Map-Based Localization Under Adversarial Attacks

Yulin Yang - yuyang@udel.edu
Guoquan Huang - ghuang@udel.edu

Department of Mechanical Engineering
University of Delaware, Delaware, USA

RPNG

Robot Perception and Navigation Group (RPNG)
Tech Reoprt - RPNG-2017-SECURE
Last Updated - Oct. 30, 2017

Contents

Abstract

Due to increasing proliferation of autonomous vehicles, securing robot navigation against malicious attacks becomes a matter of urgent societal interest, because attackers can fool these vehicles by manipulating their sensors, exposing us to unprecedented vulnerabilities and ever-increasing possibilities for malicious attacks. To address this issue, we analyze in-depth the Maximum Correntropy Criterion Extended Kalman Filter (MCC-EKF) and propose a weighted MCC-EKF (WMCC-EKF) algorithm by systematically, rather than in an ad-hoc way, inflating the noise covariance of the compromised measurements based on each measurement's quality. As a conservative alternative, we also design a secure estimator by first detecting attacks based on $\ell_0(\ell_1)$-optimization assuming that only a small number of measurements can be attacked, and then employ a sliding-window Kalman filter to update the state estimates and covariance using *only* the uncompromised measurements – the resulting algorithm is termed Secure Estimation-EKF (SE-EKF). Both Monte-Carlo simulations and experiments are performed to validate the proposed secure estimators for map-based localization.

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, 4, 5, 6, 7, 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 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 size 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. Based on that, 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 $\mathbf{x}_k \in \mathsf{R}^{m\times 1}$ 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 $\mathbf{Q}$. $\mathbf{y} \in \mathsf{R}^{p\times 1}$ denotes the measurements from p sensors, $\mathbf{h}$ represents the nonlinear measurement model function. $\mathbf{a} \in \mathsf{R}^{p\times 1}$ denotes the attack signals and is assumed to be sparse vector that at least one sensor cannot be attacked. We also define $\mathbf{z} \in \mathsf{R}^{p\times 1}$ as the un-attacked output. $\mathbf{n} \in \mathsf{R}^{p\times 1}$ 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 measurement 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 Appendix A) 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 (Fig. 1) as an example to illustrate the key ideas of our proposed secure estimators, the methodology is general and readily

applicable to other systesm. Specifically, in map-based localization, the dynamic motion model is given by:

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

where $\mathbf{v} = [v\cos(\phi) \quad v\sin(\phi)]^\top$ is the linear velocity and ω is the angular velocity of the robot. ϕ denotes the orientation of the robot. Note that we assume a more challenging localization scenario than [16, 13] that the robot cannot get GPS signals. Instead, only the relative range and bearing measurements of the features are available for localization, and the measurements can be described as:

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

where $\mathbf{h}^{(r)}$ and $\mathbf{h}^{(b)}$ represent the range and bearing measurements respectively, ${}^{s}\mathbf{p_f} = [{}^{s}x_\mathbf{f} \ {}^{s}y_\mathbf{f}]^\top$ represents the map feature in the sensor frame of reference.

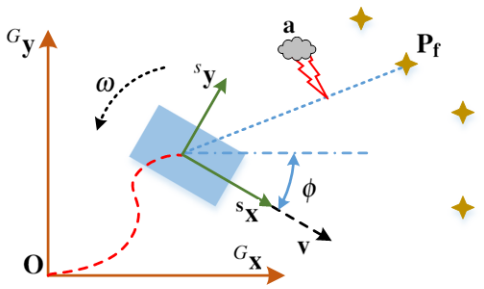

Figure 1: 2D map-based localization with adversarial attacks

It is important to note that, instead of assuming a fixed set of attacked sensors [4, 12], we assume that the attacker can attack different sensors randomly at different time steps [see (49)]. Note also that as compared to [16, 15], 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 $a_i (i = 1 \dots p)$ 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:

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

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, $\mathbf{P}_{k+1|k}$ are 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, 21], we analytically derive the MCC-EKF for the case of *nonlinear* systems such as map-based localization. In particular, given the prior knowledge of the 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:

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

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

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

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

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

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

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

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

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

where d_{k+1} is a ratio scalar computed from Gaussian kernel. Based on these derivations, the detailed MCC-EKF algorithm can be found as Algorithm 1.

Algorithm 1 MCC-EKF Algorithm

Prior Information $\hat{\mathbf{x}}_{0|0}$, $\mathbf{P}_{0|0}$
for $k \leftarrow 0$, $N-1$ **do**
  $\hat{\mathbf{x}}_{k+1|k} \leftarrow Eq.(10)$ {Propagation}
  $\mathbf{P}_{k+1|k} \leftarrow Eq.(11)$
  $\hat{\mathbf{z}}_{k+1|k} \leftarrow Eq.(12)$ {Update}
  $d_{k+1} \leftarrow Eq.(13)$
  $K_{k+1|k} \leftarrow Eq.(14)$
  $\hat{\mathbf{x}}_{k+1|k+1} \leftarrow Eq.(16)$
  $\mathbf{P}_{k+1|k+1} \leftarrow Eq.(17)$
end for

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:

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

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 strengthened [see(19)], which indicates

large uncertainties from the measurements. Thus, $\mathbf{S}_{k+1|k}$ increases accordingly and the updated state covariance $\mathbf{P}_{k+1|k+1}$ will be inflated due to (18). Lemma 3.1 summarizes our analysis:

Lemma 3.1. *For the MCC-EKF, if the attack $\mathbf{a}_{k+1}$ goes unbounded, the Iter will not perform measurement update and output the propagated estimates.*

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 $\hat{\mathbf{x}}_{k+1|k+1} \to \hat{\mathbf{x}}_{k+1|k}$ respectively. 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 cancel *all* the observation updates even if only one measurement is attacked at time step $k+1$, which clearly is too conservative and loses much useful information. Thus, in order to enable the MCC-EKF to still 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:

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

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]:

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

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

where $\hat{\sigma}_{i,k+1}, i = 1 \dots p$ 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 each 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:

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

where $\mathbf{H}_{i,k+1}, i = 1 \dots p$ represents each row of the Jacobian $\mathbf{H}_{k+1}$, $\hat{\mathbf{x}}_{k+1|k} = \mathbf{f}(\hat{\mathbf{x}}_{k|k}, \mathbf{0})$, and $\tilde{\mathbf{x}}_{k+1} = \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k}$. 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:

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

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

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

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

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

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

To this step, we have the new state update as (30), which is highly similar to (16). Now comes how to choose appropriate bandwidths. In order to design the bandwidths with physical meanings, we fix the ratio of $\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2}$ as σ_i^2, where σ_i denotes the standard deviation of the i-th measurement 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, we choose $\sigma_i^2=\lambda_\sigma\hat{\sigma}_i^2$, 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:

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

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

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

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

Algorithm 2 WMCC-EKF Algorithm

Prior Information $\hat{\mathbf{x}}_{0|0}$, $\mathbf{P}_{0|0}$
for $k\leftarrow 0, N-1$ **do**
  $\hat{\mathbf{x}}_{k+1|k}\leftarrow Eq.(10)$ {Propagation}
  $\mathbf{P}_{k+1|k}\leftarrow Eq.(11)$
  $\hat{\mathbf{z}}_{k+1|k}\leftarrow Eq.(12)$ {Update}
  for $i\leftarrow 1, p$ **do**
    $d_{i,k+1}\leftarrow Eq.(27)$ {Construct $\mathbf{D}_{k+1}$}
  end for
  $K_{k+1|k}\leftarrow Eq.(32)$
  $\hat{\mathbf{x}}_{k+1|k+1}\leftarrow Eq.(34)$
  $\mathbf{P}_{k+1|k+1}\leftarrow Eq.(35)$
end for

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:

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

where Σ_{w1} and Σ_{w2} denotes the information from motion model (1) and the measurement model (2), respectively. Note that $d_{i,k+1}\frac{\mathbf{H}_{i,k+1}^{\top}\mathbf{H}_{i,k+1}}{\sigma_{i,k+1}^{2}}$ 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 weakened. However, this process will not affect the information contribution from other sensors. Therefore, different with 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], we also perform the convergence analysis for the proposed WMCC-EKF 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 equation (34) can be rewritten as:

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1|k}(\mathbf{z}_{k+1} - \bar{\mathbf{z}}_{k+1} + \mathbf{h}(\bar{\mathbf{x}}_{k+1}) - \mathbf{h}(\hat{\mathbf{x}}_{k+1|k}) + \mathbf{a}_{k+1}) \tag{38}$$

$$= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1|k}(\mathbf{z}_{k+1} - \bar{\mathbf{z}}_{k+1}) + \mathbf{K}_{k+1|k}\mathbf{s}_{k+1} \tag{39}$$

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, in the following lemma, we in fact show that the WMCC-EKF can constrain the attack innovation to a small bound even under unbounded attacks.

Lemma 3.2. *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. See Appendix B. □

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, 22] 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 [23]. 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:

$$\mathbf{x}_{c_k} = \begin{bmatrix} \mathbf{x}_k^\top & \mathbf{x}_{k-1}^\top & \cdots & \mathbf{x}_{k-N+1}^\top & \mathbf{x}_{k-N}^\top \end{bmatrix}^\top \tag{41}$$

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 $\mathbf{x}_{c_k}$ in the state vector and accumulated their related measurements, we can linearize and stack all the measurements together as:

$$\begin{bmatrix} \tilde{\mathbf{z}}_k \\ \tilde{\mathbf{z}}_{k-1} \\ \vdots \\ \tilde{\mathbf{z}}_{k-N} \end{bmatrix} \simeq \begin{bmatrix} \mathbf{H}_k \\ \mathbf{H}_{k-1} \\ \vdots \\ \mathbf{H}_{k-N} \end{bmatrix} \tilde{\mathbf{x}}_{c_k} + \begin{bmatrix} \mathbf{n}_k \\ \mathbf{n}_{k-1} \\ \vdots \\ \mathbf{n}_{k-N} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_k \\ \mathbf{a}_{k-1} \\ \vdots \\ \mathbf{a}_{k-N} \end{bmatrix} \tag{42}$$

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

$$\tilde{\mathbf{x}}_k = \mathbf{F}_{k-1} \cdots \mathbf{F}_{k-N} \tilde{\mathbf{x}}_{k-N} = \mathbf{F}_{k-1,k-N} \tilde{\mathbf{x}}_{k-N} \tag{43}$$

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, (42) can be written as:

$$\underbrace{\begin{bmatrix} \tilde{\mathbf{z}}_k \\ \tilde{\mathbf{z}}_{k-1} \\ \vdots \\ \tilde{\mathbf{z}}_{k-N} \end{bmatrix}}_{\tilde{\mathbf{Z}}} \simeq \underbrace{\begin{bmatrix} \mathbf{H}_k \\ \mathbf{H}_{k-1} \\ \vdots \\ \mathbf{H}_{k-N} \end{bmatrix} \begin{bmatrix} \mathbf{F}_{k,k-N} \\ \mathbf{F}_{k-1,k-N} \\ \vdots \\ \mathbf{I} \end{bmatrix}}_{\mathbf{\Phi}} \tilde{\mathbf{x}}_{k-N} + \underbrace{\begin{bmatrix} \mathbf{n}_k \\ \mathbf{n}_{k-1} \\ \vdots \\ \mathbf{n}_{k-N} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_k \\ \mathbf{a}_{k-1} \\ \vdots \\ \mathbf{a}_{k-N} \end{bmatrix}}_{\mathbf{E}} \tag{44}$$

where $\tilde{\mathbf{Z}}$ represents the stacked measurement residuals, and $\mathbf{E}$ denotes the sum of stacked noise and attack vectors, $\mathbf{\Phi}$ denotes the stacked state transition matrix from $\tilde{\mathbf{x}}_{k-N}$ to each state in the window. Inspired by the attack detection techniques in [16, 22] we apply left null space operation to $\mathbf{\Phi}$ to simplify (44). Let $\mathbf{U}_n$ be the left null space of $\mathbf{\Phi}$, that is $\mathbf{U}_n^\top \mathbf{\Phi} = \mathbf{0}$, then we can have:

$$\mathbf{Z}_o = \mathbf{U}_n^\top \mathbf{Z} = \mathbf{U}_n^\top \mathbf{E} \tag{45}$$

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

$$\mathbf{\Phi} = \mathbf{U}_e \mathbf{R}_\Delta = \begin{bmatrix} \mathbf{U}_e & \mathbf{U}_n \end{bmatrix} \begin{bmatrix} \mathbf{R}_\Delta \\ \mathbf{0} \end{bmatrix} \tag{46}$$

Given the strong sparse attack assumption that less than a half of the all the sensors can be attacked, $\mathbf{E}$ can be solved from (45) by formulating the following optimization problem with ℓ_1 norm regularization[24] as:

$$\hat{\mathbf{E}} = \arg\min_{\mathbf{E}} \left[\left\| \mathbf{Z}_o - \mathbf{U}_n^\top \mathbf{E} \right\|_2^2 + \lambda \left\| \mathbf{E} \right\|_{\ell_1} \right] \tag{47}$$

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 (45) will not be perfectly sparse. In order to minimize this side effects, 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:

$$\|e_i\| = \|n_i + a_i\| \leq \|n_i\| + \|a_i\| \leq \|n_i\| \tag{48}$$

Based on the white Gaussian noise assumption [*i.e.*, $n_i \sim \mathcal{N}(\mathbf{0}, \sigma_i^2)$], 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 all the attacked measurements and perform the state update only with un-attacked measurements. The Algorithm 3 describes the main procedures of SE-EKF as:

Algorithm 3 SE-EKF Algorithm

Prior Information $\hat{\mathbf{x}}_{0|0}$, $\mathbf{P}_{0|0}$
for $i \leftarrow 0$, N **do**
  Stack and construct $\tilde{\mathbf{Z}}$, $\mathbf{E}$ and $\mathbf{\Phi}$ as in Eq. (44) {Until current window size i reaches N}
  if *i reaches N* **then**
    $\hat{\mathbf{E}} \leftarrow Eq.(47)$
    $\mathbf{a} \leftarrow$Enforce sparsity by threshold λ_t and determine the attacks
    Kalman filter update with un-attacked measurements
    $i \leftarrow 0$ and clear current window
  end if
end for

5 Nonlinear Observability Analysis With Attacks

According to the [25, 26, 27], the non-linear observability analysis for attacked map-based localization system can be concluded with the following lemma:

Lemma 5.1. *Given a map-based localization system of (7) and (8), if at least one bearing or range measurement is un-attacked, the system's observability are determined by the number of features observed, that is:*

- *If only one feature is observed, the system is unobservable with unobservable direction as* $\mathbf{U}$*;*
- *If more than one feature are observed, the system is fully observable.*

where $\mathbf{U} = \begin{bmatrix} \left[-\mathbf{J}\left({}^{G}\mathbf{p_f} - {}^{G}\mathbf{p_x}\right)\right]^\top & 1 \end{bmatrix}^\top$, $\mathbf{J} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Proof. See Appendix C □

6 Simulation Results

To validate the proposed secure estimators, we consider a map-based localization scenario where a wheeled robot moves in a circle trajectory with diameter of 5 meters. 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 have also considered 3 different attack modes (49), 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

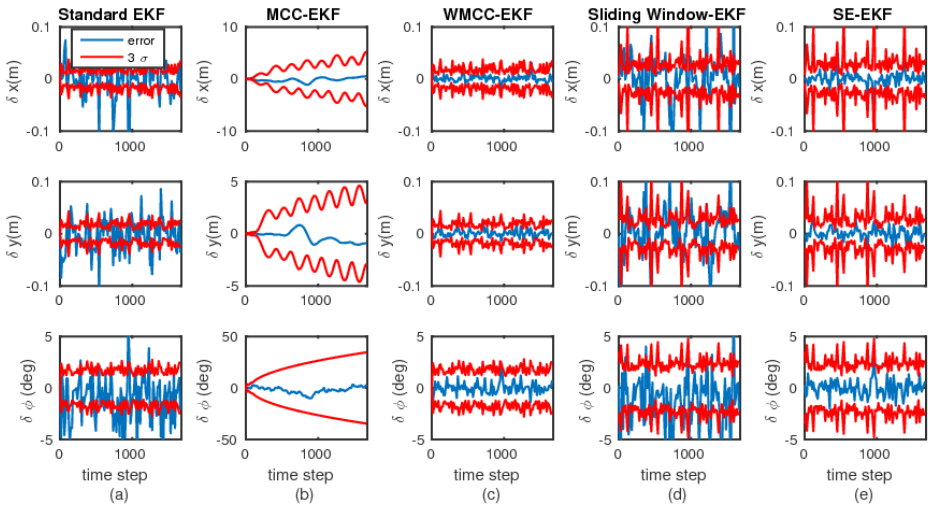

Figure 2: Comparison of the Standard EKF, MCC-EKF, WMCC-EKF, Sliding Window-EKF and SE-EKF under attacks.

attacked, there are i attacked sensors for Attack Mode i, and the set of attacked sensors are changing randomly over time.

$$\left.\begin{array}{l} Sensor\ 1: \ range \\ Sensor\ 2: \ bearing \\ Sensor\ 3: \ range \\ Sensor\ 4: \ bearing \end{array}\right\} \Leftarrow \underbrace{\begin{bmatrix} a_* & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & a_* & \cdots \\ \mathbf{0} & a_* & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & a_* & \mathbf{0} & \cdots \end{bmatrix}}_{\textbf{Attack Mode 1}}, \underbrace{\begin{bmatrix} a_* & a_* & \mathbf{0} & \mathbf{0} & \cdots \\ a_* & \mathbf{0} & \mathbf{0} & a_* & \cdots \\ \mathbf{0} & a_* & a_* & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & a_* & a_* & \cdots \end{bmatrix}}_{\textbf{Attack Mode 2}}, \underbrace{\begin{bmatrix} a_* & a_* & \mathbf{0} & a_* & \cdots \\ a_* & \mathbf{0} & a_* & a_* & \cdots \\ a_* & a_* & a_* & \mathbf{0} & \cdots \\ \mathbf{0} & a_* & a_* & a_* & \cdots \end{bmatrix}}_{\textbf{Attack Mode 3}} \tag{49}$$

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 rad for bearing measurement if not specified.

Fig. 2 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.

According to [16], the SE can have stable performance 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. Fig. 3 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) [28] 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:

$$d_m = \mathbf{r}^\top \mathbf{S}^{-1} \mathbf{r} \tag{50}$$

where $\mathbf{r}$ is the measurement residual and $\mathbf{S}$ is the corresponding innovation covariance. The d_m is

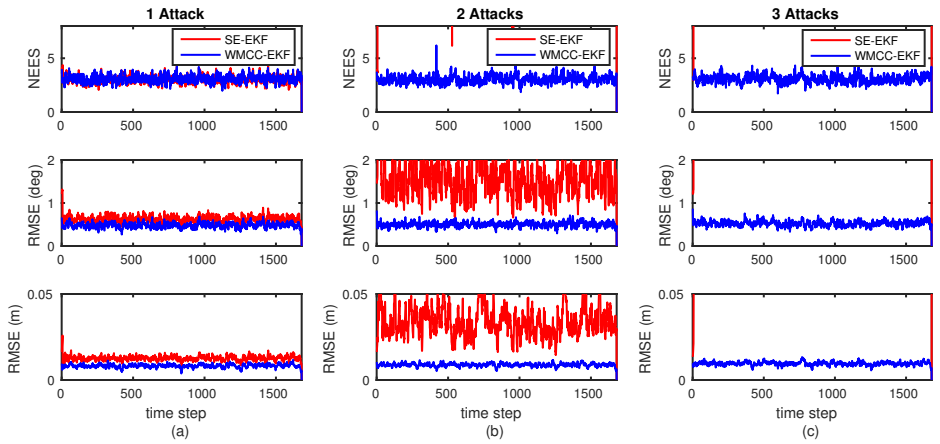

Figure 3: Full capacity test of the WMCC-EKF and SE-EKF in 50 Monte-Carlo simulations.

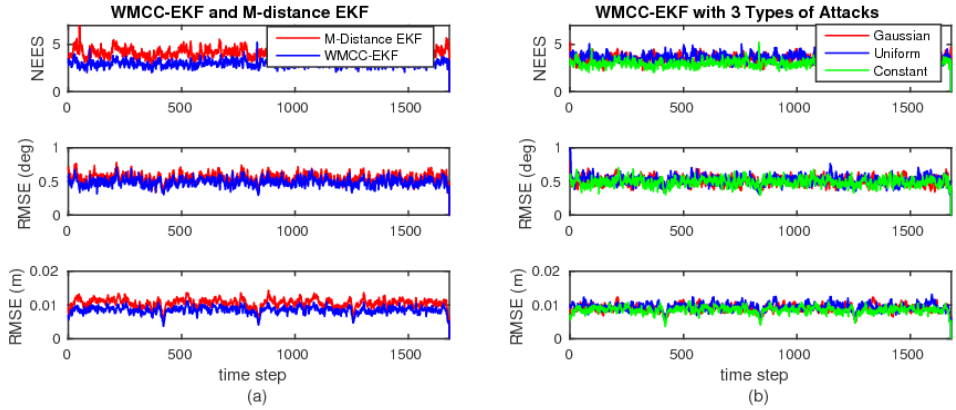

Figure 4: (a) Comparison of the Standard EKF and M-distance test based EKF under attacks; (b) Performance of WMCC-EKF with Gaussian, uniform and constant attacks.

assumed to follow the $\boldsymbol{\chi}^2$ distribution, thus we can define a threshold γ for d_m to identify outliers. We perform 50 Monte-Carlo runs (Fig. 4) 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.

7 Experimental Results

We further test the proposed WMCC-EKF and SE-EKF with a real dataset, the Victoria Park dataset [29], which includes wheel odometry measurements between robot poses and 2D range-bearing observations to landmarks (trees). Specifically, we first run a batch MAP optimization using GTSAM [30] 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 probability of 20% at each time step. Both range and bearing attack signals follows a uniform distribution, with magnitude c of 15m for range and $0.5rad$ for bearing, respectively. The results are shown in

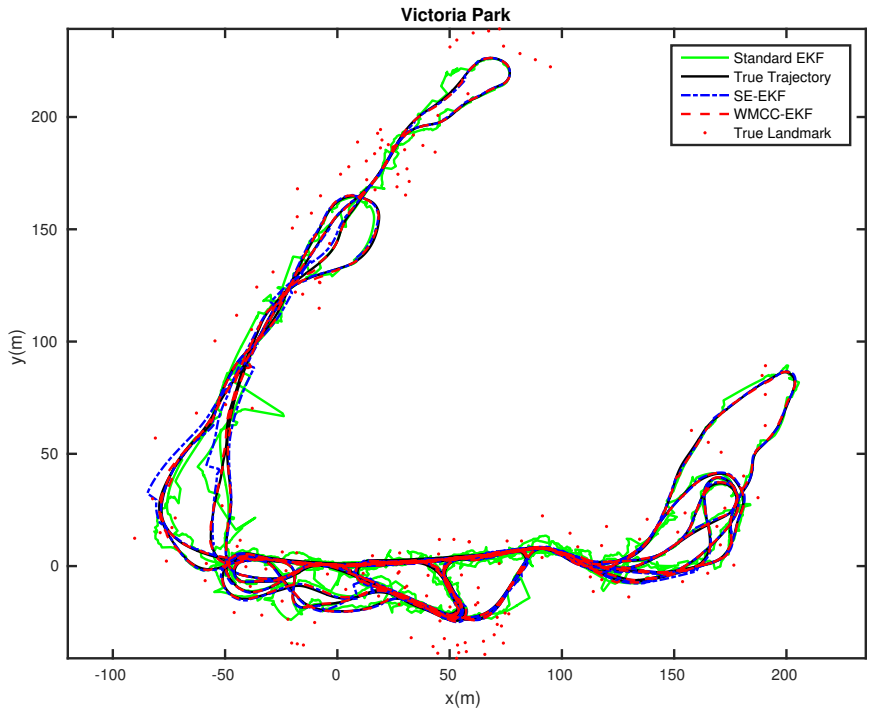

Figure 5: Estimated trajectories of the WMCC-EKF, SE-EKF and the Standard EKF with synthetic attacks on the Victoria Park dataset.

Figs. 5 and 6. It is clear from these plots that the green trajectory estimated by the Standard EKF is not acceptable, while the trajectories estimated by the proposed WMCC-EKF and SE-EKF are close to the true trajectory, which demonstrate that the proposed algorithms are able to secure the robot localization.

8 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 measurements. More conservatively, 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.

9 Acknowledgement

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

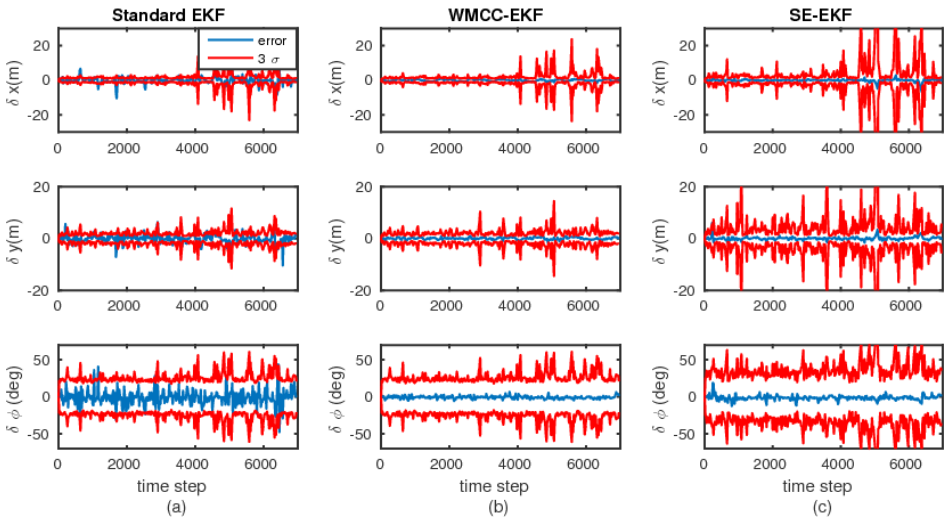

Figure 6: Estimation errors of the WMCC-EKF, SE-EKF and the Standard EKF with synthetic attacks on the Victoria Park dataset.

the DTRA (HDTRA1-16-1-0039).

Appendix A: Noise Pre-whitening

If the noise covariances matrix $\mathbf{R}$ are full matrix, we can perform pre-whitening. Since $\mathbf{R}$ is symmetrical, positive and definite (SPD) matrix, we can factorized $\mathbf{R}$ in the following way:

$$\mathbf{R} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top = (\mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}})(\mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}})^\top = \widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top \tag{51}$$

$$\Rightarrow \widehat{\mathbf{V}}^{-1}\mathbf{R}(\widehat{\mathbf{V}}^\top)^{-1} = \mathbf{I}_{\mathbf{\Lambda}} \tag{52}$$

where $\mathbf{\Lambda}$ is a diagonal matrix and $\widehat{\mathbf{V}} = \mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}}$. The pre-whitening is to apply $\widehat{\mathbf{V}}^{-1}$ with the linearized measurement equation as:

$$\underbrace{\widehat{\mathbf{V}}^{-1}\widetilde{\mathbf{z}}_{k+1}}_{\check{\mathbf{z}}_{k+1}} = \underbrace{\widehat{\mathbf{V}}^{-1}\mathbf{H}_{k+1}}_{\check{\mathbf{H}}_{k+1}}\widetilde{\mathbf{x}}_{k+1} + \underbrace{\widehat{\mathbf{V}}^{-1}\mathbf{n}_{k+1}}_{\check{\mathbf{n}}_{k+1}} \tag{53}$$

After pre-whitening, the new measurement noise becomes $\check{\mathbf{n}}_{k+1}$ s $\mathcal{N}(\mathbf{0}, \mathbf{I}_{\mathbf{\Lambda}})$, where $\mathbf{I}_{\mathbf{\Lambda}}$ is identity (diagonal) matrix.

Appendix B: Proof of Lemma 3.2

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

$$\left\|\mathbf{K}_{k+1|k}\mathbf{s}_{k+1}\right\|^2 = \left\|\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^\top\left(\mathbf{H}_{k+1}\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^\top + \mathbf{D}_{k+1}^{-1}\mathbf{R}_{k+1}\right)^{-1}\mathbf{s}_{k+1}\right\|^2 \tag{54}$$

$$\leq \left\|\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^\top\right\|^2\left\|\left(\mathbf{H}_{k+1}\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^\top + \mathbf{D}_{k+1}^{-1}\mathbf{R}_{k+1}\right)^{-1}\mathbf{s}_{k+1}\right\|^2 \tag{55}$$

$$= \left\|\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^\top\right\|^2\|\tau\|^2 \tag{56}$$

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 boundaries for $\|\tau\|$ as:

$$\begin{aligned}
\|\tau\|^2 &= \left\|\left(\mathbf{H}_{k+1}\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^{\top} + \mathbf{D}_{k+1}^{-1}\mathbf{R}_{k+1}\right)^{-1}\mathbf{s}_{k+1}\right\|^2 && (57)\\
&\leq \left\|\left(\sigma_{\min}^2\mathbf{I} + \mathbf{D}_{k+1}^{-1}\mathbf{R}_{k+1}\right)^{-1}\mathbf{s}_{k+1}\right\|^2 && (58)\\
&= \sum_{j=1}^{p}\left(\frac{s_j}{\sigma_{\min}^2 + d_j^{-1}\sigma_j^2}\right)^2 && (59)
\end{aligned}$$

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:

$$\mathbf{Pr}\left(\|\tilde{z}_{j,k+1}\| \leq 3\bar{\sigma}_{j,k+1}\right) = 99.7\% \tag{60}$$

Eq. (60) indicates that $\|\tilde{z}_{j,k+1}\|$ is bounded almost surely, and the bound $3\bar{\sigma}_{j,k+1}$ is accurate enough for engineering application. Then, we drop the time stamps for simplicity and arrive at:

$$\begin{aligned}
\left[\frac{\mathbf{s}_j}{\sigma_{\min}^2 + d_j^{-1}\sigma_j^2}\right]^2 &= \left[\frac{s_j}{\sigma_{\min}^2 + \exp(\frac{(y_j - \mathbf{h}(\hat{\mathbf{x}}))^2}{2\hat{\sigma}_j^2})\sigma_j^2}\right]^2 && (61)\\
&= \left[\frac{s_j}{\sigma_{\min}^2 + \exp(\frac{(z_j - \bar{z}_j + s_j)^2}{2\hat{\sigma}_j^2})\sigma_j^2}\right]^2 && (62)\\
&\leq \left[\frac{s_j}{\sigma_{\min}^2 + \exp\left(\frac{(\|s_j\| - \|\tilde{z}_j\|)^2}{2\hat{\sigma}_j^2}\right)\sigma_j^2}\right]^2 && (63)
\end{aligned}$$

If the j^{th} sensor attack a_j goes unbounded, according to (39), $\|\mathbf{s}_j\| \to \infty$ and hence $\|s_j\| > 3\bar{\sigma}_j$. Then, it is not difficult to see that $(\|s_j\| - \|\tilde{z}_j\|)^2 \geq (\|s_j\| - 3\bar{\sigma}_j)^2$, and we can arrive at:

$$\begin{aligned}
\left[\frac{s_j}{\sigma_{\min}^2 + \exp\left(\frac{(\|s_j\| - \|\tilde{z}_j\|)^2}{2\hat{\sigma}_j^2}\right)\sigma_j^2}\right]^2 &\leq \left[\frac{s_j}{\sigma_{\min}^2 + \exp\left(\frac{(\|s_j\| - 3\bar{\sigma}_j)^2}{2\hat{\sigma}_j^2}\right)\sigma_j^2}\right]^2 && (64)\\
&< \frac{\frac{\hat{\sigma}_j^2}{\sigma_j^4}\left(\frac{\|s_j\|}{\hat{\sigma}_j}\right)^2}{\left[\exp\left(\frac{1}{2}\left(\frac{\|s_j\|}{\hat{\sigma}_j} - 3\frac{\bar{\sigma}_j}{\hat{\sigma}_j}\right)^2\right)\right]^2} && (65)\\
&= \frac{\hat{\sigma}_j^2}{\sigma_j^4}\frac{\zeta^2}{\left[\exp\left(\frac{1}{2}(\zeta - \mu)^2\right)\right]^2} && (66)
\end{aligned}$$

where $\zeta = \frac{\|s_j\|}{\hat{\sigma}_j}$, and $\mu = 3\frac{\bar{\sigma}_j}{\hat{\sigma}_j}$. Obviously, as $\|s_j\| \to \infty$, $\zeta \to \infty$, and the right side of (64) will finally approach 0. Besides, if we take derivative of the right side of (64), we can have:

$$\frac{\partial}{\partial \zeta}\left(\frac{\hat{\sigma}_j^2}{\sigma_j^4}\frac{\zeta^2}{\left[\exp\left(\frac{1}{2}\left(\zeta-\mu\right)^2\right)\right]^2}\right) = \frac{\hat{\sigma}_j^2}{\sigma_j^4}\left(\frac{-2\zeta\left(\zeta^2-\mu\zeta-1\right)}{\left[\exp\left(\frac{1}{2}\left(\zeta-\mu\right)^2\right)\right]^2}\right) = 0 \tag{67}$$

The maximum value is when $\zeta' = \frac{\mu+\sqrt{\mu^2+8}}{2}$, thus:

$$\left[\frac{s_j}{\sigma_{\min}^2 + d_j^{-1}\sigma_j^2}\right]^2 \leq \frac{\hat{\sigma}_j^2}{\sigma_j^4}\frac{\zeta'^2}{\left[\exp\left(\frac{1}{2}\left(\zeta'-\mu\right)^2\right)\right]^2} \tag{68}$$

Since ζ' is independent of the attack innovation s_j, thus we can bound the (68) by appropriate design of bandwidth $\hat{\sigma}_j$. Therefore, according to (57) and (60), $\|\tau\|^2$ is the summation of (68) and is bounded by the design of $\mathbf{D}_{k+1}$ with probability 99.7%. In (56), $\left\|\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^\top\right\|^2$ is independent from the $\mathbf{a}_{k+1}$, and thus it is bounded. Therefore, with $\|\tau\|$ also being bounded, according to (56), we can easily find a ξ that satisfies (40).

Appendix C: Non-linear Observability Analysis Under Attacks

Before the observability analysis, we will first briefly go over the method in [25] and [27]. A non-linear continuous-time system can be written as:

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}_0(\mathbf{x}) + \sum_{i=1}^{l}\mathbf{f}_i(\mathbf{x})\mu_i \\ \mathbf{y} = \mathbf{h}(\mathbf{x}) \end{cases} \tag{69}$$

where $\mu_i, (i = 1 \ldots l)$ is the control input and $\mathbf{f}_i, (i = 1 \ldots l)$ are the process functions. Given the zeroth-order and the $i+1$-th order Lie derivative ([25]) of the measurement function $\mathbf{h}$, we have:

$$\mathcal{L}^0\mathbf{h} = \mathbf{h}(\mathbf{x}) \tag{70}$$

$$\mathcal{L}_{\mathbf{f}_j}^{i+1}\mathbf{h} = \nabla\mathcal{L}^i\mathbf{h}\cdot\mathbf{f}_j \tag{71}$$

with the span of the i-th order Lie derivative is defined as:

$$\nabla\mathcal{L}^i\mathbf{h} = \begin{bmatrix}\frac{\partial\mathcal{L}^i\mathbf{h}}{\partial x_1} & \frac{\partial\mathcal{L}^i\mathbf{h}}{\partial x_2} & \cdots & \frac{\partial\mathcal{L}^i\mathbf{h}}{\partial x_m}\end{bmatrix} \tag{72}$$

Thus, we can define the observability matrix $\mathcal{O}$ with block rows being the span of Lie derivatives of (69) based on [25]:

$$\mathcal{O} = \begin{bmatrix} \nabla\mathcal{L}^0\mathbf{h} \\ \nabla\mathcal{L}_{\mathbf{f}_i}^1\mathbf{h} \\ \nabla\mathcal{L}_{\mathbf{f}_i\mathbf{f}_j}^2\mathbf{h} \\ \nabla\mathcal{L}_{\mathbf{f}_i\mathbf{f}_j\mathbf{f}_k}^3\mathbf{h} \\ \vdots \end{bmatrix} \tag{73}$$

where $i, j, k = 0 \ldots l$. Based on [25], the system is observable if the matrix $\mathcal{O}$ is of full column rank. In the meanwhile, if we want to show the system is unobservable and determine the unobservable direction, we need to infinitely many Lie derivatives to construct a sub-matrix $\mathcal{O}'$, which is quite challenging, and the null space of $\mathcal{O}'$ will be the unobservable direction for the system. In order to address this issue, Guo et al. [27] proposed the following theorem to decompose the matrix $\mathcal{O}$ and simplify the problem.

Theorem 1. *Assume that there exists a non-linear transformation*

$$\boldsymbol{\beta}(\mathbf{x}) = \left[\boldsymbol{\beta}_1^\top(\mathbf{x}) \dots \boldsymbol{\beta}_t^\top(\mathbf{x})\right]^\top$$

of variable $\mathbf{x}$ *in* (69)*, such that:*

- $\mathbf{h}(\mathbf{x}) = \mathbf{h}'(\boldsymbol{\beta})$ *is a function of* $\boldsymbol{\beta}$*;*
- $\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} \cdot \mathbf{f}_i, i = 0, \dots l$ *are functions of* $\boldsymbol{\beta}$*;*
- $\boldsymbol{\beta}$ *is a function of the variables of a set* $\mathbf{S}$ *comprising Lie derivatives from oder zero up to in nite order.*

Then, the observability matrix $\mathcal{O}$ *can be factorized as:* $\mathcal{O} = \boldsymbol{\Xi} \cdot \boldsymbol{\Omega}$*, where* $\Omega \triangleq \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}}$ *and* $\boldsymbol{\Xi}$ *is the observability matrix of the following system* (74)*:*

$$\begin{cases} \dot{\boldsymbol{\beta}} = \mathbf{g}_0(\boldsymbol{\beta}) + \sum_{i=1}^{l} \mathbf{g}_i(\boldsymbol{\beta})\mu_i \\ \mathbf{y} = \mathbf{h}'(\boldsymbol{\beta}) \end{cases} \tag{74}$$

where $\mathbf{g}_i(\boldsymbol{\beta}) \triangleq \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} \mathbf{f}_i(\mathbf{x}), i = 1 \dots l$*. Therefore, the following statements are equivalent:*

- *System* (74) *is observable.*
- $null(\mathcal{O}) = null(\boldsymbol{\Omega})$*.*

Please refer to [27] for the complete proof of Theorem 1. In this paper we are utilizing this conclusion to analyze the non-linear observability for system under attacks.

C.1: Map-based Localization System Under Attacks

We first define the state vector for the Map-based localization problem (7) and (8) as:

$$\mathbf{x} = \begin{bmatrix} {}^G\mathbf{p}_x{}^\top & \phi \end{bmatrix}^\top \tag{75}$$

Thus, we can have the system dynamic model and equivalent measurement model under attacks as:

$$\dot{\mathbf{x}} = \underbrace{\begin{bmatrix} \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \end{bmatrix} \\ 0 \end{bmatrix}}_{\mathbf{f}_1} v + \underbrace{\begin{bmatrix} \mathbf{0}_{2\times 1} \\ 1 \end{bmatrix}}_{\mathbf{f}_2} \omega \tag{76}$$

$$\mathbf{h}' = \begin{bmatrix} \mathbf{h}_1' \\ \mathbf{h}_2' \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} + \mathbf{a} = \begin{bmatrix} \sqrt{{}^s\mathbf{p}_\mathbf{f}^\top {}^s\mathbf{p}_\mathbf{f}} \\ \frac{\mathbf{e}_2^\top {}^s\mathbf{p}_\mathbf{f}}{\mathbf{e}_1^\top {}^s\mathbf{p}_\mathbf{f}} \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{77}$$

where $\mathbf{h}_1'$ represents the attacked range measurement, $\mathbf{h}_2'$ represents the attacked bearing measurement. Based on sparse attack assumption, not all sensors are attacked. Hence, not all elements in attack vector $\mathbf{a} = \begin{bmatrix} a_1 & a_2 \end{bmatrix}^\top$ will be non-zeros. According to the Theorem 1 of [27], we will do the rank test in the following way:

$$\mathcal{O} = \boldsymbol{\Xi} \cdot \boldsymbol{\Omega} \tag{78}$$

where $\mathcal{O}$ is the observability rank matrix. It can be decomposed as $\mathbf{\Omega}$ and $\mathbf{\Xi}$. If $\mathbf{\Xi}$ is of full rank, then $\mathbf{\Omega}$ will have the same unobservable direction as the $\mathcal{O}$. After explaining how to construct the 74, we will elaborate the matrix rank test for $\mathbf{\Xi}$ in the next subsection. Based on our experiences, a good choice of $\boldsymbol{\beta}$ is the relative position measurement:

$$\begin{aligned} \boldsymbol{\beta}_1 &= [\mathbf{R}(\phi)]^\top ({}^G\mathbf{p_f} - {}^G\mathbf{p}_x) && (79) \\ \frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{x}} &= \begin{bmatrix} \frac{\partial \boldsymbol{\beta}_1}{\partial {}^G\mathbf{p}_x} & \frac{\partial \boldsymbol{\beta}_1}{\partial \phi} \end{bmatrix} = \begin{bmatrix} -[\mathbf{R}(\phi)]^\top & -\mathbf{J}[\mathbf{R}(\phi)]^\top ({}^G\mathbf{p_f} - {}^G\mathbf{p}_x) \end{bmatrix} && (80) \end{aligned}$$

Then, the new basis functions can be constructed accordingly as:

$$\begin{aligned} \mathbf{g}_1 = \frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{x}} \mathbf{f}_1 &= -\begin{bmatrix} 1 \\ 0 \end{bmatrix} && (81) \\ \mathbf{g}_2 = \frac{\partial \boldsymbol{\beta}_1}{\partial \mathbf{x}} \mathbf{f}_2 &= -\mathbf{J}[\mathbf{R}(\phi)]^\top ({}^G\mathbf{p_f} - {}^G\mathbf{p}_x) = -\mathbf{J}\boldsymbol{\beta}_1 && (82) \end{aligned}$$

Therefore, the new dynamic and measurement model can be expressed by the basis functions as:

$$\begin{aligned} \dot{\boldsymbol{\beta}}_1 &= -\begin{bmatrix} 1 \\ 0 \end{bmatrix} v - \mathbf{J}\boldsymbol{\beta}_1 \omega && (83) \\ \mathbf{h}' &= \begin{bmatrix} \mathbf{h}_1' \\ \mathbf{h}_2' \end{bmatrix} = \begin{bmatrix} \sqrt{\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1} + a_1 \\ \frac{\mathbf{e}_2^\top \boldsymbol{\beta}_1}{\mathbf{e}_1^\top \boldsymbol{\beta}_1} + a_2 \end{bmatrix} && (84) \end{aligned}$$

C.2: Rank Test Under Attacks

From the definition of Ξ, we can re-write Ξ as:

$$\Xi = \begin{bmatrix} \Xi_1 \\ \Xi_2 \end{bmatrix} \tag{85}$$

where Ξ_i is derived from $\mathbf{h}_i'$, $i = 1, 2$ respectively. As long as Ξ_1 or Ξ_2 is of full column rank, the matrix Ξ of (85) will be of full column rank. Now we will inspect the column rank for the Ξ_1 and Ξ_2 respectively. For Ξ_1, we have:

- The zeroth-order Lie derivative:

$$\mathcal{L}^0 \mathbf{h}_1 = \sqrt{\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1} + a_1 \tag{86}$$

$$\nabla \mathcal{L}^0 \mathbf{h}_1 = \frac{\boldsymbol{\beta}_1^\top}{\sqrt{\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1}} + \left(\frac{\partial a_1}{\partial \boldsymbol{\beta}_1} \right)^\top \tag{87}$$

- The first-order Lie derivative:

$$\mathcal{L}^1_{\mathbf{g}_1} \mathbf{h}_1 = -\mathbf{e}_1^\top \left(\frac{\boldsymbol{\beta}_1}{\sqrt{\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1}} + \frac{\partial a_1}{\partial \boldsymbol{\beta}_1} \right) \tag{88}$$

$$\nabla \mathcal{L}^1_{\mathbf{g}_1} \mathbf{h}_1 = -\mathbf{e}_1^\top \frac{(\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1 \mathbf{I}_2 - \boldsymbol{\beta}_1 \boldsymbol{\beta}_1^\top)}{(\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1)^{\frac{3}{2}}} - \mathbf{e}_1^\top \frac{\partial \left(\frac{\partial a_1}{\partial \boldsymbol{\beta}_1} \right)}{\partial \boldsymbol{\beta}_1} \tag{89}$$

- The Ξ_1 can be constructed as:

$$\Xi_1 = \begin{bmatrix} \nabla \mathcal{L}^0 \mathbf{h}_1 \\ \nabla \mathcal{L}^1_{\mathbf{g}_1} \mathbf{h}_1 \end{bmatrix} = - \begin{bmatrix} \frac{\boldsymbol{\beta}_1^\top}{\sqrt{\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1}} + \left(\frac{\partial a_1}{\partial \boldsymbol{\beta}_1} \right)^\top \\ -\mathbf{e}_1^\top \frac{(\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1 \mathbf{I}_2 - \boldsymbol{\beta}_1 \boldsymbol{\beta}_1^\top)}{(\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1)^{\frac{3}{2}}} - \mathbf{e}_1^\top \frac{\partial \left(\frac{\partial a_1}{\partial \boldsymbol{\beta}_1} \right)^\top}{\partial \boldsymbol{\beta}_1} \end{bmatrix} \tag{90}$$

Similarly, for Ξ_2, we can have:

- The zeroth-order Lie derivative:

$$\mathcal{L}^0 \mathbf{h}_2 = \frac{\mathbf{e}_2^\top \boldsymbol{\beta}_1}{\mathbf{e}_1^\top \boldsymbol{\beta}_1} + a_2 \tag{91}$$

$$\nabla \mathcal{L}^0 \mathbf{h}_2 = \frac{\boldsymbol{\beta}_1^\top \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}{\left(\boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\beta}_1 \right)^2} + \left(\frac{\partial a_2}{\partial \boldsymbol{\beta}_1} \right)^\top \tag{92}$$

- The first order Lie derivative :

$$\mathcal{L}^1_{\mathbf{g}_1} \mathbf{h}_2 = -\mathbf{e}_1^\top \left(\frac{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \boldsymbol{\beta}_1}{\left(\boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\beta}_1 \right)^2} + \frac{\partial a_2}{\partial \boldsymbol{\beta}_1} \right) \tag{93}$$

$$\nabla \mathcal{L}^1_{\mathbf{g}_1} \mathbf{h}_2 = -\mathbf{e}_1^\top \left(\frac{\left(\boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\beta}_1 \right) \mathbf{I}_2 - 4 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \boldsymbol{\beta}_1 \boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}{\left(\boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\beta}_1 \right)^3} \right) - \mathbf{e}_1^\top \frac{\partial \left(\frac{\partial a_2}{\partial \boldsymbol{\beta}_1} \right)^\top}{\partial \boldsymbol{\beta}_1} \tag{94}$$

- The Ξ_2 can be constructed as:

$$\Xi_2 = \begin{bmatrix} \nabla \mathcal{L}^0 \mathbf{h}_2 \\ \nabla \mathcal{L}^1_{\mathbf{g}_1} \mathbf{h}_2 \end{bmatrix} = \begin{bmatrix} \frac{\boldsymbol{\beta}_1^\top \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}{\left(\boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\beta}_1 \right)^2} + \left(\frac{\partial a_2}{\partial \boldsymbol{\beta}_1} \right)^\top \\ -\mathbf{e}_1^\top \left(\frac{\left(\boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\beta}_1 \right) \mathbf{I}_2 - 4 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \boldsymbol{\beta}_1 \boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}{\left(\boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\beta}_1 \right)^3} \right) - \mathbf{e}_1^\top \frac{\partial \left(\frac{\partial a_2}{\partial \boldsymbol{\beta}_1} \right)^\top}{\partial \boldsymbol{\beta}_1} \end{bmatrix} \tag{95}$$

Under the assumption of sparse attack, a_1 and a_2 will not be non-zero at the same time. Therefore, we can have the following analysis for the rank of $\boldsymbol{\Xi}$ of (85):

- If $a_1 = 0$, then the Ξ_1 will be:

$$\Xi_1 = \begin{bmatrix} \nabla \mathcal{L}^0 \mathbf{h}_1 \\ \nabla \mathcal{L}^1_{\mathbf{g}_1} \mathbf{h}_1 \end{bmatrix} = - \begin{bmatrix} \frac{\boldsymbol{\beta}_1^\top}{\sqrt{\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1}} \\ -\mathbf{e}_1^\top \frac{(\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1 \mathbf{I}_2 - \boldsymbol{\beta}_1 \boldsymbol{\beta}_1^\top)}{(\boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1)^{\frac{3}{2}}} \end{bmatrix} \tag{96}$$

It is not difficult to see that $\mathbf{\Xi}$ is of full column rank. Thus, from (85), we know $\mathbf{\Xi}$ is of full column rank;

- If $a_2 = 0$, then the Ξ_2 will be:

$$\Xi_2 = \begin{bmatrix} \nabla\mathcal{L}^0\mathbf{h}_2 \\ \nabla\mathcal{L}^1_{\mathbf{g}_1}\mathbf{h}_2 \end{bmatrix} = \begin{bmatrix} \dfrac{\boldsymbol{\beta}_1^\top \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}{\left(\boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\beta}_1\right)^2} \\ -\mathbf{e}_1^\top \left(\dfrac{\left(\boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\beta}_1\right) \mathbf{I}_2 - 4 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \boldsymbol{\beta}_1 \boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}{\left(\boldsymbol{\beta}_1^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\beta}_1\right)^3} \right) \end{bmatrix} \tag{97}$$

Similarly, we can find that $\mathbf{\Xi}$ is of full column rank. Thus, from (85), we know $\mathbf{\Xi}$ is also of full column rank.

Based on the above analysis, We can make the conclusion that, under sparse-attack assumption, no-matter whether the attack signal is related to the state $\mathbf{x}$ or not, the Ξ will be of full column rank. Therefore, according to the Theorem 1, the unobservable direction of $\mathcal{O}$ is the same as that of $\mathbf{\Omega}$.

C.3: Unobservable Direction

In the above section, we have proved that under sparse attack assumption the system (74)'s $\mathbf{\Xi}$ matrix will be full rank. Then, according to the Theorem 1, we only need to inspect the null space of $\mathbf{\Omega}$ and we have the following conclusions:

If only one feature ${}^G\mathbf{p_f}$ is observed from the map, the matrix Ω and its null space $\mathbf{U}$ can be expressed as:

$$\mathbf{\Omega} = \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} = \begin{bmatrix} -\left[\mathbf{R}(\phi)\right]^\top & -\mathbf{J}\left[\mathbf{R}(\phi)\right]^\top \left({}^G\mathbf{p_f} - {}^G\mathbf{p_x}\right) \end{bmatrix} \tag{98}$$

$$\mathbf{U} = \begin{bmatrix} -\mathbf{J}\left({}^G\mathbf{p_f} - {}^G\mathbf{p_x}\right) \\ 1 \end{bmatrix} \tag{99}$$

In this case we know that the map-based localization system (76) and (77) is unobservable, and the unobservable direction $\mathbf{U}$ is related to the rotation between the robot frame and the global frame.

If more than 1 features have been observed (e.g., ${}^G\mathbf{p_{f_1}}$ and ${}^G\mathbf{p_{f_2}}$), then the matrix Ω can be expressed as:

$$\mathbf{\Omega} = \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} = \begin{bmatrix} -\left[\mathbf{R}(\phi)\right]^\top & -\mathbf{J}\left[\mathbf{R}(\phi)\right]^\top \left({}^G\mathbf{p_{f_1}} - {}^G\mathbf{p_x}\right) \\ -\left[\mathbf{R}(\phi)\right]^\top & -\mathbf{J}\left[\mathbf{R}(\phi)\right]^\top \left({}^G\mathbf{p_{f_2}} - {}^G\mathbf{p_x}\right) \end{bmatrix} \tag{100}$$

In this case, the $\mathbf{\Omega}$ is of full column rank and thus, the map-based localization system is fully observable even under attacks.

References

[1] Mark Harris. "Researcher Hacks Self-driving Car Sensors". In: *IEEE Spectrum* (Sept. 2015). URL: http://spectrum.ieee.org/cars-that-think/transportation/self-driving/researcher-hacks-selfdriving-car-sensors.

[2] Robert N. Charette. "Commercial Drones and GPS Spoofers a Bad Mix". In: *IEEE Spectrum* (June 2012). URL: http://spectrum.ieee.org/riskfactor/aerospace/aviation/commercial-drones-and-gps-spoofers-a-bad-mix.

[3] Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo. "Attack detection and identification in cyber-physical systems". In: *IEEE Transactions on Automatic Control* 58.11 (2013), pp. 2715–2729.

[4] H. Fawzi, P. Tabuada, and S. Diggavi. "Secure Estimation and Control for Cyber-Physical Systems Under Adversarial Attacks". In: *IEEE Transactions on Automatic Control* 59.6 (June 2014), pp. 1454–1467.

[5] Yilin Mo and Bruno Sinopoli. "Secure estimation in the presence of integrity attacks". In: *IEEE Transactions on Automatic Control* 60.4 (2015), pp. 1145–1151.

[6] M. Pajic et al. "Robustness of attack-resilient state estimators". In: *Proc. of the ACM/IEEE Conf. on Cyber-Physical Systems*. 2014, pp. 163–174.

[7] Yasser Shoukry et al. "Sound and complete state estimation for linear dynamical systems under sensor attacks using satisfiability modulo theory solving". In: *American Control Conference*. IEEE. 2015, pp. 3818–3823.

[8] Yilin Mo and Richard M. Murray. "Multi-dimensional state estimation in adversarial environment". In: *Proc. of the Chinese Control Conference*. Hangzhou, China, 2015.

[9] R. Langner. "Stuxnet: Dissecting a Cyberwarfare Weapon". In: *IEEE Security Privacy* 9.3 (May 2011), pp. 49–51.

[10] Yao Liu, Peng Ning, and Michael K. Reiter. "False data injection attacks against state estimation in electric power grids". In: *ACM Transactions on Information and System Security* 14.1 (May 2011), pp. 1–33.

[11] Aviva Hope Rutkin. *Spoofers Use Fake GPS Signals to Knock a Yacht O Course*. http://www.udel.edu/003938. Aug. 2013.

[12] Miroslav Pajic et al. "Attack-resilient state estimation in the presence of noise". In: *Conference on Decision and Control*. IEEE. 2015, pp. 5827–5832.

[13] Miroslav Pajic, Insup Lee, and George J Pappas. "Attack-resilient state estimation for noisy dynamical systems". In: *IEEE Transactions on Control of Network Systems* 4.1 (2017), pp. 82–92.

[14] Michelle S Chong, Masashi Wakaiki, and Joao P Hespanha. "Observability of linear systems under adversarial attacks". In: *American Control Conference*. IEEE. 2015, pp. 2439–2444.

[15] Nicola Bezzo et al. "Attack resilient state estimation for autonomous robotic systems". In: *Proc. of IEEE Conf. on Intelligent Robots and Systems*. IEEE. 2014, pp. 3692–3698.

[16] Qie Hu, Young Hwan Chang, and Claire J Tomlin. "Secure Estimation for Unmanned Aerial Vehicles against Adversarial Cyber Attacks". In: *arXiv preprint arXiv:1606.04176* (2016).

[17] Emmanuel J Candes and Terence Tao. "Decoding by linear programming". In: *IEEE transactions on information theory* 51.12 (2005), pp. 4203–4215.

[18] Yasser Shoukry et al. “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, 2015.

[19] R. Izanloo et al. “Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise”. In: *Conference on Information Science and Systems (CISS)*. 2016, pp. 500–505. DOI: 10.1109/CISS.2016.7460553.

[20] X. Liu et al. “Extended Kalman filter under maximum correntropy criterion”. In: *Inter. Joint Conf. on Neural Networks*. 2016, pp. 1733–1737. DOI: 10.1109/IJCNN.2016.7727408.

[21] MV Kulikova. “Square-root algorithms for maximum correntropy estimation of linear discrete-time systems in presence of non-Gaussian noise”. In: *arXiv preprint arXiv:1610.00257* (2016).

[22] Young Hwan Chang, Qie Hu, and Claire J Tomlin. “Secure estimation based Kalman Filter for Cyber-Physical Systems against adversarial attacks”. In: *arXiv preprint arXiv:1512.03853* (2015).

[23] S. I. Roumeliotis and J. W. Burdick. “Stochastic Cloning: A generalized framework for processing relative state measurements”. In: *Proc. of IEEE Conf. on Robotics and Automation*. Washington, DC, 2002, pp. 1788–1795.

[24] S. J. Kim et al. “An Interior-Point Method for Large-Scale l_1-Regularized Least Squares”. In: *IEEE Journal of Selected Topics in Signal Processing* 1.4 (2007), pp. 606–617. ISSN: 1932-4553.

[25] R. Hermann and A. Krener. “Nonlinear controllability and observability”. In: *IEEE Transactions on Automatic Control* 22.5 (Oct. 1977), pp. 728 –740.

[26] Guoquan Huang, Anastasios I. Mourikis, and Stergios I. Roumeliotis. “Observability-based Rules for Designing Consistent EKF SLAM Estimators”. In: *International Journal of Robotics Research* 29.5 (Apr. 2010), pp. 502–528.

[27] Chao Guo and Stergios Roumeliotis. “IMU-RGBD Camera 3D Pose Estimation and Extrinsic Calibration: Observability Analysis and Consistency Improvement”. In: *Proc. of the IEEE International Conference on Robotics and Automation*. Karlsruhe, Germany, 2013.

[28] Yaakov Bar-Shalom, X Rong Li, and Thiagalingam Kirubarajan. *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[29] J. E. Guivant and E. M. Nebot. “Optimization of the Simultaneous Localization and Map Building Algorithm for Real Time Implementation”. In: *IEEE Transactions on Robotics and Automation* 17.3 (June 2001), pp. 242–257.

[30] Frank Dellaert. “Factor graphs and GTSAM: A hands-on introduction”. In: *Technical Report* (2012).