate the key ideas of our proposed secure estimators, the methodology is general and readily applicable to other systems. Specifically, in map-based localization, the dynamic motion model of the robot pose is given by:

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{\mathbf{p}}_R \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} v\cos(\phi) \\ v\sin(\phi) \\ \omega \end{bmatrix} = \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \\ 0 \end{bmatrix} v + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \omega \tag{7}$$

where $\mathbf{v}$ is the linear velocity and ω is the angular velocity of the robot. $\mathbf{p}_R$ and ϕ denote the position and orientation of the robot, respectively. Note that we assume a more challenging localization scenario than [13, 16] that the robot does not have access to GPS signals. Instead, only the relative range and bearing measurements of the features are available for localization, and the measurements can be described as:

$$\mathbf{h}(\mathbf{x}) = \begin{bmatrix} \mathbf{h}^{(r)}(\mathbf{x}) \\ \mathbf{h}^{(b)}(\mathbf{x}) \end{bmatrix} + \mathbf{a} = \begin{bmatrix} \sqrt{{}^s\mathbf{p}_\mathbf{f}^\top {}^s\mathbf{p}_\mathbf{f}} \\ \arctan\left(\frac{{}^s y_\mathbf{f}}{{}^s x_\mathbf{f}}\right) \end{bmatrix} + \mathbf{a} \tag{8}$$

where $\mathbf{h}^{(r)}$ and $\mathbf{h}^{(b)}$ represent the range and bearing measurements respectively. Given the rotation matrix $\mathbf{C}(\phi)$ between global and sensor frames, ${}^{s}\mathbf{p}_{\mathbf{f}} = [{}^{s}x_{\mathbf{f}} \ {}^{s}y_{\mathbf{f}}]^{\top} = \mathbf{C}(\phi)\,(\mathbf{p}_{\mathbf{f}} - \mathbf{p}_{R})$ represents the map feature in the sensor frame of reference.

It is important to note that, instead of assuming a fixed set of attacked sensors [4, 12], we consider that the attacker can attack different sensors randomly at different time steps [see (54)]. Note also that as compared to [15, 16], instead of assuming that less than a half of the sensors can be attacked, we only assume that at least one bearing or range sensor is not attacked. Moreover, attack signals can even go unbounded—that is, some of the sensor attacks $a_i\,(i \in \{1 \ldots p\})$ might go unbounded, i.e., $\|a_i\| \to \infty$.

3 Maximum Correntropy Criterion (MCC)-Based Filters

In this section, we present in detail our secure filters based on the maximum correntropy criterion. The correntropy can be defined as a statistical metric of similarity between two random variables [19], and one can pose a cost function $\mathbf{J}_m$ for robust filters based on the correntropy with Gaussian kernels as follows:

$$\mathbf{J}_m(\mathbf{x}_{k+1}) = \mathbf{G}_\sigma\left(\|\mathbf{y}_{k+1} - \mathbf{h}(\mathbf{x}_{k+1})\|_{\mathbf{R}_{k+1}^{-1}}\right) + \mathbf{G}_\sigma\left(\|\mathbf{x}_{k+1} - \mathbf{f}(\mathbf{x}_k, \mathbf{0})\|_{\mathbf{P}_{k+1|k}^{-1}}\right) \tag{9}$$

where $\mathbf{G}_\sigma$ is the Gaussian kernel in the form of $\mathbf{G}_\sigma(\|x_i - y_i\|) = \exp(-\frac{\|x_i - y_i\|^2}{2\sigma^2})$ with σ as bandwidth, $\mathbf{P}_{k+1|k}$ is the propagated covariance [see (11)]. Minimization of the cost function (9) can lead to the derivation of correntropy based filters [19]. Correntropy based filter is proved to be robust when having large disturbances or outliers and can work well with non-Gaussian noise.

3.1 MCC-EKF

Based on [19, 22], we analytically derive the MCC-EKF for the case of *nonlinear* systems such as map-based localization. In particular, given the initial state in the form of Gaussian distribution, $\mathcal{N}(\hat{\mathbf{x}}_{0|0}, \mathbf{P}_0)$, state estimate and covariance propagation based on the motion model (1) from time step k to $k+1$ is:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{f}(\hat{\mathbf{x}}_{k|k}, \mathbf{0}) \tag{10}$$

$$\mathbf{P}_{k+1|k} = \mathbf{F}_k \mathbf{P}_{k|k} \mathbf{F}_k^{\top} + \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^{\top} \tag{11}$$

Then, EKF-like update based on the measurement model (2) can be written as:

$$\hat{\mathbf{y}}_{k+1|k} = \hat{\mathbf{z}}_{k+1|k} = \mathbf{h}(\hat{\mathbf{x}}_{k+1|k}) \tag{12}$$

$$d_{k+1} = \frac{\mathbf{G}_\sigma\left(\left\|\mathbf{y}_{k+1} - \mathbf{h}(\hat{\mathbf{x}}_{k+1|k})\right\|_{\mathbf{R}_{k+1}^{-1}}\right)}{\mathbf{G}_\sigma\left(\left\|\hat{\mathbf{x}}_{k+1|k} - \mathbf{f}(\hat{\mathbf{x}}_{k|k}, \mathbf{0})\right\|_{\mathbf{P}_{k+1|k}^{-1}}\right)} \tag{13}$$

$$\mathbf{K}_{k+1|k} = \left(\mathbf{P}_{k+1|k}^{-1} + \mathbf{H}_{k+1}^\top (d_{k+1}\mathbf{R}_{k+1}^{-1})\mathbf{H}_{k+1}\right)^{-1} \mathbf{H}_{k+1}^\top (d_{k+1}\mathbf{R}_{k+1}^{-1}) \tag{14}$$

$$= \mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^\top \left(\mathbf{H}_{k+1}\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^\top + d_{k+1}^{-1}\mathbf{R}_{k+1}\right)^{-1} \tag{15}$$

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1|k}(\mathbf{y}_{k+1} - \hat{\mathbf{y}}_{k+1|k}) \tag{16}$$

$$\mathbf{P}_{k+1|k+1} = \left(\mathbf{P}_{k+1|k}^{-1} + \mathbf{H}_{k+1}^\top (d_{k+1}\mathbf{R}_{k+1}^{-1})\mathbf{H}_{k+1}\right)^{-1} \tag{17}$$

where d_{k+1} is a ratio scalar computed from Gaussian kernel. Based on these derivations, the detailed MCC-EKF algorithm can be found in the companion technical report [21]. With an in-depth inspection of the MCC-EKF, the updated covariance (17) can also be written as:

$$\mathbf{P}_{k+1|k+1} = \mathbf{P}_{k+1|k} - \mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^\top \mathbf{S}_{k+1|k}^{-1}\mathbf{H}_{k+1}\mathbf{P}_{k+1|k} \tag{18}$$

with the innovation covariance $\mathbf{S}_{k+1|k}$ defined as:

$$\mathbf{S}_{k+1|k} = \underbrace{\mathbf{H}_{k+1}\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^\top}_{\mathbf{S}_1} + \underbrace{d_{k+1}^{-1}\mathbf{R}_{k+1}}_{\mathbf{S}_2} \tag{19}$$

where $\mathbf{S}_1$ and $\mathbf{S}_2$ denote the covariance contribution from the motion (1) and measurement (2), respectively. Note that the MCC-EKF can be viewed as using the scalar d_{k+1} to control the covariance inflation from the attacked measurements. As shown in (13), d_{k+1} decreases if system has been attacked, and the covariance contribution $\mathbf{S}_2$ will be increased [see(19)], implying that the measurement becomes more uncertain. As a result, $\mathbf{S}_{k+1|k}$ and thus the updated state covariance $\mathbf{P}_{k+1|k+1}$, will be inflated due to (18). Lemma 1 summarizes our analysis:

Lemma 1 *For the MCC-EKF, if the attack* $\mathbf{a}_{k+1}$ *goes unbounded, the filter will not perform measurement update.*

Proof If the attack goes unbounded, that is $\|\mathbf{a}_{k+1}\| \to \infty$, then $\|\mathbf{y}_{k+1} - \mathbf{h}(\hat{\mathbf{x}}_{k+1|k})\|_{\mathbf{R}_{k+1}^{-1}} \to \infty$, and hence $d_k \to 0$. According to (14) and (16), $\mathbf{K}_{k+1} \to \mathbf{0}$ and $\hat{\mathbf{x}}_{k+1|k+1} \to \hat{\mathbf{x}}_{k+1|k}$. Finally, with (17), $\mathbf{P}_{k+1|k+1} \to \mathbf{P}_{k+1|k}$.

This result essentially shows that the scalar d_{k+1} will dismiss *all* the observation updates even if only one measurement is attacked at time step $k+1$, which clearly is too conservative. In order to enable the MCC-EKF to utilize the information contained in un-attacked measurements, we propose the *weighted* MCC-EKF derived from multiple Gaussian kernels.

3.2 Weighted MCC-EKF

Compared to (9), we define the cost function for the maximum correntropy criterion with multiple Gaussian kernels as:

$$\mathbf{J}(\mathbf{x}_{k+1}) = \sum_{i=1}^{p} \mathbf{G}_{\hat{\sigma}_{i,k+1}} \left(\left\| \mathbf{y}_{i,k+1} - \mathbf{h}_{i,k+1}(\mathbf{x}_{k+1}) \right\| \right) + \mathbf{G}_{\hat{\sigma}_{0,k+1}} \left(\left\| \mathbf{x}_{k+1} - \mathbf{f}(\hat{\mathbf{x}}_{k|k}, \mathbf{0}) \right\|_{\mathbf{P}_{k+1|k}^{-1}} \right) \tag{20}$$

where we have defined the Gaussian kernel $\mathbf{G}_{\hat{\sigma}_{i,k+1}}$ and $\mathbf{G}_{\hat{\sigma}_{0,k+1}}$ according to [19]:

$$\mathbf{G}_{\hat{\sigma}_{i,k+1}} \left(\left\| \mathbf{y}_{i,k+1} - \mathbf{h}_{i,k+1}(\mathbf{x}_{k+1}) \right\| \right) = \exp\left(- \frac{\left\| \mathbf{y}_{i,k+1} - \mathbf{h}_{i,k+1}(\mathbf{x}_{k+1}) \right\|^2}{2\hat{\sigma}_{i,k+1}^2} \right) \tag{21}$$

$$\mathbf{G}_{\hat{\sigma}_{0,k+1}} \left(\left\| \mathbf{x}_{k+1} - \mathbf{f}(\hat{\mathbf{x}}_{k|k}, \mathbf{0}) \right\|_{\mathbf{P}_{k+1|k}^{-1}} \right) = \exp\left(- \frac{\left\| \mathbf{x}_{k+1} - \mathbf{f}(\hat{\mathbf{x}}_{k|k}, \mathbf{0}) \right\|_{\mathbf{P}_{k+1|k}^{-1}}^2}{2\hat{\sigma}_{0,k+1}^2} \right) \tag{22}$$

where $\hat{\sigma}_{i,k+1}, i = 1 \ldots p$ denotes the Gaussian kernel bandwidth of the i-th measurement at time step $k+1$, and $\hat{\sigma}_{0,k+1}$ denotes the Gaussian kernel bandwidth of the motion model. $\mathbf{y}_{i,k+1}$ and $\mathbf{h}_{i,k+1}(\mathbf{x}_{k+1})$ represents the i-th row of $\mathbf{y}_{k+1}$ and $\mathbf{h}_{k+1}$. Aiming to meet the maximum correntropy criterion, we linearize and take the derivatives of the cost function $\mathbf{J}(\mathbf{x}_{k+1})$ as:

$$\frac{\partial \mathbf{J}(\mathbf{x}_{k+1})}{\partial \tilde{\mathbf{x}}_{k+1}} \simeq -\frac{1}{2} \sum_{i=1}^{p} \frac{\mathbf{G}_{\hat{\sigma}_{i,k+1}}}{\hat{\sigma}_{i,k+1}^2} \frac{\partial \left(\left\| \tilde{\mathbf{y}}_{i,k+1} - \mathbf{H}_{i,k+1} \tilde{\mathbf{x}}_{k+1} \right\|^2 \right)}{\partial \tilde{\mathbf{x}}_{k+1}} - \frac{1}{2} \frac{\mathbf{G}_{\hat{\sigma}_{0,k+1}}}{\hat{\sigma}_{0,k+1}^2} \frac{\partial \left(\left\| \tilde{\mathbf{x}}_{k+1} \right\|_{\mathbf{P}_{k+1|k}^{-1}}^2 \right)}{\partial \tilde{\mathbf{x}}_{k+1}} = 0 \tag{23}$$

where $\mathbf{H}_{i,k+1}, i = 1 \ldots p$, represents each row of the Jacobian $\mathbf{H}_{k+1} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}_{k+1}} \Big|_{\mathbf{x}_{k+1} = \hat{\mathbf{x}}_{k+1}}$ and $\tilde{\mathbf{x}}_{k+1} = \mathbf{x}_{k+1|k} - \mathbf{f}(\hat{\mathbf{x}}_{k|k}, \mathbf{0}) = \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k}$. Then we can arrive at:

$$\sum_{i=1}^{p} \frac{\mathbf{G}_{\hat{\sigma}_{i,k+1}}}{\mathbf{G}_{\hat{\sigma}_{0,k+1}}} \frac{\mathbf{H}_{i,k+1}^{\top} \mathbf{H}_{i,k+1}}{\frac{\hat{\sigma}_{i,k+1}^2}{\hat{\sigma}_{0,k+1}^2}} \tilde{\mathbf{x}}_{k+1} - \sum_{i=1}^{p} \frac{\mathbf{G}_{\hat{\sigma}_{i,k+1}}}{\mathbf{G}_{\hat{\sigma}_{0,k+1}}} \frac{\mathbf{H}_{i,k+1}^{\top}}{\frac{\hat{\sigma}_{i,k+1}^2}{\hat{\sigma}_{0,k+1}^2}} \tilde{\mathbf{y}}_{i,k+1} + \mathbf{P}_{k+1|k}^{-1} \tilde{\mathbf{x}}_{k+1} = \mathbf{0} \tag{24}$$

$$\Rightarrow \left[\sum_{i=1}^{p} \frac{\mathbf{G}_{\hat{\sigma}_{i,k+1}}}{\mathbf{G}_{\hat{\sigma}_{0,k+1}}} \frac{\mathbf{H}_{i,k+1}^{\top} \mathbf{H}_{i,k+1}}{\frac{\hat{\sigma}_{i,k+1}^2}{\hat{\sigma}_{0,k+1}^2}} + \mathbf{P}_{k+1|k}^{-1} \right] \tilde{\mathbf{x}}_{k+1} = \sum_{i=1}^{p} \frac{\mathbf{G}_{\hat{\sigma}_{i,k+1}}}{\mathbf{G}_{\hat{\sigma}_{0,k+1}}} \frac{\mathbf{H}_{i,k+1}^{\top}}{\frac{\hat{\sigma}_{i,k+1}^2}{\hat{\sigma}_{0,k+1}^2}} \tilde{\mathbf{y}}_{i,k+1} \tag{25}$$

Then (25) can be written in matrix form as:

$$\left[\mathbf{H}_{k+1}^{\top} \mathbf{D}_{k+1} \hat{\mathbf{R}}_{k+1}^{-1} \mathbf{H}_{k+1} + \mathbf{P}_{k+1|k}^{-1} \right] \tilde{\mathbf{x}}_{k+1} = \mathbf{H}_{k+1}^{\top} \mathbf{D}_{k+1} \hat{\mathbf{R}}_{k+1}^{-1} \tilde{\mathbf{y}}_{k+1} \tag{26}$$

where we have defined $d_{i,k+1}$, $\mathbf{D}_{k+1}$ and $\hat{\mathbf{R}}_{k+1}$ as:

$$d_{i,k+1} = \frac{\mathbf{G}_{\hat{\sigma}_{i,k+1}}\left(\left\|\mathbf{y}_{i,k+1} - \mathbf{h}_{i,k+1}(\mathbf{x}_{k+1})\right\|\right)}{\mathbf{G}_{\hat{\sigma}_{0,k+1}}\left(\left\|\mathbf{x}_{k+1} - \mathbf{f}(\hat{\mathbf{x}}_{k|k}, \mathbf{0})\right\|_{\mathbf{P}^{-1}_{k+1|k}}\right)} \tag{27}$$

$$\mathbf{D}_{k+1} = \mathbf{diag}\{d_{1,k+1}, \ldots, d_{i,k+1}, \ldots, d_{p,k+1}\} \tag{28}$$

$$\hat{\mathbf{R}}_{k+1} = \mathbf{diag}\{\frac{\hat{\sigma}^2_{1,k+1}}{\hat{\sigma}^2_{0,k+1}}, \ldots, \frac{\hat{\sigma}^2_{i,k+1}}{\hat{\sigma}^2_{0,k+1}}, \ldots, \frac{\hat{\sigma}^2_{p,k+1}}{\hat{\sigma}^2_{0,k+1}}\} \tag{29}$$

Hence, the new state and covariance update can be expressed as:

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \left[\mathbf{H}^\top_{k+1}\mathbf{D}_{k+1}\hat{\mathbf{R}}^{-1}_{k+1}\mathbf{H}_{k+1} + \mathbf{P}^{-1}_{k+1|k}\right]^{-1}\mathbf{H}^\top_{k+1}\mathbf{D}_{k+1}\hat{\mathbf{R}}^{-1}_{k+1}\left(\mathbf{y}_{k+1} - \hat{\mathbf{y}}_{k+1|k}\right) \tag{30}$$

$$\mathbf{P}_{k+1|k+1} = \left[\mathbf{H}^\top_{k+1}\mathbf{D}_{k+1}\hat{\mathbf{R}}^{-1}_{k+1}\mathbf{H}_{k+1} + \mathbf{P}^{-1}_{k+1|k}\right]^{-1} \tag{31}$$

Up to this step, we have the new state update as (30), which is highly similar to (16). Now comes how to choose appropriate bandwidths. We fixed the ratio of $\frac{\hat{\sigma}^2_i}{\hat{\sigma}^2_0}$ as σ^2_i, where σ_i denotes the standard deviation of the i-th measurement obtained from noise covariance $\mathbf{R}_{k+1}$. Therefore, $\hat{\mathbf{R}}_{k+1} = \mathbf{R}_{k+1}$, and $\mathbf{D}_{k+1}$ can just be seen as a weight matrix for the measurement noise. During the implementation of the WMCC-EKF [21], we choose $\sigma^2_i = \lambda_\sigma \hat{\sigma}^2_i$, with $\lambda_\sigma \in (0.125, 0.5)$ which are shown to work well in our simulation and experiments. Upon this choice, the state and covariance update of the proposed WMCC-EKF can be finally described as:

$$\mathbf{K}_{k+1|k} = \left[\mathbf{H}^\top_{k+1}\mathbf{D}_{k+1}\mathbf{R}^{-1}_{k+1}\mathbf{H}_{k+1} + \mathbf{P}^{-1}_{k+1|k}\right]^{-1}\mathbf{H}^\top_{k+1}\mathbf{D}_{k+1}\mathbf{R}^{-1}_{k+1} \tag{32}$$

$$= \mathbf{P}_{k+1|k}\mathbf{H}^\top_{k+1}\left(\mathbf{H}_{k+1}\mathbf{P}_{k+1|k}\mathbf{H}^\top_{k+1} + \mathbf{R}_{k+1}\mathbf{D}^{-1}_{k+1}\right)^{-1} \tag{33}$$

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1|k}\left(\mathbf{y}_{k+1} - \hat{\mathbf{y}}_{k+1|k}\right) \tag{34}$$

$$\mathbf{P}_{k+1|k+1} = \left[\mathbf{H}^\top_{k+1}\mathbf{D}_{k+1}\mathbf{R}^{-1}_{k+1}\mathbf{H}_{k+1} + \mathbf{P}^{-1}_{k+1|k}\right]^{-1} \tag{35}$$

Now we will inspect WMCC-EKF from an information perspective. Compared to the MCC-EKF, the information matrix for the WMCC-EKF can be written as:

$$\mathbf{P}^{-1}_{k+1|k+1} = \mathbf{P}^{-1}_{k+1|k} + \mathbf{H}^\top_{k+1}(\mathbf{D}_{k+1}\mathbf{R}^{-1}_{k+1})\mathbf{H}_{k+1} = \underbrace{\mathbf{P}^{-1}_{k+1|k}}_{\boldsymbol{\Sigma}_{w1}} + \underbrace{\sum_{i=1}^{p} d_{i,k+1}\frac{\mathbf{H}^\top_{i,k+1}\mathbf{H}_{i,k+1}}{\sigma^2_{i,k+1}}}_{\boldsymbol{\Sigma}_{w2}} \tag{36}$$

where Σ_{w1} and Σ_{w2} denote the information from motion model (1) and the measurement model (2), respectively. Note that $d_{i,k+1}\frac{\mathbf{H}^\top_{i,k+1}\mathbf{H}_{i,k+1}}{\sigma^2_{i,k+1}}$ represents the information contribution from the i-th sensor's measurement, and thus, Σ_{w2} in (36) can be seen as the sum of single information matrix from all the p sensors. If the i-th sensor is

attacked, $d_{i,k+1}$ will decrease exponentially and the corresponding information contribution $d_{i,k+1}\frac{\mathbf{H}_{i,k+1}^{\top}\mathbf{H}_{i,k+1}}{\sigma_{i,k+1}^{2}}$ will be dramatically reduced. However, this process will not affect the information contribution from other sensors. Therefore, different from the MCC-EKF, the WMCC-EKF is able to utilize the information from un-attacked sensor measurements.

3.3 Convergence Analysis Under Unbounded Attacks

Inspired by [15], to further understand the proposed WMCC-EKF, we perform the convergence analysis when the system is suffering from unbounded attacks. We first define $\bar{\mathbf{x}}_{k+1}$ as the state estimate with un-attacked measurement $\mathbf{z}_{k+1}$, and the predicted measurement based on $\bar{\mathbf{x}}_{k+1}$ can be denoted as:

$$\bar{\mathbf{z}}_{k+1} = \mathbf{h}(\bar{\mathbf{x}}_{k+1}) \tag{37}$$

Hence, with (2) and (3), the update Eq. (34) can be rewritten as:

$$\begin{aligned}\hat{\mathbf{x}}_{k+1|k+1} &= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1|k}(\mathbf{z}_{k+1} - \bar{\mathbf{z}}_{k+1} + \mathbf{h}(\bar{\mathbf{x}}_{k+1}) - \mathbf{h}(\hat{\mathbf{x}}_{k+1|k}) + \mathbf{a}_{k+1}) \quad (38)\\ &= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1|k}(\mathbf{z}_{k+1} - \bar{\mathbf{z}}_{k+1}) + \mathbf{K}_{k+1|k}\mathbf{s}_{k+1} \quad (39)\end{aligned}$$

where $\mathbf{s}_{k+1} = \mathbf{h}(\bar{\mathbf{x}}_{k+1}) - \mathbf{h}(\hat{\mathbf{x}}_{k+1|k}) + \mathbf{a}_{k+1}$ describes the difference of measurement estimates from un-attacked and attacked measurements. Since $\mathbf{s}_{k+1}$ also includes the attack vector $\mathbf{a}_{k+1}$, the term $\mathbf{K}_{k+1|k}\mathbf{s}_{k+1}$ can be seen as *Attack Innovation*. We would like to shrink this term, so that the attacked estimate $\hat{\mathbf{x}}_{k+1|k+1}$ will approach the ideal estimate $\bar{\mathbf{x}}_{k+1}$ as close as possible. Interestingly, the WMCC-EKF can constrain the attack innovation to a small bound even under unbounded attacks.

Lemma 2 *Given an unbounded attack* $\mathbf{a}_{k+1}$ *and an arbitrarily small positive constant value* ξ*, there exists a correntropy weight matrix* $\mathbf{D}_{k+1}$ *for the WMCC-EKF such that:*

$$\mathbf{Pr}\left(\left\|\mathbf{K}_{k+1|k}\mathbf{s}_{k+1}\right\|^{2} \leq \xi\right) > 99.7\% \tag{40}$$

Proof From (33), we can write attack innovation $\mathbf{K}_{k+1|k}\mathbf{s}_{k+1}$ as:

$$\left\|\mathbf{K}_{k+1|k}\mathbf{s}_{k+1}\right\|^{2} = \left\|\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^{\top}\right\|^{2}\|\tau\|^{2} \tag{41}$$

where we define $\tau = \left(\mathbf{H}_{k+1}\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^{\top} + \mathbf{D}_{k+1}^{-1}\mathbf{R}_{k+1}\right)^{-1}\mathbf{s}_{k+1}$. We can observe that in oder to show bounded attack innovation, we only need to show that $\|\tau\|$ is bounded. We consider the worst case and compute the boundary for $\|\tau\|$ as:

$$\|\tau\|^2 \leq \left\| \left(\sigma_{\min}^2 \mathbf{I} + \mathbf{D}_{k+1}^{-1} \mathbf{R}_{k+1}\right)^{-1} \mathbf{s}_{k+1} \right\|^2 = \sum_{j=1}^{p} \left(\frac{s_j}{\sigma_{\min}^2 + d_j^{-1} \sigma_j^2} \right)^2 \tag{42}$$

We define the ideal estimate residual as $\tilde{\mathbf{z}}_{j,k+1} = \mathbf{z}_{j,k+1} - \hat{\mathbf{z}}_{j,k+1}$, and $\tilde{\mathbf{z}}_{j,k+1} \sim \mathcal{N}(\mathbf{0}, \bar{\sigma}_{j,k+1}^2)$. Based on Gaussian distribution, we have:

$$\mathbf{Pr}\left(\left\| \tilde{z}_{j,k+1} \right\| \leq 3\bar{\sigma}_{j,k+1} \right) = 99.7\% \tag{43}$$

Equation (43) indicates that $\left\| \tilde{z}_{j,k+1} \right\|$ is almost bounded by $3\bar{\sigma}_{j,k+1}$. If the j-th sensor attack a_j goes unbounded, $\left\| s_j \right\| \to \infty$ and hence $\left\| s_j \right\| > 3\bar{\sigma}_j$. Then, we drop the timestamps for simplicity and arrive at:

$$\left[\frac{s_j}{\sigma_{\min}^2 + d_j^{-1} \sigma_j^2} \right]^2 \leq \left[\frac{s_j}{\sigma_{\min}^2 + \exp\left(\frac{(\|s_j\| - \|\tilde{z}_j\|)^2}{2\hat{\sigma}_j^2} \right) \sigma_j^2} \right]^2 < \frac{\hat{\sigma}_j^2}{\sigma_j^4} \frac{\zeta^2}{\left[\exp\left(\frac{1}{2} (\zeta - \mu)^2 \right) \right]^2} \tag{44}$$

where $\zeta = \frac{\|s_j\|}{\hat{\sigma}_j}$, and $\mu = 3\frac{\bar{\sigma}_j}{\hat{\sigma}_j}$. Obviously, as $\left\| s_j \right\| \to \infty$, $\zeta \to \infty$, and the right side of (44) will finally approach 0. Besides, if we take derivative of the right side of (44) regarding to ζ, we can have the maximum value of (44) when $\zeta' = \frac{\mu + \sqrt{\mu^2 + 4}}{2}$, that is:

$$\left[\frac{s_j}{\sigma_{\min}^2 + d_j^{-1} \sigma_j^2} \right]^2 \leq \frac{\hat{\sigma}_j^2}{\sigma_j^4} \frac{\zeta'^2}{\left[\exp\left(\frac{1}{2} (\zeta' - \mu)^2 \right) \right]^2} \tag{45}$$

Since ζ' is independent of the attack innovation s_j, thus we can bound (45) by appropriate design of bandwidth $\hat{\sigma}_j$. According to (42) and (43), $\|\tau\|^2$ is the summation of (45) and is bounded by the design of $\mathbf{D}_{k+1}$ with probability 99.7%. In (41), $\left\| \mathbf{P}_{k+1|k} \mathbf{H}_{k+1}^{\top} \right\|^2$ is independent from the $\mathbf{a}_{k+1}$, and thus it is bounded. Therefore, we can easily find a ξ that satisfies (40).

4 Secure Estimation (SE)-EKF

Ideally, we would like to identify the attacked measurements so that we can ensure estimation security by excluding them from the EKF update. To this end, we introduce the Secure-estimation (SE)-EKF by generalizing the SE-KF [16, 23] to the nonlinear system under consideration. In particular, in order to detect sensor attacks, we adopt the sliding-window strategy. Specifically, we construct a fixed-sized window within EKF framework by stochastic cloning [24]. All the accumulated measurements within the window are used for update at certain time step. After update, the window will be cleared and start to accumulate new measurements again. We define the state vector with window size N at time step k as:

$$\mathbf{x}_{c_k} = \begin{bmatrix} \mathbf{x}_k^\top & \mathbf{x}_{k-1}^\top & \cdots & \mathbf{x}_{k-N+1}^\top & \mathbf{x}_{k-N}^\top \end{bmatrix}^\top \tag{46}$$

where $\mathbf{x}_k$ represents the current robot state, $\mathbf{x}_{k-i}$ represents the cloned robot state at time step $k-i$, $i \in \{1 \dots N\}$. Thus, $\mathbf{x}_{k-N}$ is the oldest cloned state. Similar to SE in [16], after we have cloned N robot states in the state vector and accumulated their measurements, we can linearize and stack all the measurements together as:

$$\begin{bmatrix} \tilde{\mathbf{z}}_k \\ \tilde{\mathbf{z}}_{k-1} \\ \vdots \\ \tilde{\mathbf{z}}_{k-N} \end{bmatrix} \simeq \begin{bmatrix} \mathbf{H}_k \\ \mathbf{H}_{k-1} \\ \vdots \\ \mathbf{H}_{k-N} \end{bmatrix} \tilde{\mathbf{x}}_{c_k} + \begin{bmatrix} \mathbf{n}_k \\ \mathbf{n}_{k-1} \\ \vdots \\ \mathbf{n}_{k-N} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_k \\ \mathbf{a}_{k-1} \\ \vdots \\ \mathbf{a}_{k-N} \end{bmatrix} \tag{47}$$

According to the linearized motion model (4), within the sliding-window, we have

$$\tilde{\mathbf{x}}_k = \mathbf{F}_{k-1} \cdots \mathbf{F}_{k-N} \tilde{\mathbf{x}}_{k-N} = \mathbf{F}_{k-1,k-N} \tilde{\mathbf{x}}_{k-N} \tag{48}$$

where $\mathbf{F}_{k-1,k-N} = \mathbf{F}_{k-1} \cdots \mathbf{F}_{k-N}$ represents the state transition matrix from cloned state $\tilde{\mathbf{x}}_{k-N}$ to the current robot state $\tilde{\mathbf{x}}_k$. Thus, (47) can be written as:

$$\underbrace{\begin{bmatrix} \tilde{\mathbf{z}}_k \\ \tilde{\mathbf{z}}_{k-1} \\ \vdots \\ \tilde{\mathbf{z}}_{k-N} \end{bmatrix}}_{\tilde{\mathbf{Z}}} \simeq \underbrace{\begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_{k-1} \\ \vdots \\ \mathbf{H}_{k-N} \end{bmatrix} \begin{bmatrix} \mathbf{F}_{k,k-N} \\ \mathbf{F}_{k-1,k-N} \\ \vdots \\ \mathbf{I} \end{bmatrix}}_{\boldsymbol{\Phi}} \tilde{\mathbf{x}}_{k-N} + \underbrace{\begin{bmatrix} \mathbf{n}_0 \\ \mathbf{n}_{k-1} \\ \vdots \\ \mathbf{n}_{k-N} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_{k-1} \\ \vdots \\ \mathbf{a}_{k-N} \end{bmatrix}}_{\mathbf{E}} \tag{49}$$

where $\tilde{\mathbf{Z}}$ represents the stacked measurement residuals, and $\mathbf{E}$ denotes the sum of stacked noise and attack vectors, $\boldsymbol{\Phi}$ denotes the stacked state transition matrix from $\tilde{\mathbf{x}}_{k-N}$ to each state in the window. Similar to [16, 23] we apply left null space operation to $\boldsymbol{\Phi}$ to simplify (49). Let $\mathbf{U}_n$ be the left null space of $\boldsymbol{\Phi}$, that is $\mathbf{U}_n^\top \boldsymbol{\Phi} = \mathbf{0}$, then we can have:

$$\mathbf{Z}_o = \mathbf{U}_n^\top \mathbf{Z} = \mathbf{U}_n^\top \mathbf{E} \tag{50}$$

where $\mathbf{U}_n$ can be computed from the QR decomposition of $\boldsymbol{\Phi}$ as:

$$\boldsymbol{\Phi} = \mathbf{U}_e \mathbf{R}_\Delta = \begin{bmatrix} \mathbf{U}_e & \mathbf{U}_n \end{bmatrix} \begin{bmatrix} \mathbf{R}_\Delta \\ \mathbf{0} \end{bmatrix} \tag{51}$$

Given the strong sparse attack assumption that less than a half of the all the sensors can be attacked, $\mathbf{E}$ can be solved by formulating the following optimization problem with ℓ_1 norm regularization [25] as:

$$\hat{\mathbf{E}} = \arg\min_{\mathbf{E}} \left[\left\| \mathbf{Z}_o - \mathbf{U}_n^\top \mathbf{E} \right\|_2^2 + \lambda \left\| \mathbf{E} \right\|_{\ell_1} \right] \tag{52}$$

where λ is the regularization parameter.

Different from [16], we here consider a *nonlinear* model and thus, the sparsity of $\mathbf{E}$ will be contaminated by linearization errors and noises. Therefore, the ℓ_1-optimization solution $\hat{\mathbf{E}}$ from (52) will not be perfectly sparse. In order to minimize this side effect, we propose to set a threshold t for $\hat{\mathbf{E}}$ to enforce the sparsity. Let e_i denotes the i-th element in $\mathbf{E}$, and if $e_i < t$, we set $e_i = 0$ and assume no attack to the i-th element; otherwise e_i will keep its value and the i-th element is labeled as attacking signal. Let a_i and n_i denote the corresponding i-th element of the noise and attack vector respectively. If the i-th measurement is not attacked ($a_i = 0$), then:

$$\|e_i\| = \|n_i + a_i\| \leq \|n_i\| + \|a_i\| \leq \|n_i\| \tag{53}$$

Based on the white Gaussian noise assumption [*i.e.*, $n_i \sim \mathcal{N}(0, \sigma_i^2)$], we have $\mathbf{Pr}(\|n_i\| \leq 3\sigma_i) = 99.7\%$. Considering the linearization errors, we set the threshold $t_i = \lambda_t \sigma_i$ where $\lambda_t \in (3, 6)$ is used in our simulations. With the attack identification, the SE-EKF algorithm will be able to remove the attacked measurements and perform the state update only with un-attacked measurements [21].

5 Simulation Results

To validate the proposed secure estimators, we consider a map-based localization scenario where a mobile robot moves in a circle trajectory. There are 120 landmarks randomly generated near the trajectory as the map. We assume that the robot is equipped with 4 sensors: 2 range sensors and 2 bearing sensors, and these sensors collect independent range and bearing measurements of the map points when the robot is moving on the trajectory.

Moreover, we consider 3 different attack modes (54), where Attack Mode i ($i = 1 \ldots 3$) represents the attack signals received by the 4 sensors, and each column represents a time step. a_* denotes non-zero arbitrary or unbounded attack signals and $\mathbf{0}$ indicates no attack. Note that at each time step the senors might be attacked with the probability from 33% to 50%. If attacked, there are i attacked sensors for Attack Mode i, and the set of attacked sensors are changing randomly over time.

$$\left.\begin{array}{l} Sensor\ 1: range \\ Sensor\ 2: bearing \\ Sensor\ 3: range \\ Sensor\ 4: bearing \end{array}\right\} \Leftarrow \underbrace{\begin{bmatrix} a_* & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & a_* & \cdots \\ \mathbf{0} & a_* & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & a_* & \mathbf{0} & \cdots \end{bmatrix}}_{\textbf{Attack Mode 1}}, \underbrace{\begin{bmatrix} a_* & a_* & \mathbf{0} & \mathbf{0} & \cdots \\ a_* & \mathbf{0} & \mathbf{0} & a_* & \cdots \\ \mathbf{0} & a_* & a_* & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & a_* & a_* & \cdots \end{bmatrix}}_{\textbf{Attack Mode 2}}, \underbrace{\begin{bmatrix} a_* & a_* & \mathbf{0} & a_* & \cdots \\ a_* & \mathbf{0} & a_* & a_* & \cdots \\ a_* & a_* & a_* & \mathbf{0} & \cdots \\ \mathbf{0} & a_* & a_* & a_* & \cdots \end{bmatrix}}_{\textbf{Attack Mode 3}} \tag{54}$$

We also define 3 types of attack distribution: constant attack $a_* = c$, uniform attack $a_* \sim \mathcal{U}[-c, c]$, and the Gaussian distribution $a_* \sim \mathcal{N}(0, c^2)$. For the results presented below, c is set to 1 m for range measurement and is 0.5 for bearing measurement if not specified.

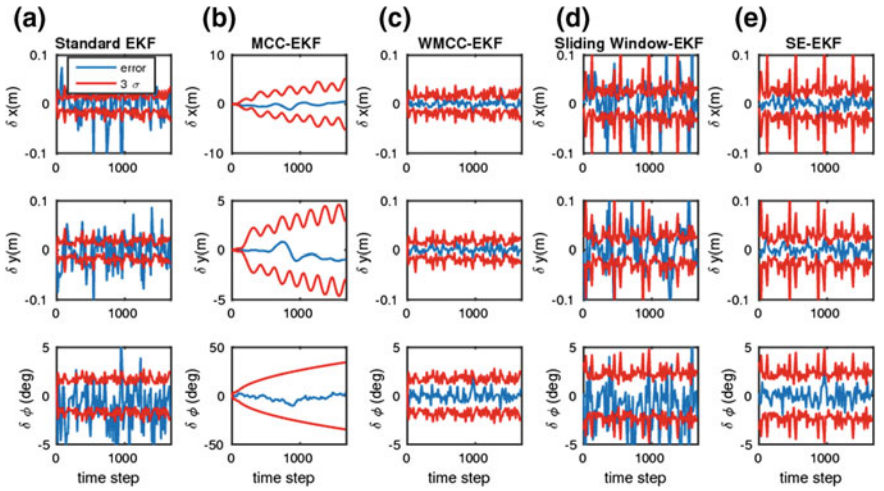

Fig. 1 Comparison of the Standard EKF, MCC-EKF, WMCC-EKF, Sliding Window-EKF and SE-EKF under attacks

Figure 1 shows the estimation errors of the Standard EKF, MCC-EKF, WMCC-EKF, Sliding Window-EKF and SE-EKF. The attacks are following Attack Mode 1 with constant attacks. We can see that the Standard EKF and Sliding Window-EKF have failed. Although the MCC-EKF can still work, the accuracy is much worse than that of the WMCC-EKF and the SE-EKF, which demonstrates the superior performance of the proposed estimators.

Note the SE can have stable performance [16] if and only if the attacked sensors number satisfies $q \leq p/2 - 1$, where p is the number of sensors and q is the number of attacked sensors. But we have relaxed this assumption for the WMCC-EKF, and Monte-Carlo tests are performed with different numbers of attacked sensors to test the full capacity of these proposed algorithms. Figure 2 shows the results of 50 Monte-Carlo runs with constant attacks of Attack Mode 1, 2 and 3. Normalized estimation error squared (NEES) and root mean square error (RMSE) [26] are used for evaluating the estimation consistency and accuracy. Clearly, the SE-EKF can only work when one of the four sensors is attacked, which conforms to [16]. In contrast, the WMCC-EKF can still perform well even when there are three out of four randomly attacked sensors.

We have also implemented the EKF with Mahalanobis-distance (M-distance) test for outliers rejection, and compared its performance with the WMCC-EKF. The M-distance test is a common outliers rejection strategy, given by:

$$d_m = \mathbf{r}^\top \mathbf{S}^{-1} \mathbf{r} \tag{55}$$

where $\mathbf{r}$ is the measurement residual and $\mathbf{S}$ is the corresponding innovation covariance. The d_m is assumed to follow the χ^2 distribution, thus we can define a threshold

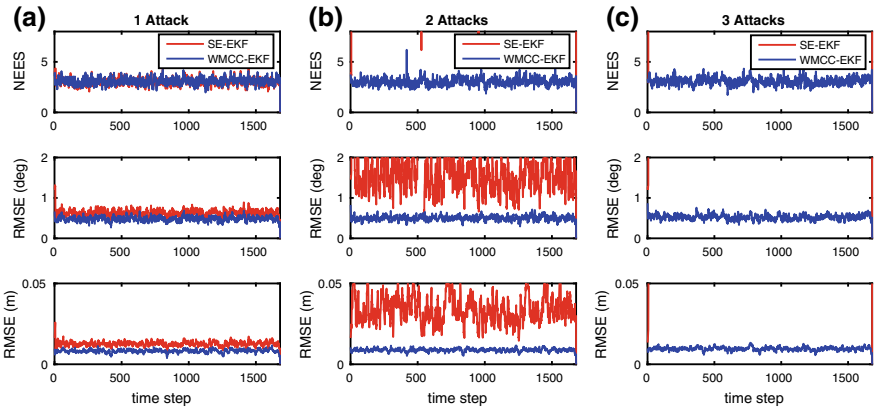

Fig. 2 Full capacity test of the WMCC-EKF and SE-EKF in 50 Monte-Carlo simulations

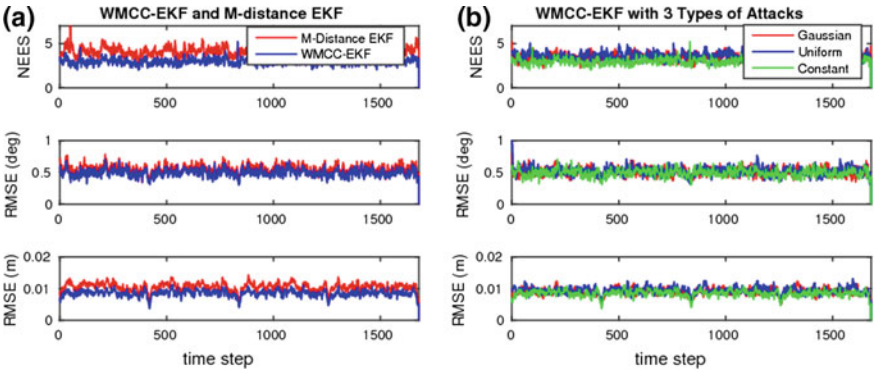

Fig. 3 **a** Comparison of the WMCC-EKF and M-distance EKF under attacks; **b** Performance of WMCC-EKF with Gaussian, uniform and constant attacks

γ for d_m to identify outliers. We perform 50 Monte-Carlo runs (Fig. 3) with both the WMCC-EKF and the M-distance based EKF. Note that the Attack Mode 1 with constant attack is applied, and the overall average NEES for the WMCC-EKF is approximately 2.97 while for M-distance based EKF is around 4.16. This shows that the proposed WMCC-EKF achieves better consistency than the M-distance test based EKF. In addition, the WMCC-EKF is shown to achieve slightly better estimation accuracy.

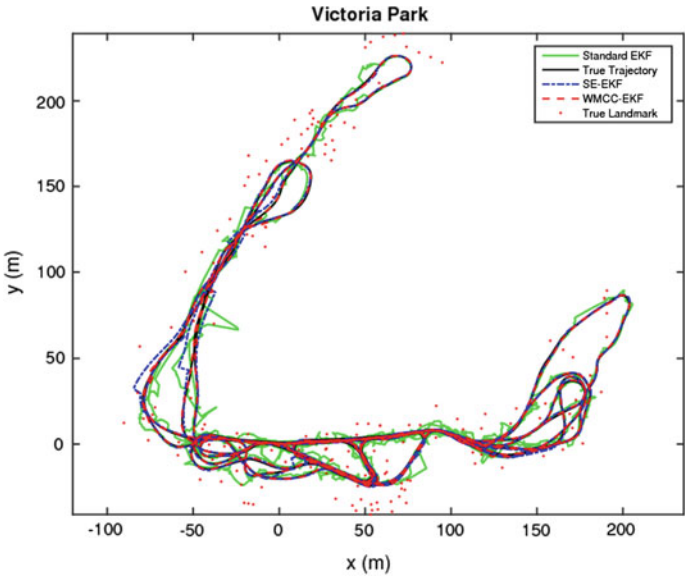

Fig. 4 Estimated trajectories of the WMCC-EKF, SE-EKF and the Standard EKF with synthetic attacks on the Victoria Park dataset

6 Experimental Results

We further test the proposed WMCC-EKF and SE-EKF with the Victoria Park dataset [27], which includes wheel odometry and 2D range-bearing observations to landmarks (trees). Specifically, we first run a batch MAP optimization using GTSAM [28] to generate both the car trajectory and the map, which are used as the ground truth. Based on this map, we validate our proposed algorithms for map-based localization. During the test, we synthetically add random attacks to the range-bearing measurements with 20% probability at each time step. Both range and bearing attack signals follows a uniform distribution, with magnitude c of 15 m for range and 0.5 for bearing, respectively. It is clear from Figs. 4 and 5 that the green trajectory estimated by the Standard EKF is not acceptable, while the blue and red trajectories estimated by the proposed WMCC- and SE-EKF are close to the true trajectory, which verify that the proposed algorithms are able to secure the robot localization.

7 Conclusions and Future Work

In this paper, we have developed the weighted MCC-EKF to secure state estimation for stochastic nonlinear systems under adversarial attacks. The key idea of this method is to design proper weights to inflate the possibly-compromised measure-

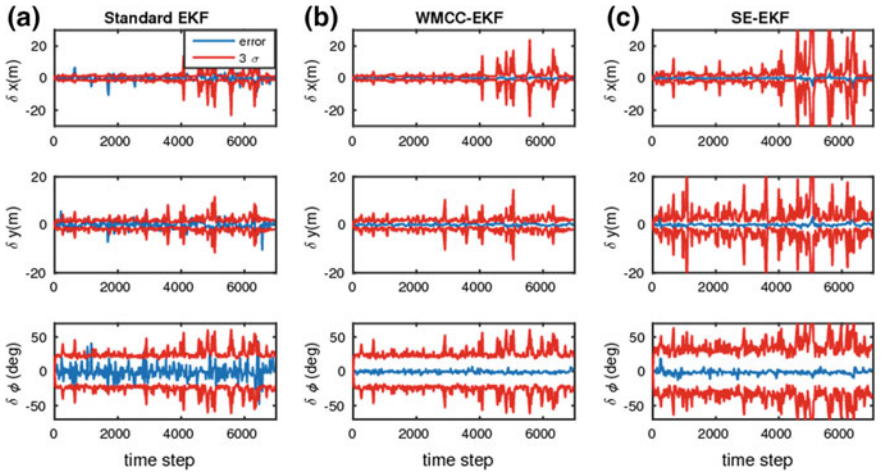

Fig. 5 Estimation errors of the WMCC-EKF, SE-EKF and the Standard EKF with synthetic attacks on the Victoria Park dataset

ments. Moreover, we have also extended the SE-KF from linear to nonlinear cases and proposed the SE-EKF within the sliding window filtering framework to identify the attacked measurements and remove them from the EKF update. The proposed algorithms have been extensively validated by Monte-Carlo simulations and experiments on a real dataset. Currently we extend the current work on 2D map-based localization to 3D simultaneous localization and mapping (SLAM). We will also investigate the signal spoofing for commonly-used sensors in SLAM, such as GPS, cameras, lidars and sonars.

Acknowledgements This work was partially supported by the University of Delaware College of Engineering, UD Cybersecurity Initiative, the Delaware NASA/EPSCoR Seed Grant, the NSF (IIS-1566129), and the DTRA (HDTRA1-16-1-0039).

References

1. Harris, M.: Researcher hacks self-driving car sensors. IEEE Spectrum (2015)
2. Charette, R.N.: Commercial drones and GPS spoofers a bad mix. IEEE Spectrum (2012)
3. Pasqualetti, F., Dörfler, F., Bullo, F.: Attack detection and identification in cyber-physical systems. IEEE Trans. Autom. Control. **58**(11), 2715–2729 (2013)
4. Fawzi, H., Tabuada, P., Diggavi, S.: Secure estimation and control for cyber-physical systems under adversarial attacks. IEEE Trans. Autom. Control. **59**, 1454–1467 (2014)
5. Mo, Y., Sinopoli, B.: Secure estimation in the presence of integrity attacks. IEEE Trans. Autom. Control. **60**(4), 1145–1151 (2015)
6. Pajic, M., Weimer, J., Bezzo, N., Tabuada, P., Sokolsky, O., Lee, I., Pappas, G.: Robustness of attack-resilient state estimators. In: Proceedings of the ACM/IEEE Conference on Cyber-Physical Systems, pp. 163–174 (2014)

7. Shoukry, Y., Puggelli, A., Nuzzo, P., Sangiovanni-Vincentelli, A.L., Seshia, S.A., Tabuada, P.: Sound and complete state estimation for linear dynamical systems under sensor attacks using satisfiability modulo theory solving. In: American Control Conference, pp. 3818–3823. IEEE (2015)
8. Mo, Y., Murray, R.M.: Multi-dimensional state estimation in adversarial environment. In: Proceedings of the Chinese Control Conference, Hangzhou, China, pp. 28–30 (2015)
9. Langner, R.: Stuxnet: dissecting a cyberwarfare weapon. IEEE Secur. Priv. **9**, 49–51 (2011)
10. Liu, Y., Ning, P., Reiter, M.K.: False data injection attacks against state estimation in electric power grids. ACM Trans. Inf. Syst. Secur. **14**, 1–33 (2011)
11. Rutkin, A.H.: Spoofers use fake GPS signals to knock a yacht off course, Aug 2013. http://www.udel.edu/003938
12. Pajic, M., Tabuada, P., Lee, I., Pappas, G.J.: Attack-resilient state estimation in the presence of noise. In: Conference on Decision and Control, pp. 5827–5832. IEEE (2015)
13. Pajic, M., Lee, I., Pappas, G.J.: Attack-resilient state estimation for noisy dynamical systems. IEEE Trans. Control. Netw. Syst. **4**(1), 82–92 (2017)
14. Chong, M.S., Wakaiki, M., Hespanha, J.P.: Observability of linear systems under adversarial attacks. In: American Control Conference, pp. 2439–2444. IEEE (2015)
15. Bezzo, N., Weimer, J., Pajic, M., Sokolsky, O., Pappas, G.J., Lee, I.: Attack resilient state estimation for autonomous robotic systems. In: Proceedings of IEEE Conference on Intelligent Robots and Systems, pp. 3692–3698. IEEE (2014)
16. Hu, Q., Chang, Y.H., Tomlin, C.J.: Secure estimation for unmanned aerial vehicles against adversarial cyber attacks (2016). arXiv:1606.04176
17. Candes, E.J., Tao, T.: Decoding by linear programming. IEEE Trans. Inf. Theory **51**(12), 4203–4215 (2005)
18. Shoukry, Y., Nuzzo, P., Bezzo, N., Sangiovanni-Vincentelli, A., Seshia, S.A., Tabuada, P.: Attack detection and state reconstruction in differentially flat systems under sensor attacks using satisfiability modulo theory solving. In: Conference on Decision and Control, Osaka, Japan, pp. 15–18 (2015)
19. Izanloo, R., Fakoorian, S.A., Yazdi, H.S., Simon, D.: Kalman filtering based on the maximum correntropy criterion in the presence of non-gaussian noise. In: Conference on Information Science and Systems (CISS), pp. 500–505 (2016)
20. Liu, X., Qu, H., Zhao, J., Chen, B.: Extended kalman filter under maximum correntropy criterion. In: International Joint Conference on Neural Networks, pp. 1733–1737 (2016)
21. Yang, Y., Huang, G.: Map-based localization under adversarial attacks," Tech. Rep. 2017-003, University of Delaware, Department of Mechanical Engineering, Oct 2017. Link: udel.edu/∼ghuang/papers/tr_secure.pdf
22. Kulikova, M.: Square-root algorithms for maximum correntropy estimation of linear discrete-time systems in presence of non-gaussian noise (2016). arXiv:1610.00257
23. Chang, Y.H., Hu, Q., Tomlin, C.J.: Secure estimation based kalman filter for cyber-physical systems against adversarial attacks. arXiv:1512.03853
24. Roumeliotis, S.I., Burdick, J.W.: Stochastic cloning: a generalized framework for processing relative state measurements. In: Proceedings of IEEE Conference on Robotics and Automation, Washington, DC, pp. 1788–1795, May 11–15 2002
25. Kim, S.J., Koh, K., Lustig, M., Boyd, S., Gorinevsky, D.: An interior-point method for large-scale l_1-regularized least squares. IEEE J. Sel. Top. Signal Process. **1**, 606–617 (2007)
26. Bar-Shalom, Y., Li, X.R., Kirubarajan, T.: Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software. Wiley (2004)
27. Guivant, J.E., Nebot, E.M.: Optimization of the simultaneous localization and map building algorithm for real time implementation. IEEE Trans. Robot. Autom. **17**, 242–257 (2001)
28. Dellaert, F.: Factor graphs and gtsam: a hands-on introduction. Technical report (2012)