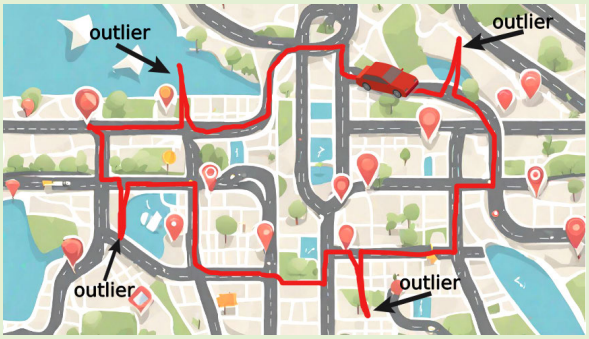
# Outlier-Insensitive Kalman Filtering: Theory and Applications

Shunit Truzman, Guy Revach, *Senior Member, IEEE*, Nir Shlezinger, *Senior Member, IEEE*, and Itzik Klein, *Senior Member, IEEE*

***Abstract*—State estimation of dynamical systems from noisy observations is a fundamental task in many applications. It is commonly addressed using the linear Kalman filter (KF), whose performance can significantly degrade in the presence of outliers in the observations, due to the sensitivity of its convex quadratic objective function. To mitigate such behavior, outlier detection algorithms can be applied. In this work, we propose a parameter-free algorithm which mitigates the harmful effect of outliers while requiring only a short iterative process of the standard KF's update step. To that end, we model each potential outlier as a normal process with unknown variance and apply online estimation through either expectation–maximization or alternating maximization algorithms. Simulations and field experiment evaluations demonstrate our method's competitive performance, showcasing its robustness to outliers in filtering scenarios compared with alternative algorithms.**




***Index Terms*— Alternating maximization, expectation–maximization, global navigation satellite systems (GNSSs), Kalman filter (KF), outlier detection.**

## I. Introduction

STATE estimation from noisy observation is a core task in various signal processing applications [2], such as localization and tracking [3], [4], [5]. This task is commonly addressed by the celebrated Kalman filter (KF) [6], a recursive and efficient algorithm providing an optimal low-complexity solution under the Gaussian noise and linear dynamics assumptions. However, KF's performance degrades significantly when observations are impaired by outliers, due to its least-squares cost function [7], [8], [9]. In real-world scenarios, measurements, especially from lower quality sensors such as global navigation satellite system (GNSS) devices, often contain outliers [5], [10], [11]. This presents a significant challenge to KF effectiveness. Therefore, an algorithm's ability to remain insensitive to outliers plays a crucial role in state estimation missions. Various techniques were proposed in the literature to cope with outliers: basic techniques, such as those in [12], [13], and [14], use statistical tests such as $\chi^2$-test to identify outliers based on prior information, and subsequently reject them. However, their robustness against outliers relies solely on the prediction step. The methods in [15], [16], and [17] suggest reweighting the observation noise covariance at each update step, but they often require extensive hyperparameter tuning. The approaches in [8], [18], [19], and [20] strive to reduce KF's sensitivity to outliers by replacing its quadratic cost function. Specifically, the works [18], [19] propose a Huber-based KF by minimizing the combined $\mathcal{L}_1$- and $\mathcal{L}_2$-norms. The nominal noise is bounded using a Huber function, but the feature of heavy tails inherent in non-Gaussian noises could limit the estimation accuracy. The techniques in [8] and [20] substitute the quadratic cost function with more suitable, often nonsmooth, convex functions, controlling outliers by promoting sparsity. However, these techniques involve smoothing algorithms rather than filtering and can be computationally complex. Methods such as [21] use heavy-tailed distributions, such as the Student's t-distribution, to model the observation noise. However, in the absence of outliers, a significant degradation is expected due to violated Gaussian assumption.

Received 25 August 2024; revised 1 October 2024; accepted 3 October 2024. Date of publication 17 October 2024; date of current version 27 November 2024. The work of Shunit Truzman was supported by the Maurice Hatter Foundation. An earlier version of this paper was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2023 [DOI:10.1109/ICASSP49357.2023.10095261]. The associate editor coordinating the review of this article and approving it for publication was Prof. Yulong Huang. *(Corresponding author: Shunit Truzman.)*

Shunit Truzman and Itzik Klein are with the Hatter Department of Marine Technologies, University of Haifa, Haifa 3498838, Israel (e-mail: shunitruzman@gmail.com; kitzik@univ.haifa.ac.il).

Guy Revach is with the Institute for Signal and Information Processing (ISI), D-ITET, 8092 ETH Zürich, Switzerland (e-mail: grevach@ethz.ch).

Nir Shlezinger is with the School of ECE, Ben-Gurion University of the Negev, Be'er Sheva 8410501, Israel (e-mail: nirshl@bgu.ac.il).

Digital Object Identifier 10.1109/JSEN.2024.3478328

The authors of [22] address this problem using a hierarchical distribution, specifically adopting a more robust distribution when the noise is skewed. Another popular technique, known as the maximum correntropy KF (MCKF) [23], [24], enhances KF's performances for state estimation in the presence of non-Gaussian noises, where correntropy is maximized. However, when the process model has uncertainties, the performance of MCKF degrades [25]. With recent advancements in neural networks (NNs), methods such as [14], [26] suggest detecting and correcting outlier observations using NNs before they enter into the KF stage. Nonetheless, these methods often require access to large amounts of data and pretraining. In [27], the use of normal with unknown variance (NUV) prior is introduced to devising an outlier-insensitive Kalman smoother (KS). Inspired by sparse Bayesian learning [28], [29], the authors propose to model each potential outlier as NUV [30], [31], [32] and estimate the unknown variance using expectation–maximization (EM) algorithm [33], [34], resulting in sparse outlier detection. Wadehn et al. [27] focus on the offline smoothing task and propose only the derivation of the EM to estimate the unknown variance, which requires the computation of second-order moments. In addition, this work focuses on a smoothing problem, commonly used for postprocessing, where all the data are available and one can run a forward and backward pass algorithm that allows to refine the state estimates simultaneously, which makes the task simpler compared with KF.

In the preliminary findings of this work, reported in the conference paper [1], we introduced the outlier-insensitive KF (OIKF), which is designed for the more commonly encountered task of online real-time filtering. In addition to presenting the EM algorithm, we also provide the alternating maximization (AM) algorithm [35] for estimating the unknown variance, which eliminates the need for computing the second-order moment of the state vector, makes its implementation simpler, and requires significantly less computation time. The main advantages of our approach, compared with other existing outlier-robust KF methods, are that it (i) is parameter-free; (ii) amounts to a short iterative process within KF's update step, i.e., we effectively stay within the linear Gaussian framework, and (iii) effectively leverages all the observation samples during the state estimation process of the KF. This article extends the preliminary findings reported in our conference paper [1] with the following additional contributions.

1) **Motivation**: A comprehensive motivation for the utilization of NUV in outlier detection within the KF framework, highlighting its benefits.
2) **Theory**: A comprehensive elucidation of the motivation behind the NUV prior representation to model outlier and tackle the problem of state estimation in the presence of outliers. To that end, we provide a complete mathematical derivation providing in-depth insights into the theoretical aspects of our outlier-insensitive Kalman filter (OIKF) with its two implementations OIKF-EM and OIKF-AM.
3) **Extensive Simulation Analysis**: A comparison of scenarios with low outlier intensity, suitable for any sensor updating the KF, which are inherently more challenging to detect and compensate for.
4) **GNSS Outlier Detection**: A real-world analysis focused on global navigation satellite system (GNSS) outlier detection using two datasets with three different platforms to highlight our approach's robustness. One dataset comprises Segway recordings [36], while the other comprises data from a quadrotor and a marine vessel [37].
5) **Open Source**: The source code and additional information on our empirical study can be found at https://github.com/KalmanNet/OIKF-NUV.git.

The rest of this article is organized as follows: Section II reviews the preliminaries for the outlier-robust state estimation task. Section III provides detailed explanations of NUV modeling and its utilization in the OIKF. Section IV presents the results of the empirical study, while Section V concludes this article with final remarks.

## II. Problem Formulation and Preliminaries

In this section, we introduce the preliminaries for the task of outlier-robust online state estimation, namely, the state-space (SS) model with outliers, and recapitulate the KF algorithm.

### A. SS Model With Outliers

We consider a scenario where noisy time-series observations, denoted as $\{\mathbf{y}_\tau\}_{\tau=1}^t$, are sequentially presented to a filter. The objective is to provide a sequence of estimates, $\{\hat{\mathbf{x}}_\tau\}_{\tau=1}^t$, corresponding to a sequence of hidden (latent) values or "states," $\{\mathbf{x}_\tau\}_{\tau=1}^t$ [38]. This scenario introduces an additional challenge: a subset of the observations may be impaired by outliers from an unknown distribution. We operate under the assumption that an anomalous observation should be considered a rare event to qualify as an outlier.

Unlike the offline state estimation task, also known as as smoothing, which is considered in [39], where all the observations are provided as a batch, we focus on real-time filtering here. In this approach, the estimate of $\mathbf{x}_t$ relies solely on current and past observations. This stands in contrast to the methodology in [27], where iterating on the entire batch of observations is used to enhance robustness to outliers and consequently improve state estimation performance.

In this work, we assume that the underlying relationship between the observed values and the hidden values is represented by an SS model [2]. We focus on an linear Gaussian SS model in discrete-time, $t \in \mathbb{Z}$, represented as follows:

$$\mathbf{x}_t = \mathbf{F} \cdot \mathbf{x}_{t-1} + \mathbf{e}_t, \quad \mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \quad \mathbf{x}_t \in \mathbb{R}^m \tag{1a}$$

$$\mathbf{y}_t = \mathbf{H} \cdot \mathbf{x}_t + \mathbf{z}_t + \mathbf{u}_t, \quad \mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \quad \mathbf{y}_t \in \mathbb{R}^n. \tag{1b}$$

Equation (1a) describes the time evolution of the state $\mathbf{x}_t$ from the previous state $\mathbf{x}_{t-1}$, governed by an system (evolution) matrix $\mathbf{F}$ and additive Gaussian noise $\mathbf{e}_t$. This noise, with a process covariance matrix $\mathbf{Q}$, represents potential modeling uncertainties. Equation (1b) portrays how observations $\mathbf{y}_t$ are generated from $\mathbf{x}_t$, the current state at time step $t$. This process involves a measurement (observation) matrix $\mathbf{H}$, additive Gaussian noise $\mathbf{z}_t$, with a measurement covariance matrix $\mathbf{R}$ accounting for uncertainties in the measurements, and potential outliers $\mathbf{u}_t$, which follows an unspecified distribution.

### B. Linear Kalman Filtering

The celebrated KF [6] is particularly noteworthy for its recursive and efficient algorithm, providing an optimal solution under Gaussian noise and linear dynamics [2], [40]. In its most general form, the KF aims to estimate the current state based on a noisy observation signal. However, KF's performance can degrade in the presence of outliers [8], [9], [20]. This sensitivity stems from the filter's objective to minimize a quadratic cost function, a structure that inherently is not able to follow fast jumps in the state dynamics [41]. For full details on how the Maximum *a posteriori* (MAP) formulation boils down to least-squares minimization, see [42], [43].

The KF estimates the state $\mathbf{x}_t$ from the observations $\{\mathbf{y}_\tau\}_{\tau\leq t}$ and can be thought of as a two-step process at each time step: *predict* and *update*. In the *predict* step, the joint probability distribution is computed using the first- and second-order moments of the Gaussian distribution, resulting in the prior distribution. The *predict* of the first- and second-order moments

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}\cdot\hat{\mathbf{x}}_{t-1}, \quad \boldsymbol{\Sigma}_{t|t-1} = \mathbf{F}\cdot\boldsymbol{\Sigma}_{t-1}\cdot\mathbf{F}^\top + \mathbf{Q} \tag{2a}$$

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{H}\cdot\hat{\mathbf{x}}_{t|t-1}, \quad \mathbf{S}_{t|t-1} = \mathbf{H}\cdot\boldsymbol{\Sigma}_{t|t-1}\cdot\mathbf{H}^\top + \mathbf{R}. \tag{2b}$$

where $\boldsymbol{\Sigma}$ represents the covariance of the state, $\mathbf{F}$ is the state-transition model, and $\mathbf{H}$ is the observation and model. The matrices $\mathbf{Q}$ and $\mathbf{R}$ are the covariance matrices of the process noise and observation noise, respectively. The KF uses this prior distribution in the *update* step in the *posterior* distribution calculation by computing the new observation $\mathbf{y}_t$ with the previously predicted *prior* $\hat{\mathbf{x}}_{t|t-1}$. And the *update* of the first- and second order statistical moment

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t|t-1} + \boldsymbol{\mathcal{K}}_t\cdot\Delta\mathbf{y}_t, \quad \boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\mathcal{K}}_t\cdot\mathbf{S}_{t|t-1}\cdot\boldsymbol{\mathcal{K}}_t^\top \tag{3}$$

$$\boldsymbol{\mathcal{K}}_t = \boldsymbol{\Sigma}_{t|t-1}\cdot\mathbf{H}^\top\cdot\mathbf{S}_{t|t-1}^{-1}, \quad \Delta\mathbf{y}_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1} \tag{4}$$

where $\boldsymbol{\mathcal{K}}_t$ is the Kalman gain matrix used to balance the contributions of both parts and produce the final posterior distribution.

## III. Outlier-Insensitive Kalman Filtering Using NUV Prior

This section introduces our OIKF algorithm. First, we present the particular property of the NUV prior that motivates us to model outliers as NUV and helps us tackle the problem of state estimation in the presence of outliers, as seen in KF. Then, we elaborate on our innovative approach of integrating the NUV prior into the KF algorithm for outlier detection, denoted as OIKF. Finally, we provide a comprehensive derivation of our two proposed algorithms to estimate the unknown variance of the NUV, namely, NUV-based EM and NUV-based AM.

### A. Motivation for NUV Prior Representation

The NUV formulation models a variable of interest as a normal distribution with unknown variance, given that the unknown variance has a prior distribution [30], [32]. The NUV representation method proves to be a robust approach with various applications, each encountering different problems, and the choice of a specific prior depends on its circumstances. For example, Ma et al. [44] propose different priors for computer imaging problems. In our specific problem, outlier-insensitive KF, we opt for a uniform prior, based on a previous work [31], which demonstrates scenarios similar to our SS model. This choice is motivated by computational convenience and the objective of resulting in sparse outlier detection, as explained in Section III-C and in Section III-E. Once the prior is set, our approach is in fact a parameter-free approach.

One well-known property of the NUV is its tendency to yield a nonconvex penalty [31], suitable for addressing sparse least-squares problems with outliers. This nonconvex penalty is motivated by the influence function [45] in residuals. This function assesses the effect of a residual's size on the loss by evaluating its derivative [20]. As the size of the residuals increases, the influence function gradually approaches zero, leading to a sparse solution.

To motivate, we use a simple example, based on the observation model (1b), which illustrates the fundamental property of NUV priors. Consider a single observation of the form

$$\mathrm{y} = \mathrm{v} + \mathrm{u}, \quad \mathrm{y}\in\mathbb{R}. \tag{5}$$

Here, v is an additive white Gaussian noise (AWGN) with variance $\mathrm{r}^2$. The variable of interest u is modeled as a zero-mean real scalar Gaussian random variable with an unknown variance $\gamma^2$ (NUV). The maximum likelihood estimation (MLE) of $\gamma^2$ from a single sample $\mathrm{y}\in\mathbb{R}$ can be computed as follows:

$$\hat{\gamma}^2 = \underset{\gamma^2\geq 0}{\arg\max}\, \mathcal{P}\left(\mathrm{y}\,\middle|\,\gamma^2\right)\mathcal{P}\left(\gamma^2\right) \tag{6}$$

when we assume a constant prior $\mathcal{P}\left(\gamma^2\right) = (2\pi)^{1/2}$ for computation convenience and $\mathcal{P}\left(\mathrm{y}\,|\gamma^2\right)$ is normally distributed, that is, $\mathcal{P}\left(\mathrm{y}\,|\gamma^2\right)\sim\mathcal{N}\left(0,\mathrm{r}^2+\gamma^2\right)$, then (6) can be rewritten as

$$\hat{\gamma}^2 = \underset{\gamma^2\geq 0}{\arg\max}\left\{\frac{1}{\sqrt{2\pi\left(\mathrm{r}^2+\gamma^2\right)}}\exp\left(\frac{-\mathrm{y}^2}{2(\mathrm{r}^2+\gamma^2)}\right)\right\}. \tag{7}$$

To simplify, the computation of (7) is written in terms of a logarithmic function

$$\hat{\gamma}^2 = \underset{\gamma^2\geq 0}{\arg\min}\left\{\ln\left(\mathrm{r}^2+\gamma^2\right) + \frac{\mathrm{y}^2}{\left(\mathrm{r}^2+\gamma^2\right)}\right\}. \tag{8}$$

To derive (8), we equate its derivative with regard to $\gamma^2$ to zero and we get the closed form of unknown variance $\gamma^2$ of Gaussian u

$$\hat{\gamma}^2 = \max\left\{\mathrm{y}^2-\mathrm{r}^2, 0\right\}. \tag{9}$$

In a subsequent step, assuming $\gamma^2$ is estimated as $\hat{\gamma}^2$ as in (9), the MAP estimate of u, denoted as $\hat{\mathrm{u}}$, is given by

$$\hat{\mathrm{u}} = \underset{\mathrm{u}}{\arg\max}\,\mathcal{P}\left(\mathrm{y}\,|\mathrm{u}\right)\cdot\mathcal{P}\left(\mathrm{u}\right) \tag{10}$$

when $\mathcal{P}\left(\mathrm{y}\,|\mathrm{u}\right)\sim\mathcal{N}\left(\mathrm{u},\mathrm{r}^2\right)$ and $\mathcal{P}\left(\mathrm{u}\right)\sim\mathcal{N}\left(0,\gamma^2\right)$, the derivation of $\hat{\mathrm{u}}$ can be accomplished simpler to $\gamma^2$ in (9), thus

$$\hat{\mathrm{u}} = \underset{\mathrm{u}}{\arg\max}\left\{\frac{1}{\sqrt{2\pi\mathrm{r}^2}}e^{\frac{-(\mathrm{y}-\mathrm{u})^2}{2\mathrm{r}^2}} + \frac{1}{\sqrt{2\pi\gamma^2}}e^{\frac{-\mathrm{u}^2}{2\gamma^2}}\right\}$$

$$= \arg\min_{\mathrm{u}} \left\{ \frac{(\mathrm{y}-\mathrm{u})^2}{2\mathrm{r}^2} + \frac{\mathrm{u}^2}{2\gamma^2} \right\}. \tag{11}$$

Maximizing the expression in (11) results in

$$\hat{\mathrm{u}} = \mathrm{y} \cdot \frac{\hat{\gamma}^2}{\hat{\gamma}^2 + \mathrm{r}^2} = \max\left\{ \frac{\mathrm{y}^2 - \mathrm{r}^2}{\mathrm{y}}, 0 \right\}. \tag{12}$$

Plugging the obtained result (9) into $\mathcal{P}(\mathrm{y})$

$$\mathcal{P}(\mathrm{y}) = \max_{\gamma^2 \geq 0} \mathcal{P}\left(\mathrm{y}\,|\gamma^2\right) \cdot \mathcal{P}\left(\gamma^2\right) \tag{13}$$

yields the equivalent cost function $\mathcal{L}(\mathrm{y}) = -\log \mathcal{P}(\mathrm{y})$, when $\mathcal{P}(\mathrm{y})$ can be derived from (8) when we also assume a constant prior for $\mathcal{P}\left(\gamma^2\right)$, and we obtain

$$\begin{aligned} \mathcal{L}(\mathrm{y}) &= -\log\left[ \frac{1}{\sqrt{\mathrm{r}^2+\gamma^2}} \exp\left( \frac{-y^2}{2\left(\mathrm{r}^2+\gamma^2\right)} \right) \right] \\ &= \frac{1}{2}\log\left(\mathrm{r}^2+\gamma^2\right) + \frac{\mathrm{y}^2}{2\left(\mathrm{r}^2+\gamma^2\right)}. \end{aligned} \tag{14}$$

Using the obtained expression for $\hat{\gamma}^2$ from (9), we obtain

$$\mathcal{L}(\mathrm{y}) = \begin{cases} \mathrm{y}^2/\left(2\mathrm{r}^2\right) + \log \mathrm{r}, & \mathrm{y}^2 < \mathrm{r}^2 \\ \log|\mathrm{y}| + 0.5, & \mathrm{y}^2 \geq \mathrm{r}^2. \end{cases} \tag{15}$$

If $\mathrm{r} > 0$, (15) results in a nonconvex function, which proves valuable for handling sparse least-squares models, such as KF, in the presence of outliers [31].

Through this example, we establish that $\gamma^2 = 0$ leads to $\mathrm{u} = 0$, indicating that no outlier is identified and the obtained $\gamma^2$ leads to a nonconvex cost function.

### B. Kalman Filtering With NUV Prior

The proposed OIKF introduces a new form of the SS model (1) by incorporating an additional variable into the observation signal, denoted as $\mathbf{u}_t$. This variable $\mathbf{u}_t$ represents the impulsive noise responsible for causing outliers. To improve the model's capability to handle heavy-tailed distributions in observations, we model this outlier $\mathbf{u}_t$ as normal with unknown variance $\boldsymbol{\gamma}_t^2$, namely,

$$\mathbf{u}_t \sim \mathcal{N}(0, \boldsymbol{\gamma}_t^2), \quad \boldsymbol{\gamma}_t^2 \sim \mathcal{P}\left(\boldsymbol{\gamma}_t^2\right) \tag{16}$$

where $\mathrm{p}(\boldsymbol{\gamma}_t^2)$ is the prior distribution of $\boldsymbol{\gamma}_t^2$. The assumption that NUV modeling follows a Gaussian distribution does not imply that the noise obeys a conventional independent and identically distributed (i.i.d) Gaussian model and that it is perfectly suitable for representing outliers. This approach allows us to remain within the linear Gaussian framework without making strict assumptions about the outlier distribution. Fig. 1 demonstrates visually the integration of the NUV representation into the overall model through a factor graph. The NUV representation approach results in a sparse outlier detection solution [30], [32], indicating that most values of $\boldsymbol{\gamma}_t^2$ will be zero. Consequently, as can be seen in (12), this leads to $\mathbf{u}_t = 0$, suggesting the absence of outliers, as expected.

For any given observation sample $\mathbf{y}_t$ (1b), we define $\mathbf{v}_t$ to be the error vector as the sum of two independent sources:

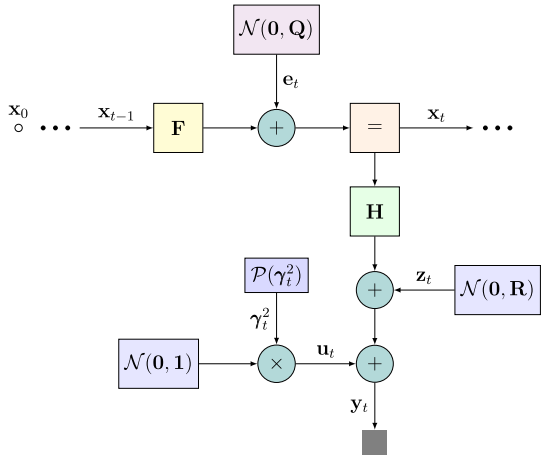



Fig. 1. Factor graph of the system model at time step $t$.

the observation noise $\mathbf{z}_t$ and the outlier noise $\mathbf{u}_t$. Thus, the covariance matrix of $\mathbf{v}_t$ is equal by definition

$$\mathbf{v}_t \triangleq \mathbf{y}_t - \mathbf{H} \cdot \mathbf{x}_t = \mathbf{z}_t + \mathbf{u}_t, \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \Gamma_t). \tag{17}$$

Here, $\Gamma_t$ is diagonal and comprises the sum of variances of the two noise sources, namely,

$$\Gamma_t = \mathrm{diag}\left(\boldsymbol{\nu}_t^2\right), \quad \boldsymbol{\nu}_t^2 \triangleq \mathbf{r}^2 + \boldsymbol{\gamma}_t^2. \tag{18}$$

In case the matrix $\Gamma_t$ is nondiagonal, the KF becomes nonlinear, requiring the use of an extended KF (EKF). In each KF iteration, the temporary estimate of $\boldsymbol{\gamma}_t^2$ is incorporated into the overall covariance $\Gamma_t$, effectively reweighting the covariance noise of the observations. Consequently, this affects the Kalman gain $\mathcal{K}_t$ in the *update* equations, allowing to extract the information from all the noisy observations and leverage their information effectively.

For the process of MAP estimation of the unknown variance, $\boldsymbol{\gamma}_t^2$, we can apply either EM (Section III-C) or AM (Section III-D) algorithms. While in the EM approach, the second-order moment $\boldsymbol{\nu}_t$ (18) is directly estimated, in the AM approach, it is obtained from estimating the first-order moment $\mathbf{v}_t$ (17) only, as summarized below

$$\hat{\boldsymbol{\gamma}}_t^2 = \max\left\{ \boldsymbol{\nu}_t^2 - \mathbf{r}^2, 0 \right\}, \quad \boldsymbol{\nu}_t^2 = \left\{ \mathrm{EM}: \hat{\boldsymbol{\nu}}_t^2, \mathrm{AM}: \hat{\mathbf{v}}_t^2 \right\}. \tag{19}$$

For both the approaches, an outlier is detected when $\hat{\boldsymbol{\gamma}}_t^2 \neq 0$; otherwise, when $\hat{\boldsymbol{\gamma}}_t^2 = 0$, it implies no outlier is present and we revert to the standard KF, preserving its optimally for data without outliers.

### C. Expectation–Maximization

For an observation $\mathbf{y}_t$, and the state vector $\mathbf{x}_t$ as defined in (1), the MAP estimation for the unknown variance $\hat{\boldsymbol{\gamma}}_t^2$ is

$$\hat{\boldsymbol{\gamma}}_t^2(\mathbf{y}_t) = \arg\max_{\boldsymbol{\gamma}_t^2 \geq 0} \mathcal{P}\left(\boldsymbol{\gamma}_t^2\,|\mathbf{y_t}\right) = \arg\max_{\boldsymbol{\gamma}_t^2 \geq 0} \mathcal{P}\left(\mathbf{y}_t \left|\boldsymbol{\gamma}_t^2\right.\right) \cdot \mathcal{P}\left(\boldsymbol{\gamma}_t^2\right) \tag{20}$$

$$= \arg\max_{\boldsymbol{\gamma}_t^2 \geq 0} \int \mathcal{P}\left(\mathbf{y}_t, \mathbf{x}_t \left|\boldsymbol{\gamma}_t^2\right.\right) dx \cdot \mathcal{P}\left(\boldsymbol{\gamma}_t^2\right). \tag{21}$$

To solve the optimization problem in (21), we devise the iterative EM algorithm, which consists of two iterating steps, namely, *E-step* and *M-step*.
The *E-step* determines the conditional expectation

$$\begin{aligned}&\mathbb{E}_{\mathbf{x}_t\left|\left(\boldsymbol{\gamma}_t^{i-1}\right)^2,\mathbf{y}_t\right.}\left[\log\left(\mathcal{P}\left(\mathbf{y}_t,\mathbf{x}_t\left|\boldsymbol{\gamma}_t^2\right.\right)\cdot\mathcal{P}\left(\boldsymbol{\gamma}_t^2\right)\right)\right]\\&=\mathbb{E}_{\mathbf{x}_t\left|\left(\boldsymbol{\gamma}_t^{i-1}\right)^2,\mathbf{y}_t\right.}\left[\log\mathcal{P}\left(\mathbf{y}_t\left|\mathbf{x}_t,\boldsymbol{\gamma}_t^2\right.\right)+\mathcal{P}\left(\mathbf{x}_t\left|\boldsymbol{\gamma}_t^2\right.\right)+\log\mathcal{P}\left(\boldsymbol{\gamma}_t^2\right)\right]\end{aligned}\tag{22}$$

Using the *Markov* property and the structure of the SS model, the term $\mathcal{P}\left(\mathbf{x}_t\left|\boldsymbol{\gamma}_t^2\right.\right)$ can be rewritten as $\mathcal{P}\left(\mathbf{x}_t\left|\mathbf{x}_{t-1}\right.\right)$.

The *M-step* goal is to maximize (22) with respect to $\boldsymbol{\gamma}_t^2$. In this problem, we assume a uniform prior on the unknown variance [30]

$$\mathcal{P}\left(\boldsymbol{\gamma}_t^2\right)\propto 1.\tag{23}$$

The choice of $\mathcal{P}\left(\boldsymbol{\gamma}_t^2\right)$ to be uniform is one of many options. As stated in [31], a uniform prior on $\mathcal{P}\left(\boldsymbol{\gamma}_t^2\right)$, also known as *plain NUV*, eventually leads to a nonconvex cost function which results in the sparse effect of the unknown variance, with most of them being zeros, as expected.

Since $\mathcal{P}\left(\boldsymbol{\gamma}_t^2\right)$ and the evolution $\mathcal{P}\left(\mathbf{x}_t\left|\mathbf{x}_{t-1}\right.\right)$ do not depend on $\boldsymbol{\gamma}_t^2$, they can thus be omitted from the optimization process (22) and we can evaluate the standard EM and compute the conditional distribution in (22). Thus, the $i$th iteration step is derived by

$$\begin{aligned}&\mathcal{Q}\left[\left(\boldsymbol{\gamma}_t^i\right)^2\right]\\&=\mathbb{E}_{\mathbf{x}_t\left|\left(\boldsymbol{\gamma}_t^{i-1}\right)^2,\mathbf{y}_t\right.}\left[\log\mathcal{P}\left(\mathbf{y}_t\left|\mathbf{x}_t,\boldsymbol{\gamma}_t^2\right.\right)\right]\\&=\mathbb{E}_{\mathbf{x}_t\left|\left(\boldsymbol{\gamma}_t^{i-1}\right)^2,\mathbf{y}_t\right.}\left[\log\left(\frac{1}{\sqrt{\Gamma_t}}\cdot\exp\left(\frac{-\left(\mathbf{y}_t-\mathbf{H}\cdot\mathbf{x}_t\right)^2}{2\cdot\Gamma_t}\right)\right)\right]\\&\propto\log\Gamma_t+\frac{1}{\Gamma_t}\cdot\mathbb{E}_{\mathbf{x}_t\left|\left(\boldsymbol{\gamma}_t^{i-1}\right)^2,\mathbf{y}_t\right.}\left[\left(\mathbf{y}_t-\mathbf{H}\cdot\mathbf{x}_t\right)^2\right].\end{aligned}\tag{24}$$

To expand the term $\mathbb{E}_{\mathbf{x}_t\left|\left(\boldsymbol{\gamma}_t^{i-1}\right)^2,\mathbf{y}_t\right.}\left[\left(\mathbf{y}_t-\mathbf{H}\cdot\mathbf{x}_t\right)^2\right]$, we use the KF, allowing to obtain the first- and second-order *posterior* moments of $\mathbf{x}_t$, which equal by definition

$$\mathbb{E}_{\mathbf{x}_t\left|\left(\boldsymbol{\gamma}_t^{i-1}\right)^2,\mathbf{y}_t\right.}\left(\mathbf{x}_t\right)\triangleq\hat{\mathbf{x}}_t^i,\tag{25a}$$

$$\mathbb{E}_{\mathbf{x}_t\left|\left(\boldsymbol{\gamma}_t^{i-1}\right)^2,\mathbf{y}_t\right.}\left(\mathbf{x}_t\cdot\mathbf{x}_t^\top\right)=\boldsymbol{\Sigma}_t^i+\mathbf{X}_t^{\mathrm{II}}\triangleq\hat{\mathbf{X}}_t^i.\tag{25b}$$

The KF defines $\hat{\mathbf{x}}_t$ as the *posteriori* state estimate mean of $\mathbf{x}$ at time $t$, considering all observations up to and including time $t$ [6]. Therefore, given $\mathbf{y}_t$ and $\left(\boldsymbol{\gamma}_t^{i-1}\right)^2$, (25a) holds true. Equation (25b) represents the squared first moment of $\mathbf{x}_t$, as defined in (26), plus $\hat{\boldsymbol{\Sigma}}_t$, the *posteriori* estimated covariance matrix of $\mathbf{x}_t$, derived from the KF [6].

To simplify computations, the following expressions are equal by definition

$$\begin{aligned}&\mathbf{X}_t^{\mathrm{II}}\triangleq\hat{\mathbf{x}}_t^i\cdot\hat{\mathbf{x}}_t^{i\top},\quad\mathbf{Y}_t^{\mathrm{II}}\triangleq\mathbf{y}_t\cdot\mathbf{y}_t^\top,\\&\mathbf{XY}_t\triangleq\hat{\mathbf{x}}_t^i\cdot\mathbf{y}_t^\top,\quad\mathbf{YX}_t\triangleq\mathbf{y}_t\cdot\hat{\mathbf{x}}_t^{i\top}.\end{aligned}\tag{26}$$

Finally, the expectation step (24) is reduced to the following expression:

$$\mathcal{Q}\left[\left(\boldsymbol{\gamma}_t^i\right)^2\right]=\log\Gamma_t+\frac{1}{\Gamma_t}\cdot\mathbf{V}_t\tag{27}$$

where $\mathbf{V}_t$ is equal by definition

$$\mathbf{V}_t\triangleq\left\{\mathbf{Y}_t^{\mathrm{II}}-\mathbf{H}\cdot\mathbf{XY}_t-\mathbf{YX}_t\cdot\mathbf{H}^\top-\mathbf{H}\cdot\hat{\mathbf{X}}_t^i\cdot\mathbf{H}^\top\right\}.\tag{28}$$

In *M-Step*, we maximize (27) with respect to $\Gamma_t$, and thus, for the $i$th iteration

$$\hat{\Gamma}_t^i=\underset{\Gamma_t^2\geq 0}{\arg\max}\left\{\ln\Gamma_t+\frac{1}{\Gamma_t}\cdot\mathbf{V}_t\right\}=\mathbf{V}_t.\tag{29}$$

We further exploit the fact that $\Gamma_t$ in (18) is diagonal, to expand $\Gamma_t$ to its components and estimate the variance $\left(\hat{\nu}_{t,k}^i\right)^2$ for each dimension $k\in\{1,\ldots,n\}$ in a scalar manner using (28)

$$\left(\hat{\nu}_{t,k}^i\right)^2=\mathbf{y}_{t,k}^2-2\cdot\mathbf{y}_{t,k}\cdot\mathbf{H}\cdot\hat{\mathbf{x}}_{t,k}^i+\mathbf{H}\cdot\left(\hat{\mathbf{x}}_{t,k}^i\right)^{\mathrm{II}}\cdot\mathbf{H}^\top\tag{30}$$

when $\hat{\mathrm{x}}_t$ is the *posteriori* state estimate. From (18), (30), and the fact that variance must be positive, we can calculate $\gamma_{t,k}^2$ in the $i$th iteration

$$\left(\hat{\gamma}_{t,k}^i\right)^2=\max\left\{\left(\hat{\nu}_{t,k}^i\right)^2-\mathrm{r}_k^2,0\right\}.\tag{31}$$

Thus, when an outlier is detected at a time step $t$, $\gamma_{t,k}^2>0$, otherwise its 0, which may lead to a sparse solution. As a consequence, the outlier will be estimated as $\hat{\mathrm{u}}_{t,k}=0$.

The above procedure is repeated iteratively for a fixed $K$ iterations, or until convergence is achieved. Algorithm 1 provides the suggested *pseudocode* for the OIKF-based NUV-EM.

**Algorithm 1** OIKF-Based NUV-EM for Time Instance $t$

```
Number of iterations K
Predict: Estimate a priori for x̂^{i=0}_{t|t-1}, Σ^{i=0}_{t|t-1} via (2a)
for i = 0, ..., K−1 do
    EM: Estimate (γ̂^i_t)^2 via (31) with the 2nd-order
        moment X̂^i_t as in (25b)
    Compute Γ^i_t = diag(r^2 + (γ̂^i_t)^2)
    Compute ŷ^i_{t|t-1}, S^i_{t|t-1} via (2b) with R = Γ^i_t.
    Update: Estimate a posteriori for x̂^i_t, Σ^i_t via (3)
end for
```

### D. Alternating Maximization

As in the EM approach, our goal is to estimate $\boldsymbol{\gamma}_t^2$ but here using AM instead. To that end, we use the iterative AM algorithm based on the plain smoothed NUV [31] to compute the joint MAP estimate for $\boldsymbol{\gamma}_t^2$, when the variable of interest

is $\mathbf{v}_t$. Consider the use of an NUV prior on variable $\mathbf{v}_t$ in an SS model with observation $\mathbf{y}_t$, we aim to determine their joint MAP estimate

$$[\hat{\mathbf{v}}_t, \hat{\boldsymbol{\gamma}}_t^2](\mathbf{y}_t) = \underset{\mathbf{v}_t, \boldsymbol{\gamma}_t^2 \geq 0}{\arg\max}\, \mathcal{P}\left(\mathbf{y}_t, \mathbf{v}_t, \boldsymbol{\gamma}_t^2\right) = \underset{\mathbf{v}_t, \boldsymbol{\gamma}_t^2 \geq 0}{\arg\max}\, \mathcal{P}\left(\mathbf{y}_t \,|\mathbf{v}_t\right) \cdot \mathcal{P}\left(\mathbf{v}_t | \boldsymbol{\gamma}_t^2\right) \cdot \mathcal{P}\left(\boldsymbol{\gamma}_t^2\right). \quad (32)$$

The latter is valid because as for certain continuous random variables, the joint probability density function is defined as the derivative of the joint cumulative distribution function.

To compute (32), we derive the AM algorithm, which iterates between a maximization step over the error state $\mathbf{v}_t$ with a fixed variance $\boldsymbol{\gamma}_t^2$

$$\hat{\mathbf{v}}_t = \underset{\mathbf{v}_t}{\arg\max}\, \mathcal{P}\left(\mathbf{y}_t \,|\mathbf{v}_t\right) \cdot \mathcal{P}\left(\mathbf{v}_t \,|\boldsymbol{\gamma}_t^2\right). \quad (33)$$

In particular, we replace $\mathbf{v}_t$ in (32) with its instantaneous estimate $\hat{\mathbf{v}}_t^i = \mathbf{y}_t - \mathbf{H} \cdot \hat{\mathbf{x}}_t^i$, which can be extracted from the KF. The next step in the AM is maximization (32) over the unknown variance $\boldsymbol{\gamma}_t^2$ based on $\mathbf{v}_t$, resulting in

$$\hat{\boldsymbol{\gamma}}_t^2 = \underset{\boldsymbol{\gamma}_t^2 \geq 0}{\arg\max}\, \mathcal{P}\left(\mathbf{v}_t \,|\boldsymbol{\gamma}_t^2\right) \cdot \mathcal{P}\left(\boldsymbol{\gamma}^2\right). \quad (34)$$

Note that since $\mathcal{P}\left(\mathbf{y}_t \,|\mathbf{v}_t\right)$ does not depend on $\boldsymbol{\gamma}_t^2$, and as we assume a uniform prior for $\mathcal{P}\left(\boldsymbol{\gamma}_t^2\right)$ as in (23), these expressions are not relevant for the maximization process (34) and can be omitted.

For convenience, we formulate (34) in a scalar manner, which extends to multivariate observations. To do that, we assume that the observation noise $\mathbf{z}_t$ and the outlier $\mathbf{u}_t$ in each dimension $k$ are independent, allowing to treat their sum, $\mathbf{v}_{t,k}$, in dimension $k$ as a scalar, leading to the scalar rule

$$\hat{\gamma}_{t,k}^2 = \underset{\gamma_{t,k}^2 \geq 0}{\arg\max}\, \mathcal{P}\left(\mathrm{v_{t,k}} | \gamma_{t,k}^2\right). \quad (35)$$

This maximization rule for $\gamma^2$ is to the one in (6), when here the variable of interest is the Gaussian $\mathrm{v}_{t,k}$. Therefore, we can use the result obtained in (9), and find the closed-form expression for the unknown variance of the $i$th iteration and for the $k$th entry

$$\left(\hat{\gamma}_{t,k}^i\right)^2 = \max\left\{\left(\hat{\mathrm{v}}_{t,k}^i\right)^2 - \mathrm{r}_k^2, 0\right\}. \quad (36)$$

We obtain an analytic expression of $\hat{\gamma}_{t,k}^2$ for the *update* step, where we alternate between $\hat{\mathbf{v}}_t^i$ and $\hat{\gamma}_{t,k}^2$ until convergence. Equation (36) is parameter-free and solely relies on the *posterior* estimate of the state $\hat{\mathbf{x}}_t^i$. This is in contrast to EM, which incorporates estimates of both the first- and second-order moments of the states, represented as $\hat{\mathbf{x}}_t^i$ and $\hat{\mathbf{X}}_t^i$, (25a) and (25b), respectively. Similar to EM, this procedure is repeated iteratively. Algorithm 2 provides the suggested *pseudocode* for the OIKF-based NUV-AM.

**Algorithm 2** OIKF-Based NUV-AM for Time Instance $t$

```
Number of iterations K
Predict: Estimate a priori for x̂_{t|t-1}^{i=0}, Σ_{t|t-1}^{i=0} via (2a)
for i = 0, ..., K-1 do
    AM: Compute v̂_t^i = y_t − H · x̂_t^i
        Estimate (γ̂_t^i)^2 via (36)
    Compute Γ_t^i = diag(r^2 + (γ̂_t^i)^2)
    Compute ŷ_{t|t-1}^i, S_{t|t-1}^i via (2b) with R = Γ_t^i.
    Update: Estimate a posteriori for x̂_t^i, Σ_t^i via (3)
end for
```

### E. Discussion

The NUV modeling is particularly efficient due to its sparse features, suitable for handling KF in the presence of outliers, which can be achieved by selecting the prior distribution of $\boldsymbol{\gamma}^2$. To that end, we opt for a uniform prior for the computations' convenience, which effectively adjusts the overall loss function to accommodate very sparse outliers' detection. Different choices for this prior would lead to alternative loss functions, such as the convex Huber cost function [7]. Furthermore, the efficiency of NUV modeling comes from being parameter-free and not requiring hyperparameter tuning, and from a computational perspective, it involves a short iterative process in the KF update step, leveraging all the observation samples.

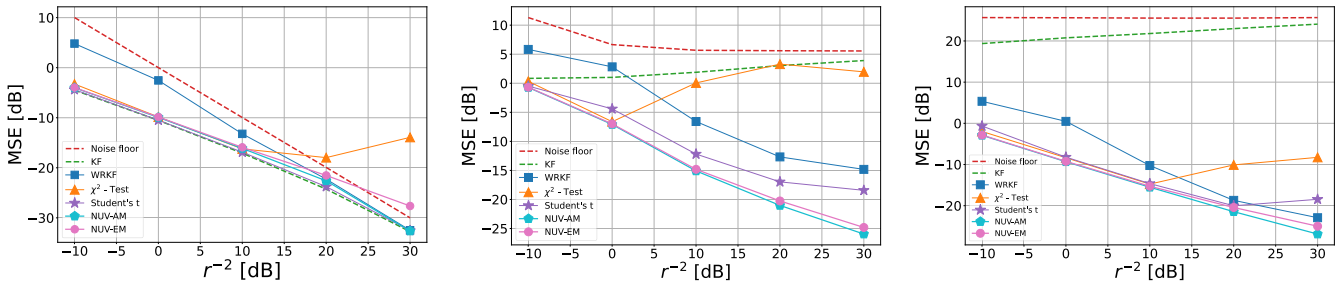
To integrate the NUV within the KF, we use either the EM or AM algorithms to estimate the unknown variance of the NUV. The estimated result combines to enhance the KF update step. The main limitation of the EM is that it requires the posterior variances of the state $\mathbf{x}_t$ in each iteration, which may be infeasible for large problems, while in AM version it is obtained from the first-order moment. However, the AM proves more effective compared to EM, as evident from the empirical evaluation in Section IV.

It is important to note that in certain applications, EM empirically gives better results than AM, primarily due to its accounting for the accuracy of the state estimate. Its simplicity relies on empirical second-order moments and holds potential for augmentation with trainable data-driven variations of the KF, for instance [38], [39]. Such fusion leverages robust filtering in partially known SS models and helps manage sensitivity to outliers.

## IV. Analysis and Results

In this section, we present a comprehensive assessment of the effectiveness of our proposed approaches: OIKF-based NUV-EM and OIKF-based NUV-AM, for outlier detection within various KF setups. Their performance is evaluated across different outlier intensities and tasks while comparing their effectiveness to other established works in the literature.

(a) **Simulations**: Our first experimental study considers a standard localization task with generated data. The synthetic dataset is generated using the white noise acceleration (WNA) [40] model, and the observation signal is subject to varying degrees of outlier corruption. Such models are commonly used in several applications such as navigation and target tracking.




(a) (b) (c)

Fig. 2. (a)–(c) MSE of the estimated position for the tracking application in the KF setup. Our NUV methods were compared to different well-established robust KF algorithms in the literature. (a) Noisy data clean of outliers. (b) Noisy data with outliers p = 0.2 scl = 3. (c) Noisy data with outliers p = 0.2 scl = 30.

(b) **North Campus Long-Term (NCLT) Dataset**: In our second study, we examine localization use case based on real-world data—the Michigan NCLT [36] dataset. Here, we compare our methods with different algorithms for tracking real-world dynamic data of a moving Segway robot using GNSS noisy measurements.

(c) **Autonomous Platforms Inertial (API) Dataset**: The third and fourth studies involve another localization use case, based on the real-world data—the API [37], [46] dataset. Here, we demonstrate the performances of our algorithms in tracking a quadrotor and a marine vessel using GNSS noisy measurements.

### A. Simulations

In this study, we use a KF algorithm where the state vector is represented by

$$\mathbf{x} = \begin{bmatrix} \mathrm{p} & \mathrm{v} \end{bmatrix}^\top \tag{37}$$

where p and v denote the position and velocity states, respectively. For the experiments that involve generated data, we establish the dynamic white noise acceleration model, followed by a linear SS model. For the filtering process, we assume both position and velocity measurements are available, and thus, the observation matrix is the identity matrix and the observation noise covariance matrix is diagonal, i.e.,

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1. \end{pmatrix}, \quad \mathbf{R} = \mathrm{r}^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{38}$$

The process noise covariance matrix is

$$\mathbf{Q} = \mathrm{q}^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{39}$$

where $\mathrm{q}^2$ is the process noise variance, set to a constant value of $-10\,[\mathrm{dB}]$.

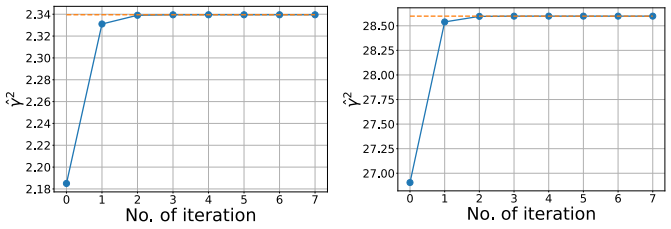
To evaluate our approach, we consider three scenarios for generating measurements: noisy data without outliers, noisy data with mild outliers, and noisy data with significant outliers. The presence of outliers within the measurement vector is modeled with intensities drawn from a Rayleigh distribution with parameter $\sigma_u^2$, where we use two scale parameters $\sigma_u^2 = [3, 30]$, representing low and high outlier intensities, respectively. The occurrence of outliers in the dataset is determined using a Bernoulli distribution, $\mathcal{B}(\mathrm{p})$, where we set the probability of an outlier to be $\mathrm{p} = 0.2$, indicating that roughly 20% of the data points are considered outliers. In Fig. 2, we compare our proposed OIKF based on NUV-EM and based on NUV-AM with the following algorithms: classical KF, reweighted algorithm (WRKF [15]), $\chi^2$-test [13], [14], and Student's t-distribution [21]. The results are presented in dB units to accentuate the distinctions between the obtained results. The unit conversion to dB is as follows:

$$\mathrm{MSE}\,[\mathrm{dB}] = 10 \cdot \log_{10}(\mathrm{MSE}). \tag{40}$$

This decision was motivated by the fact that in some cases, the differences can be subtle, making them challenging to note without the emphasis provided by dB units.

Fig. 2(a) presents the results of evaluating the performance of the NUV-AM algorithm on synthetic data that are clean of outliers, wherein the algorithm achieves the optimal minimal MSE bound by estimating a significant proportion of $\boldsymbol{\gamma}_t^2$ values as zero. Consequently, the model reverts to the KF, which is optimal for data without outliers. However, the Student's t outperforms our method because we optimize its hyperparameter such that it tends toward a normally distributed model (as the KF), which is also its drawback because it is not an adaptive algorithm. In contrast, $\chi^2$-test is deviating significantly from the performance of the KF. This is due to the necessity of determining its confidence level. In cases of low-noise observations, it may incorrectly identify normal samples as outliers, leading to their rejection and rendering the model reliant solely on the prediction step. The relative efficiencies of our OIKF and the KF can be assessed based on the resulting MSE. Notably, as depicted in Fig. 2(a), the average ratio of MSE values between the KF and OIKF-based NUV-EM yields a relative efficiency of 92%, while OIKF-based NUV-AM results in a relative efficiency of 96%. This outcome underscores the effectiveness of our algorithms even in the absence of outliers, due to the fact that a Gaussian framework is maintained.

When using synthetic data with outliers, as shown in Fig. 2(b) and (c) with outlier intensities of 3 and 30, respectively, the NUV-AM algorithm exhibits superior performance in terms of MSE when compared with other algorithms, across varying values of observation noise variance $r^2$, and for both high and low outlier intensities. Notably, it demonstrates similar performance to the NUV-EM algorithm and even surpasses




(a) (b)

Fig. 3. Convergence plots of the estimated outliers' variance computed using the NUV-AM algorithm. (a) Low outlier intensity. (b) High outlier intensity.

it for low observation noise without using a second-order moment as in EM.

In Fig. 3, we exhibit the convergence plots of the estimated variance $\hat{\gamma}^2$, using the NUV-AM algorithm when an outlier was identified. It is evident that the NUV-AM algorithm achieves rapid convergence after approximately three iterations, regardless of whether the outlier intensities are high or low.

### B. NCLT Dataset

For real-world data, we make use of the NCLT dataset [36]. The NCLT dataset is collected from a session with the date *2013-04-05*, which GNSS readings sampled at 5 Hz with a degree of noise and the corresponding ground location information of a Segway robot in motion. In the simulation setting, we process the measured vehicle position directions independently. To filter out these processes, we use the dynamic white noise acceleration model [40] for each direction separately. Since only the GNSS position is observable in this dataset, the measurement matrix is

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{41}$$

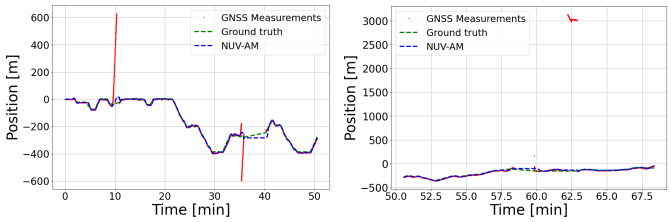
and the process and measurement noise covariances are

$$\mathbf{R} = \mathrm{r}^2, \quad \mathbf{Q} = \mathrm{q}^2 \begin{pmatrix} 1 & 0 \\ 0 & 1. \end{pmatrix} \tag{42}$$

Figs. 4 and 5 depict the trajectory of the Segway in the east and north directions, respectively, when the GNSS measurements are subjected to outliers, which result in readings deviating significantly from the ground truth (GT). We have divided each trajectory into two time intervals, with the first interval displaying outliers with lower intensity and the second interval with higher intensity.

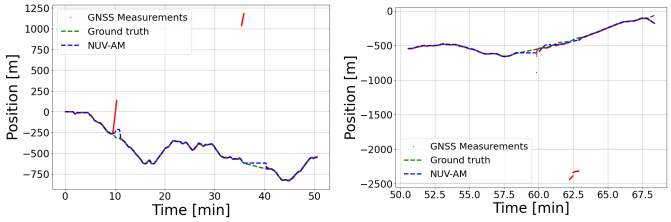
In addition, Fig. 6 presents the spatial trajectory of the moving Segway in both the directions. In the $x$-axis, we depict the trajectory from Fig. 4, while the $y$-axis depicts the trajectory from Fig. 5. Our analysis focuses on evaluating the effectiveness of NUV-AM in estimating position from real-world data and removing outliers reliably. In Table I, we compare our proposed OIKF based on NUV-EM and based on NUV-AM with the following algorithms: classical KF, reweighted algorithm (ORKF [16]), $\chi^2$-test [13], [14], and Student's t-distribution [21].

In addition, Table I presents the root mean-squared error (RMSE) and MSE for each algorithm, while the process



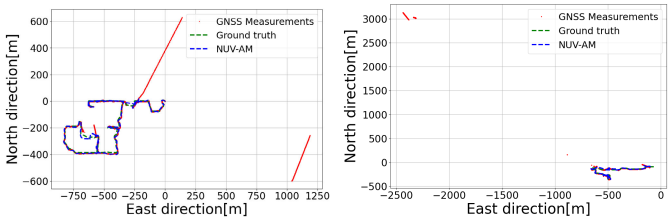

(a) (b)

Fig. 4. Measured vehicle position in *east direction* obtained from the noisy GNSS NCLT dataset (red points) is compared with the estimated trajectory by our NUV-AM (blue dashed line), which succeeded in passing the outliers and achieved performance comparable to the ground truth (green dashed line). (a) Time interval 0–50 min. (b) Time interval 50–68 min.




(a) (b)

Fig. 5. Measured vehicle position in *north direction* obtained from the noisy GNSS NCLT dataset (red points) is compared with the estimated trajectory by our NUV-AM (blue dashed line), which succeeded in passing the outliers and achieved performance comparable to the ground truth (green dashed line) (a) Time interval 0–50 min. (b) Time interval 50–68 min.




(a) (b)

Fig. 6. Measured vehicle position structure obtained from the noisy GNSS NCLT dataset (red points) is compared with the estimated trajectory by our NUV-AM (blue dashed line), which succeeded in passing the outliers and achieved performance comparable to the ground truth (green dashed line). (a) Time interval 0–50 min. (b) Time interval 50–68 min.

noise $\mathrm{q}^2$ and observation noise variance $\mathrm{r}^2$ of each algorithm are optimized separately through grid search to yield the lowest MSE. As shown in Table I, OIKF with both NUV-EM and NUV-AM has the lowest estimation errors in both the directions, with NUV-AM performing slightly better when it coincides with NUV-EM, even without using the second-order moment.

In terms of computation time, algorithms that combine KF with outlier detection techniques achieve higher accuracy but at the cost of longer computation runtimes, as outlier detection is applied. Compared with other algorithms, the $\chi^2$-test stands out as the shortest, stemming from the fact that it only detects and rejects outliers during the prediction step. On the other hand, a more sophisticated technique, such as the ORKF, requires more time due to its increased computational

TABLE I
POSITION ERROR FOR OPTIMAL VALUES OF $r^2$ AND $q^2$ FOR THE NCLT DATASET

| | North direction | | East direction | | Runtime |
|---|---|---|---|---|---|
| | RMSE[m] | MSE[dB] | RMSE[m] | MSE[dB] | [ms] |
| Noisy GNSS | 349.3 | 50.8 | 266.1 | 48.5 | - |
| KF | 92.3 | 39.3 | 164.4 | 44.3 | 0.05 |
| ORKF | 27.7 | 28.8 | 28 | 28.9 | 2.8 |
| $\chi^2$-test | 12.3 | 21.8 | 14.2 | 23 | 0.1 |
| Student's t | 11.5 | 21.2 | 13.8 | 22.8 | 0.5 |
| NUV-AM | **10.4** | **20.3** | **13** | **22.3** | **0.3** |
| NUV-EM | **10.3** | **20.3** | **13** | **22.3** | 0.4 |

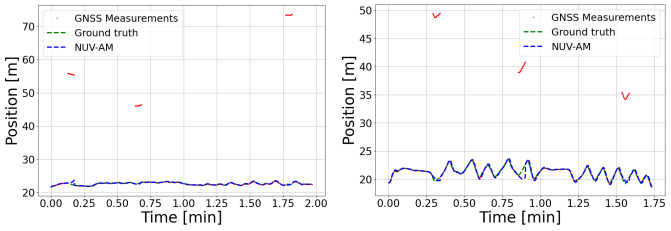



Fig. 7. Measured quadrotor trajectory obtained from the noisy GNSS measurements (red points), with the estimated trajectory by our NUV-AM (blue dashed line), and ground truth (green dashed line). (a) Horizontal direction. (b) Vertical direction.

complexity, which requires tuning multiple parameters. In contrast, the runtime computation for the Student's t distribution is significantly shorter compared with the ORKF because it requires tuning only one parameter. Our algorithms present relatively shorter computation times among outlier detection and weighting algorithms, coupled with low MSE, making them suitable for real-time tasks. In addition, in comparison to our other suggested method, NUV-based EM and NUV-based AM showcase an almost 40% reduced runtime.

### C. API Dataset

To emphasize the versatility and robustness of our approaches in tracking real-world dynamic data of different platforms, which may be corrupted by various types of outliers, we evaluate the API dataset [37]. The API dataset is collected from the *MATRICE 300 quadrotor* platform containing GNSS RTK reading sampled at 10[Hz] and from a marine vessel named *"Shikmona"* containing motion reference units (MRU) with GNSS RTK receiver, sampled at 100[Hz].

Their trajectories were populated with generated outliers, sampled with intensity from a Rayleigh distribution, while their time steps within the data were drawn from a Bernoulli distribution. To filter out this process, we use the same model and parameters as in Section IV-B.

It is evident from Fig. 7 that our OIKF based on NUV-AM algorithm accurately estimates the position, effectively handling all the outliers in both the *vertical* and *horizontal* directions of the quadrotor trajectory.

In Table II, we compare our proposed OIKF based on NUV-EM and based on NUV-AM with the following algorithms: classical KF, reweighted algorithm (ORKF [16]), and $\chi^2$-test [13], [14]. In addition, Table II presents the RMSE and MSE for each algorithm, while the process noise $q^2$ and

TABLE II
POSITION ERROR FOR OPTIMAL VALUES OF $r^2$ AND $q^2$ FOR THE API DATASET--QUADROTOR RECORDINGS

| | Horizontal direction | | Vertical direction | |
|---|---|---|---|---|
| | RMSE[m] | MSE[dB] | RMSE[m] | MSE[dB] |
| Noisy GNSS | 10.4 | 20.3 | 6.4 | 16.1 |
| KF | 0.9 | -1.2 | 1.8 | 5.3 |
| ORKF | 0.8 | -1.9 | 1 | -0.4 |
| $\chi^2$-test | 0.3 | -11.4 | 0.5 | -6.3 |
| NUV-AM | **0.1** | **-17.9** | **0.3** | **-10** |
| NUV-EM | **0.1** | **-17.9** | **0.3** | **-10.2** |

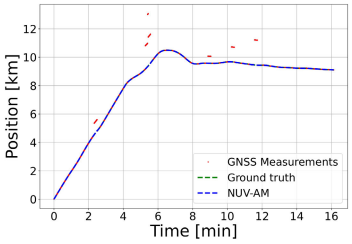



Fig. 8. Measured *"Shikmona"* marine vessel trajectory obtained from the noisy GNSS measurements (red points), with the estimated trajectory by our NUV-AM (blue dashed line), and ground truth (green dashed line).

TABLE III
POSITION ERROR FOR OPTIMAL VALUES OF $r^2$ AND $q^2$ FOR THE API DATASET--MARINE VESSEL RECORDINGS

| | RMSE[m] | MSE[dB] |
|---|---|---|
| Noisy GNSS | 373 | 51 |
| KF | 267 | 48 |
| ORKF | 48 | 34 |
| $\chi^2$-test | 6 | 16 |
| NUV-AM | **2.4** | **7.52** |
| NUV-EM | **2.4** | **7.52** |

observation noise variance $r^2$ of each algorithm are optimized at the same procedure as in Table I. As shown in Table II, OIKF with both NUV-EM and NUV-AM has the lowest estimation errors for both the directions and emphasize the use in NUV method for outlier detection.

Fig. 8 demonstrates the performance of our OIKF-based NUV-AM in estimating the position of the *"Shikmona"* marine vessel. The examined trajectory includes straight line segments and turns. As can be seen, OIKF-based NUV-AM successfully tracks the ground truth and surpasses outliers, even during turns. Table III underscores the superior performance of our proposed OIKF based on NUV-EM and based on NUV-AM compared with the following algorithms: classical KF, reweighted algorithm (ORKF [16]), and $\chi^2$-test [13], [14], revealing significantly low RMSE and MSE values for both our NUV-EM and NUV-AM algorithms.

## V. CONCLUSION

In this work, we have proposed an innovative outlier-insensitive KF that offers improved performance to tackle the problem of state estimation in the presence of outliers. Based on Bayesian learning concepts, we model the outlier as NUV and estimate the unknown variance using either EM or AM algorithms, resulting in sparse outlier detection. Both the algorithms are parameter-free and amount essentially to a short iterative process during the update step of the KF.

Our numerical study demonstrates the effectiveness of our algorithms and highlights the robustness and wide applicability in addressing a variety of applications. We demonstrate superior performances competing with other algorithms in terms of MSE and RMSE across synthetic and real-world datasets. These findings emphasize the robustness and accuracy of our OIKF approach, making it especially suitable for systems reliant on high-quality sensory data.

## ACKNOWLEDGMENT

The authors thank Prof. Hans-Andrea Loeliger for helpful discussions.

## REFERENCES

[1] S. Truzman, G. Revach, N. Shlezinger, and I. Klein, "Outlier-insensitive Kalman filtering using NUV priors," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[2] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*. London, U.K.: Oxford Univ. Press, May 2012.

[3] H. Zhu, K. Zou, Y. Li, and H. Leung, "Robust sensor fusion with heavy-tailed noises," *Signal Process.*, vol. 175, Oct. 2020, Art. no. 107659.

[4] Y. Yuan, Y. Wang, W. Gao, and F. Shen, "Vehicular relative positioning with measurement outliers and GNSS outages," *IEEE Sensors J.*, vol. 23, no. 8, pp. 8556–8567, Apr. 2023.

[5] E. Navon and B. Z. Bobrovsky, "An efficient outlier rejection technique for Kalman filters," *Signal Process.*, vol. 188, Nov. 2021, Art. no. 108164.

[6] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar. 1960.

[7] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. Hoboken, NJ, USA: Wiley, 2009.

[8] S. Farahmand, G. B. Giannakis, and D. Angelosante, "Doubly robust smoothing of dynamical processes via outlier sparsity constraints," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4529–4543, Oct. 2011.

[9] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, "Sparse/robust estimation and Kalman smoothing with nonsmooth log-concave densities: Modeling, computation, and theory," *J. Mach. Learn. Res.*, vol. 14, pp. 2689–2728, Jan. 2013.

[10] N. L. Knight and J. Wang, "A comparison of outlier detection procedures and robust estimation methods in GPS positioning," *J. Navigat.*, vol. 62, no. 4, pp. 699–709, Oct. 2009.

[11] F. Zhu, Z. Hu, W. Liu, and X. Zhang, "Dual-antenna GNSS integrated with MEMS for reliable and continuous attitude determination in challenged environments," *IEEE Sensors J.*, vol. 19, no. 9, pp. 3449–3461, May 2019.

[12] N. Ye and Q. Chen, "An anomaly detection technique based on a chi-square statistic for detecting intrusions into information systems," *Qual. Rel. Eng. Int.*, vol. 17, no. 2, pp. 105–112, Mar. 2001.

[13] A. Lekkas, M. Candeloro, and I. Schjølberg, "Outlier rejection in underwater acoustic position measurements based on prediction errorr," in *Proc. 4th IFAC Workshop Navigat., Guid. Control Underwater Vehicles*, 2015, vol. 48, no. 2, pp. 82–87.

[14] F. van Wyk, Y. Wang, A. Khojandi, and N. Masoud, "Real-time sensor anomaly detection and identification in automated vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1264–1276, Mar. 2020.

[15] J.-A. Ting, E. Theodorou, and S. Schaal, "A Kalman filter for robust outlier detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2007, pp. 1514–1519.

[16] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "An outlier-robust Kalman filter," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 1551–1558.

[17] Y. Tao and S. S. Yau, "Outlier-robust iterative extended Kalman filtering," *IEEE Signal Process. Lett.*, vol. 30, pp. 743–747, 2023.

[18] C. D. Karlgaard and H. Schaub, "Huber-based divided difference filtering," *J. Guid. Control Dyn.*, vol. 30, no. 3, pp. 885–891, May 2007.

[19] M. A. Gandhi and L. Mili, "Robust Kalman filter based on a generalized maximum-likelihood-type estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2509–2520, May 2010.

[20] A. Aravkin, J. V. Burke, L. Ljung, A. Lozano, and G. Pillonetto, "Generalized Kalman smoothing: Modeling and algorithms," *Automatica*, vol. 86, pp. 63–86, Dec. 2017.

[21] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5024–5037, Oct. 2012.

[22] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, "A novel robust student's t-based Kalman filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 1, pp. 1545–1554, Feb. 2017.

[23] G. Wang, Y. Zhang, and X. Wang, "Maximum correntropy Rauch–Tung–Striebel smoother for nonlinear and non-Gaussian systems," *IEEE Trans. Autom. Control*, vol. 66, no. 3, pp. 1270–1277, Mar. 2021.

[24] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, Feb. 2017.

[25] J. Saha and S. Bhaumik, "Robust maximum correntropy Kalman filter," 2023, *arXiv:2302.02694*.

[26] N. Davari and A. P. Aguiar, "Real-time outlier detection applied to a Doppler velocity log sensor based on hybrid autoencoder and recurrent neural network," *IEEE J. Ocean. Eng.*, vol. 46, no. 4, pp. 1288–1301, Oct. 2021.

[27] F. Wadehn, L. Bruderer, J. Dauwels, V. Sahdeva, H. Yu, and H.-A. Loeliger, "Outlier-insensitive Kalman smoothing and marginal message passing," in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2016, pp. 1242–1246.

[28] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[29] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[30] H.-A. Loeliger, L. Bruderer, H. Malmberg, F. Wadehn, and N. Zalmai, "On sparsity by NUV-EM, Gaussian message passing, and Kalman smoothing," in *Proc. Inf. Theory Appl. Workshop (ITA)*, La Jolla, CA, USA, Jan. 2016, pp. 1–10.

[31] H.-A. Loeliger, B. Ma, H. Malmberg, and F. Wadehn, "Factor graphs with NUV priors and iteratively reweighted descent for sparse least squares and more," in *Proc. IEEE 10th Int. Symp. Turbo Codes Iterative Inf. Process. (ISTC)*, Dec. 2018, pp. 1–5.

[32] H.-A. Loeliger, "On NUV priors and Gaussian message passing," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2023, pp. 1–6.

[33] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–22, 1977.

[34] J. A. Palmer, D. P. Wipf, K. Kreutz-Delgado, and B. D. Rao, "Variational EM algorithms for non-Gaussian latent variable models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, pp. 1059–1066.

[35] V. S. A Andresen, "Convergence of an alternating maximization procedure," *J. Mach. Learn. Res.*, vol. 17, no. 63, pp. 1–53, 2016.

[36] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of Michigan North Campus long-term vision and LiDAR dataset," *Int. J. Robot. Res.*, vol. 35, no. 9, pp. 1023–1035, Aug. 2016.

[37] A. Shurin et al., "The autonomous platforms inertial dataset," *IEEE Access*, vol. 10, pp. 10191–10201, 2022.

[38] G. Revach, N. Shlezinger, X. Ni, A. López Escoriza, R. J. G. van Sloun, and Y. C. Eldar, "KalmanNet: Neural network aided Kalman filtering for partially known dynamics," *IEEE Trans. Signal Process.*, vol. 70, pp. 1532–1547, 2022.

[39] G. Revach, X. Ni, N. Shlezinger, R. J. G. van Sloun, and Y. C. Eldar, "RTSNet: Learning to smooth in partially known state-space models," *IEEE Trans. Signal Process.*, vol. 71, pp. 4441–4456, 2023.

[40] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation: THeory Algorithms and Software*. Hoboken, NJ, USA: Wiley, 2004.

[41] H. Ohlsson, F. Gustafsson, L. Ljung, and S. Boyd, "Smoothed state estimates under abrupt changes using sum-of-norms regularization," *Automatica*, vol. 48, no. 4, pp. 595–605, Apr. 2012.

[42] B. M. Bell and F. W. Cathey, "The iterated Kalman filter update as a Gauss-Newton method," *IEEE Trans. Autom. Control*, vol. 38, no. 2, pp. 294–297, Feb. 1993.

[43] J. Humpherys, P. Redd, and J. West, "A fresh look at the Kalman filter," *SIAM Rev.*, vol. 54, no. 4, pp. 801–823, Jan. 2012.

[44] B. Ma, N. Zalmai, R. Torfason, C. Striti, and H.-A. Loeliger, "Color image segmentation using iterative edge cutting, NUV-EM, and Gaussian message passing," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 161–165.

[45] P. J. Rousseeuw and M. Hubert, "Robust statistics for outlier detection," *Data Mining Knowl. Discovery*, vol. 1, no. 1, pp. 73–79, 2011.

[46] A. Shurin and I. Klein, "QuadNet: A hybrid framework for quadrotor dead reckoning," *Sensors*, vol. 22, no. 4, p. 1426, Feb. 2022.

**Shunit Truzman** received the B.Sc. degree in aerospace engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 2019. She is currently pursuing the M.Sc. degree with the Autonomous Navigation and Sensor Fusion Laboratory, The Hatter Department of Marine Technologies, University of Haifa, Haifa.

Her main research interests are anomaly detection problems and solutions from the intersection of statistical signal processing with machine learning algorithms.

**Guy Revach** (Senior Member, IEEE) received the B.Sc. (cum laude) and M.Sc. degrees in electrical and computer engineering from the Andrew and Erna Viterbi Department of Electrical and Computer Engineering at the Technion—Israel Institute of Technology, Haifa, Israel, in 2008 and 2017, respectively. He is now pursuing the Ph.D. degree in electrical and computer engineering with the Institute for Signal and Information Processing, ETH Zürich, Switzerland, under the supervision of Prof. Dr. Hans-Andrea Loeliger.

His master's thesis, supervised by Prof. Nahum Shimkin, focused on planning for cooperative multiagents.

**Nir Shlezinger** (Senior Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical and computer engineering, from Ben-Gurion University, Beersheba, Israel, in 2011, 2013, and 2017, respectively.

From 2017 to 2019, he was a Postdoctoral Researcher with the Technion, Haifa, Israel. From 2019 to 2020, he was a Postdoctoral Researcher with the Weizmann Institute of Science, Rehovot, Israel. He is currently an Assistant Professor with the School of Electrical and Computer Engineering, Ben-Gurion University. His research interests include communications, information theory, signal processing, and machine learning.

Dr. Shlezinger was the recipient of the FGS Prize for outstanding research achievements from the Weizmann Institute of Science.

**Itzik Klein** (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in aerospace engineering and the Ph.D. degree in geo-information engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 2004, 2007, and 2011, respectively.

He is an Associate Professor, with the Autonomous Navigation and Sensor Fusion Laboratory, Charney School of Marine Sciences, Hatter Department of Marine Technologies, University of Haifa, Haifa. Prior to joining the University of Haifa, he worked at leading companies in Israel on navigation topics for more than 15 years. He has a wide range of experience in navigation systems and sensor fusion from both industry and academic perspectives. His research interests lie in the intersection of artificial intelligence with inertial sensing, sensor fusion, and robotics.

Dr. Klein is an Editorial Board Member of the IEEE Journal of Indoor and Seamless Positioning and Navigation (J-ISPIN). He is currently a Distinguished Lecturer for the IEEE Oceanic Engineering Society.